Using MPOG data for external validation

Weill Cornell Medicine

Erasmus MC
Sophia Kinderziekenhuis

Virginia Tangel, MA MSc

Clinical Epidemiologist, PhD candidate

v.tangel@erasmusmc.nl

MPOG
MULTICENTER PERIOPERATIVE OUTCOMES GROUP

BJA

British Journal of Anaesthesia, 133 (6): 1222–1233 (2024)

doi: 10.1016/j.bja.2024.09.003
Advance Access Publication Date: 29 October 2024
Clinical Investigation

PAEDIATRIC ANAESTHESIA

International multi-institutional external validation of preoperative risk scores for 30-day in-hospital mortality in paediatric patients

Virginia E. Tangel[1,2,*], Sanne E. Hoeks[1], Robert Jan Stolker[1], Sydney Brown[3], Kane O. Pryor[2], Jurgen C. de Graaff[1,2,4], and Multicenter Perioperative Outcomes Group (MPOG) Perioperative Clinical Research Committee[†]

[1]Department of Anesthesiology, Erasmus University Medical Centre, Rotterdam, The Netherlands, [2]Department of Anesthesiology, Weill Cornell Medicine, New York, NY, USA, [3]Department of Anesthesiology, University of Michigan, Ann Arbor, MI, USA and [4]Department of Anesthesiology, Adrz–Erasmus MC, Goes, The Netherlands

*Corresponding author. E-mail: v.tangel@erasmusmc.nl
[†]Members of the Multicenter Perioperative Outcomes Group Perioperative Clinical Research Committee are listed in Appendix 1.
Prior presentations: Presented to the Perioperative Clinical Research Committee (PCRC) of the Multicenter Perioperative Outcomes Group (MPOG) on March 14, 2022, and approved following electronic revisions on April 12, 2022. Results were presented as an abstract at Euroanaesthesia 2024 (May 26, 2024, Munich, Germany).

MPOG

Objective

- Validate existing patient-level risk prognostic prediction scores for pediatric 30-day mortality which we had previously rated as having low risks of bias per the PROBAST methodology

  - **2 scores:**

    - **Pediatric Risk Assessment Score (PRAm)** (Nasr et al. 2017)
    - **Intrinsic surgical risk score** (Nasr et al. 2019)

Risk scores

Pediatric Risk Assessment (PRAm)

Table 3. PRAm Score to Predict Postoperative Mortality

Variables	Definition	Value
Urgent	Urgent surgical procedure	+1
Comorbidity	The presence of at least 1 comorbidity among the following: respiratory disease, congenital heart disease, preoperative acute or chronic kidney disease, neurologic disease, hematologic disease	+2
Critically ill	The presence of at least 1 of the following characteristics of critical illness: preoperative mechanical ventilation, inotropic support, preoperative cardiopulmonary resuscitation	+3
Age <12 mo	Age at the time of the surgical procedure <12 mo	+3
Neoplasm	Surgical procedure in a patient with a neoplasm with or without preoperative chemotherapy	+4

Abbreviation: PRAm, Pediatric Risk Assessment.

Intrinsic surgical risk score

Table 2. Multivariable Model for 30-Day Mortality

	Odds Ratio	95% CI	*P* Value
Neonate (age ≤ 28 days)	0.85	(0.72 to 1.01)	0.052
Weight < 5 kg	1.56	(1.31 to 1.85)	< 0.001
Sex–female	0.91	(0.80 to 1.03)	0.130
ASA PS			
I	Reference	–	–
II	3.4	(1.2 to 9.6)	0.021
III	17.6	(6.4 to 48.4)	< 0.001
IV	69.4	(25.1 to 191.7)	< 0.001
V	296.7	(105.4 to 834.8)	< 0.001
Sepsis	2.1	(1.8 to 2.4)	< 0.001
Inotropic support	2.9	(2.5 to 3.5)	< 0.001
Congenital heart disease	0.94	(0.82 to 1.1)	0.344
Ventilator dependence	2.6	(2.2 to 3.1)	< 0.001
CPT risk quartile			
1	Omitted (no mortalities)	–	–
2	Reference	–	–
3	4.4	(2.9 to 6.5)	< 0.001
4	7.9	(5.2 to 11.9)	< 0.001

The 2012 to 2016 National Surgical Quality Improvement Program Pediatrics data provided a sample size of 367,065, after excluding cardiac procedures and CPT codes with fewer than 25 occurrences.

ASA PS, American Society of Anesthesiologists Physical Status classification; CPT, Current Procedural Terminology.

POG

Risk scores

Pediatric Risk Assessment (PRAm)

Table 3. PRAm Score to Predict Postoperative Mortality

Variables	Definition	Value
Urgent	Urgent surgical procedure	+1
Comorbidity	The presence of at least 1 comorbidity among the following: respiratory disease, congenital heart disease, preoperative acute or chronic kidney disease, neurologic disease, hematologic disease	+2
Critically ill	The presence of at least 1 of the following characteristics of critical illness: preoperative mechanical ventilation, inotropic support, preoperative cardiopulmonary resuscitation	+3
Age <12 mo	Age at the time of the surgical procedure <12 mo	+3
Neoplasm	Surgical procedure in a patient with a neoplasm with or without preoperative chemotherapy	+4

Abbreviation: PRAm, Pediatric Risk Assessment.

Intrinsic surgical risk score

Table 2. Multivariable Model for 30-Day Mortality

	Odds Ratio	95% CI	*P* Value
Neonate (age ≤ 28 days)	0.85	(0.72 to 1.01)	0.052
Weight < 5 kg	1.56	(1.31 to 1.85)	< 0.001
Sex–female	0.91	(0.80 to 1.03)	0.130
ASA PS			
I	Reference	–	–
II	3.4	(1.2 to 9.6)	0.021
III	17.6	(6.4 to 48.4)	< 0.001
IV	69.4	(25.1 to 191.7)	< 0.001
V	296.7	(105.4 to 834.8)	< 0.001
Sepsis	2.1	(1.8 to 2.4)	< 0.001
Inotropic support	2.9	(2.5 to 3.5)	< 0.001
Congenital heart disease	0.94	(0.82 to 1.1)	0.344
Ventilator dependence	2.6	(2.2 to 3.1)	< 0.001
CPT risk quartile			
1	Omitted (no mortalities)	–	–
2	Reference	–	–
3	4.4	(2.9 to 6.5)	< 0.001
4	7.9	(5.2 to 11.9)	< 0.001

The 2012 to 2016 National Surgical Quality Improvement Program Pediatrics data provided a sample size of 367,065, after excluding cardiac procedures and CPT codes with fewer than 25 occurrences.

ASA PS, American Society of Anesthesiologists Physical Status classification; CPT, Current Procedural Terminology.

POG

Methods

1. Calculated published risk scores for each patient
2. Updated (re-estimated) models using MPOG data

For both scores:

- **Discrimination:** ROC, precision recall curve
- **Calibration:** calibration plot
- **Clinical utility:** decision curve

Methods

1. Calculated published risk scores for each patient
2. Updated (re-estimated) models using MPOG data

For both scores:

- **Discrimination:** ROC, precision recall curve
- **Calibration:** calibration plot
- **Clinical utility:** decision curve

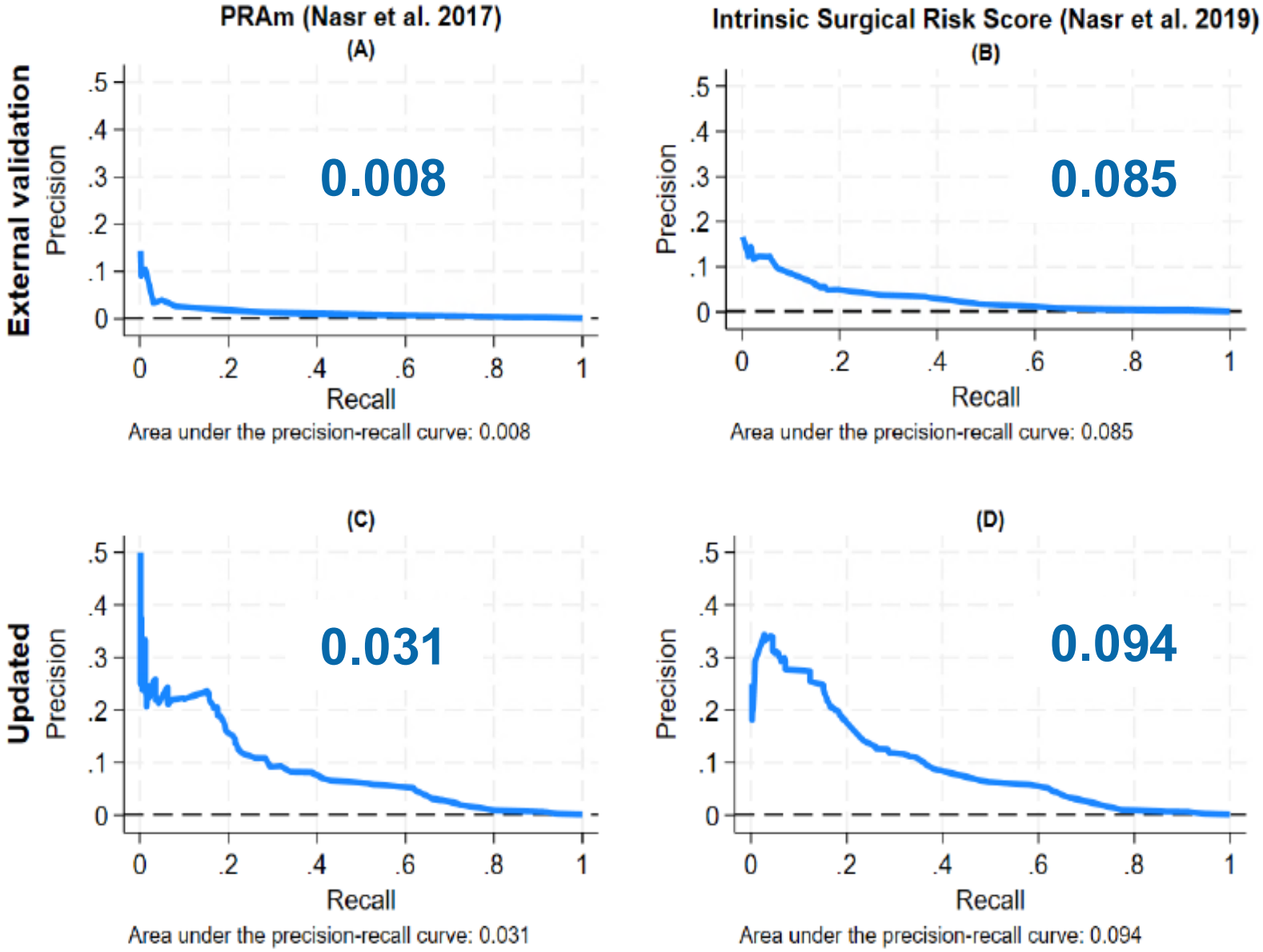
Results

- Cumulative incidence of 30-day mortality: **0.14%**
- **N=822** / 608,488 cases clustered within 56 hospitals

	Overall (N=608,488)	Non-survivor (N=822)	Survivor (N=605,666)
Neonate (age <=28 days)	6,266	167 (2.7)	6,099 (97.3)
Weight < 5 kg	16,030	253 (1.6)	15,777 (98.4)
Female sex	258,883	335 (0.1)	258,548 (99.9)
ASA Physical Status			
I	200,396	18 (> 0.0)	200,378 (100.0)
II	293,575	57 (> 0.0)	293,518 (100.0)
III	102,884	241 (0.2)	102,643 (99.8)
IV	9,094	381 (4.2)	8,713 (95.8)
V	531	125 (23.5)	406 (76.5)

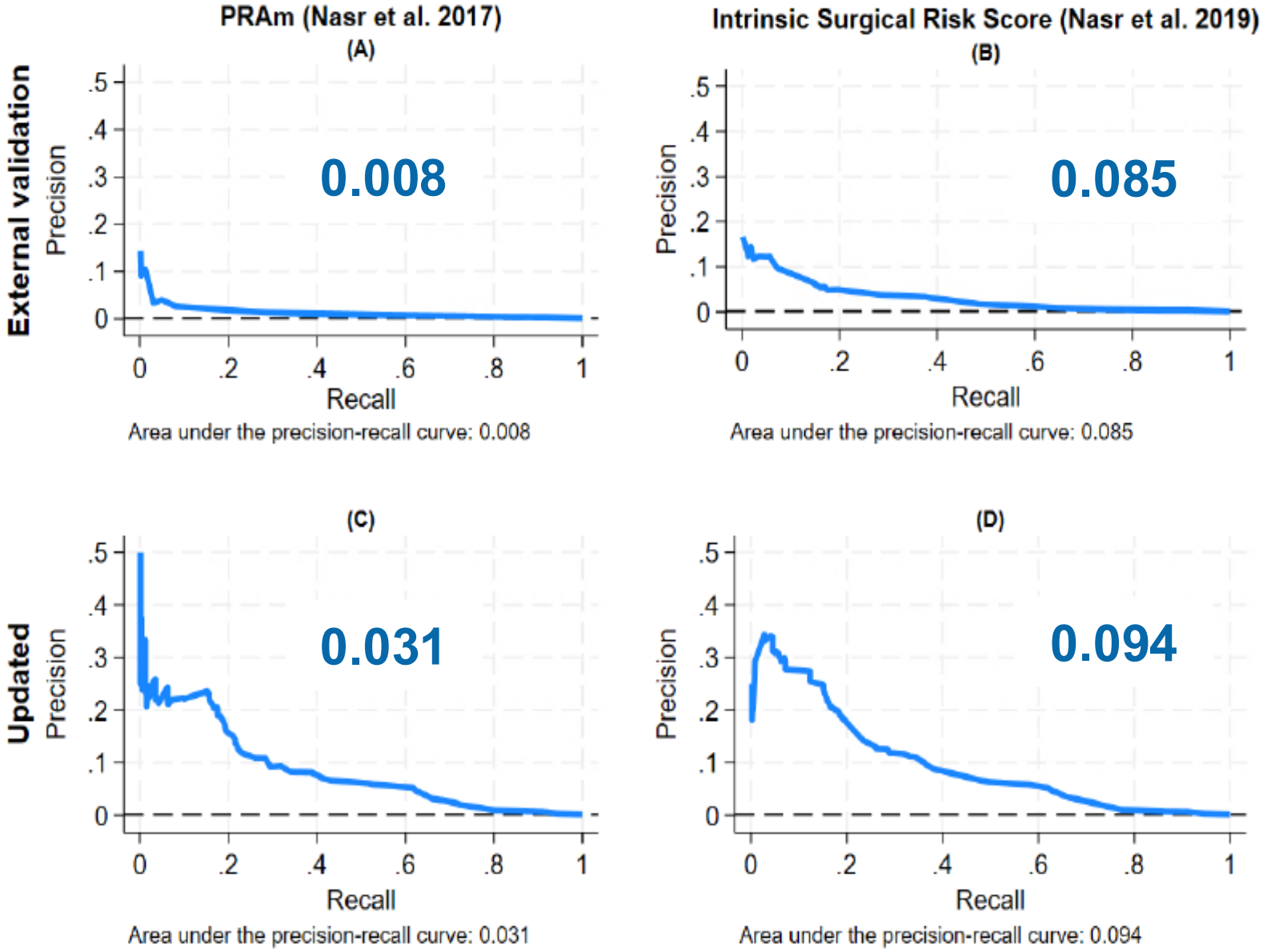
Model discrimination

- Precision-recall curve
- Random performance = 0.0014



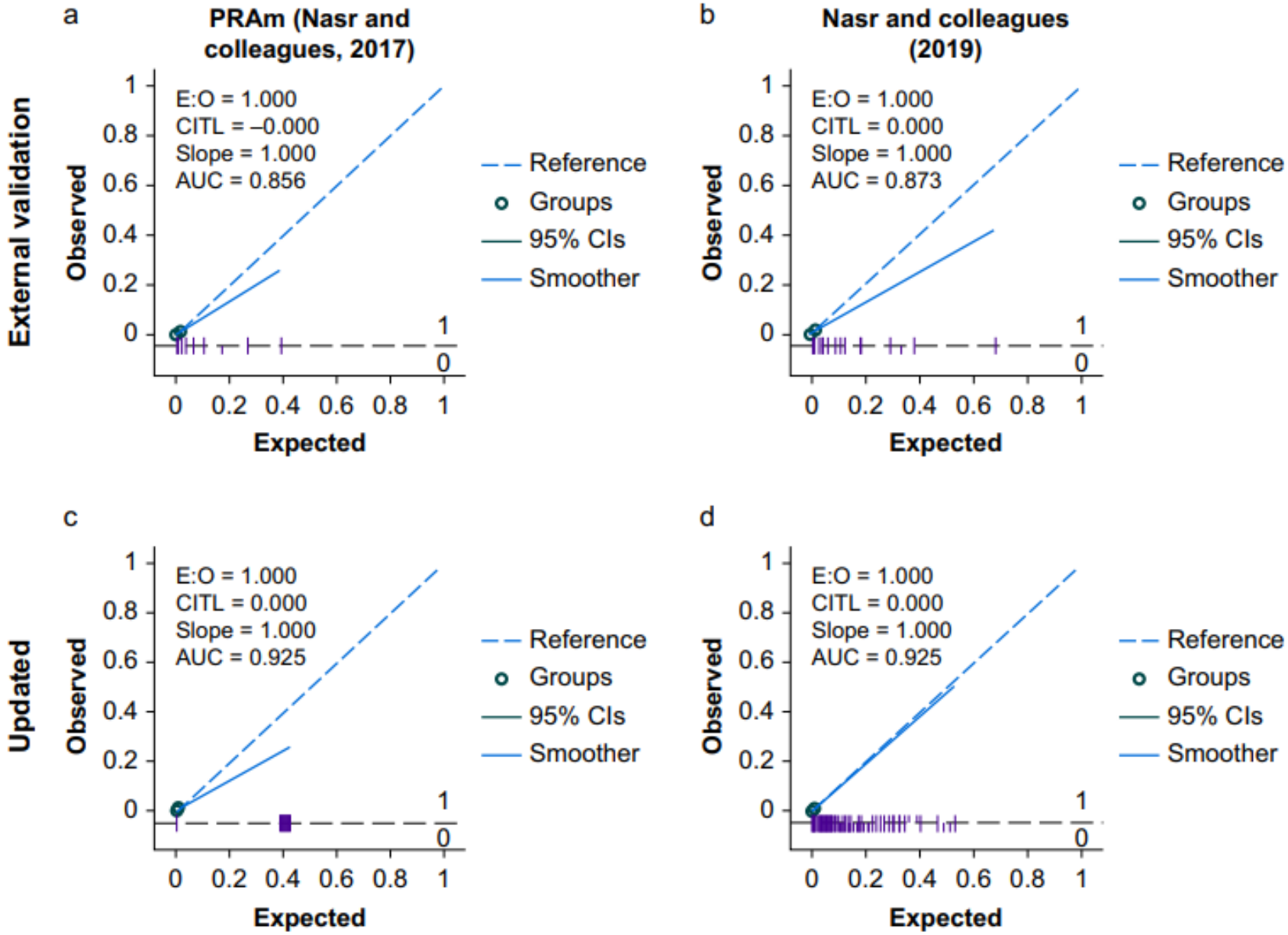
Model discrimination

- Precision-recall curve
- Random performance = 0.0014



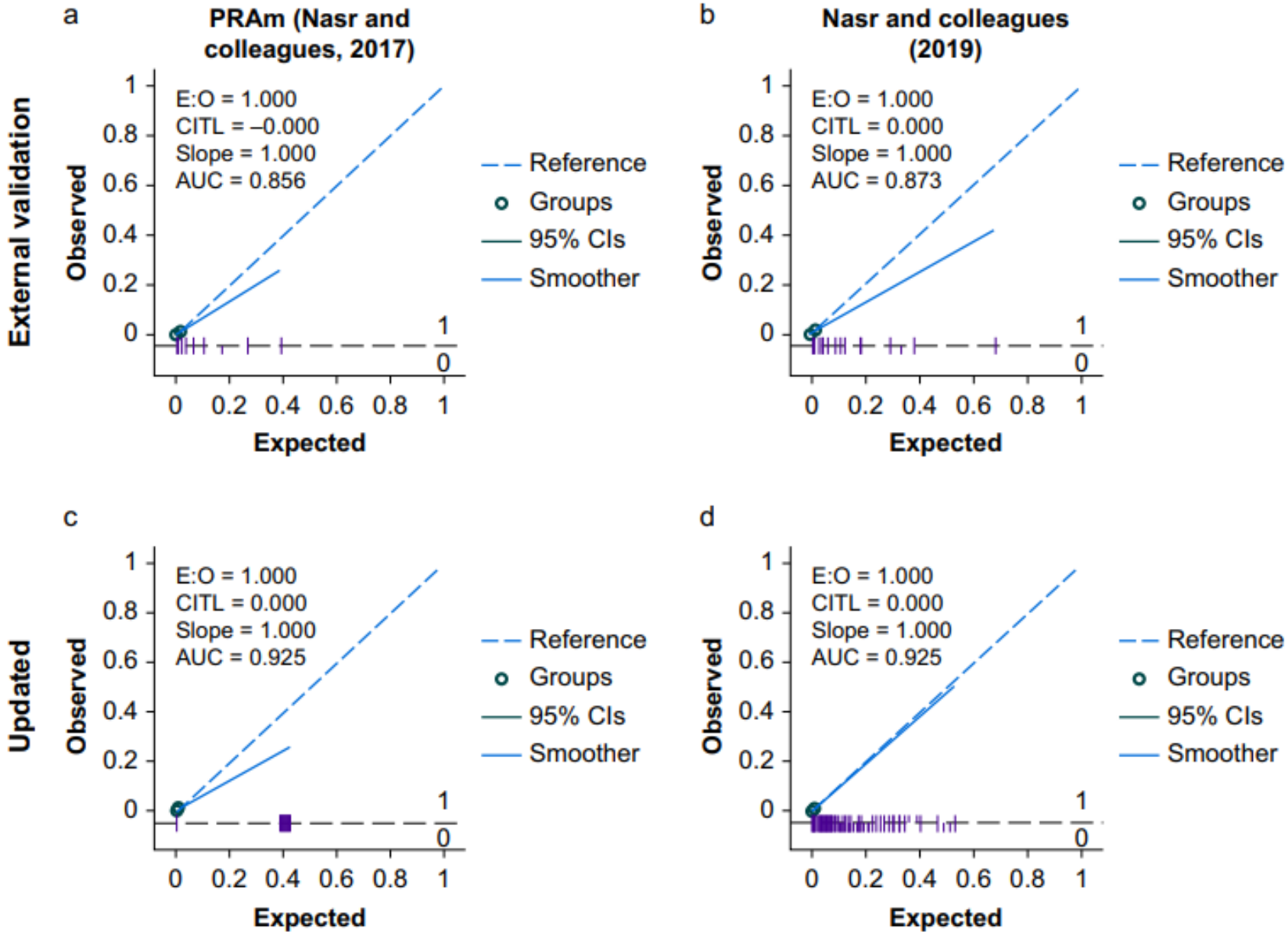
Model calibration

Calibration plots



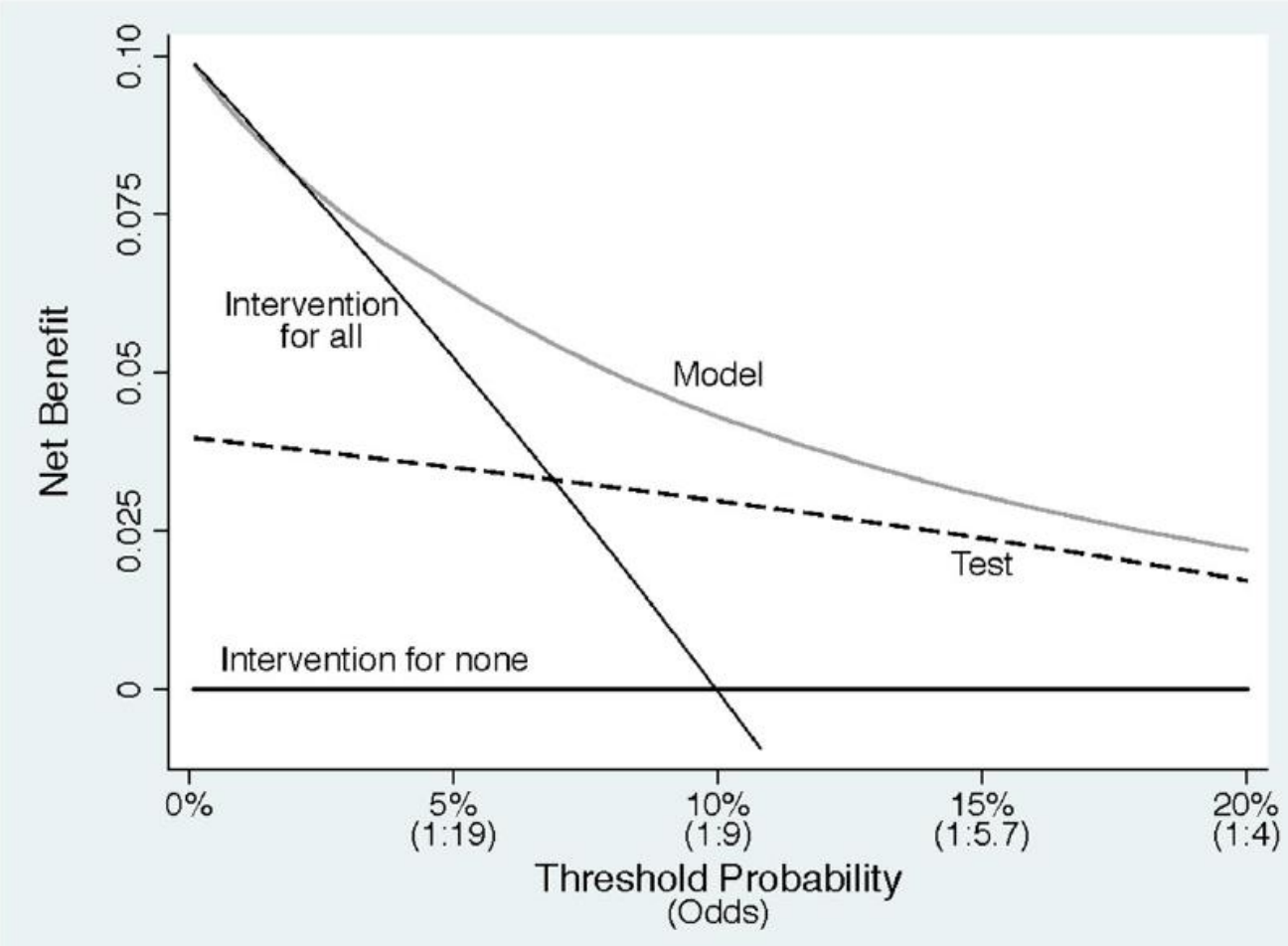
Model calibration

Calibration plots

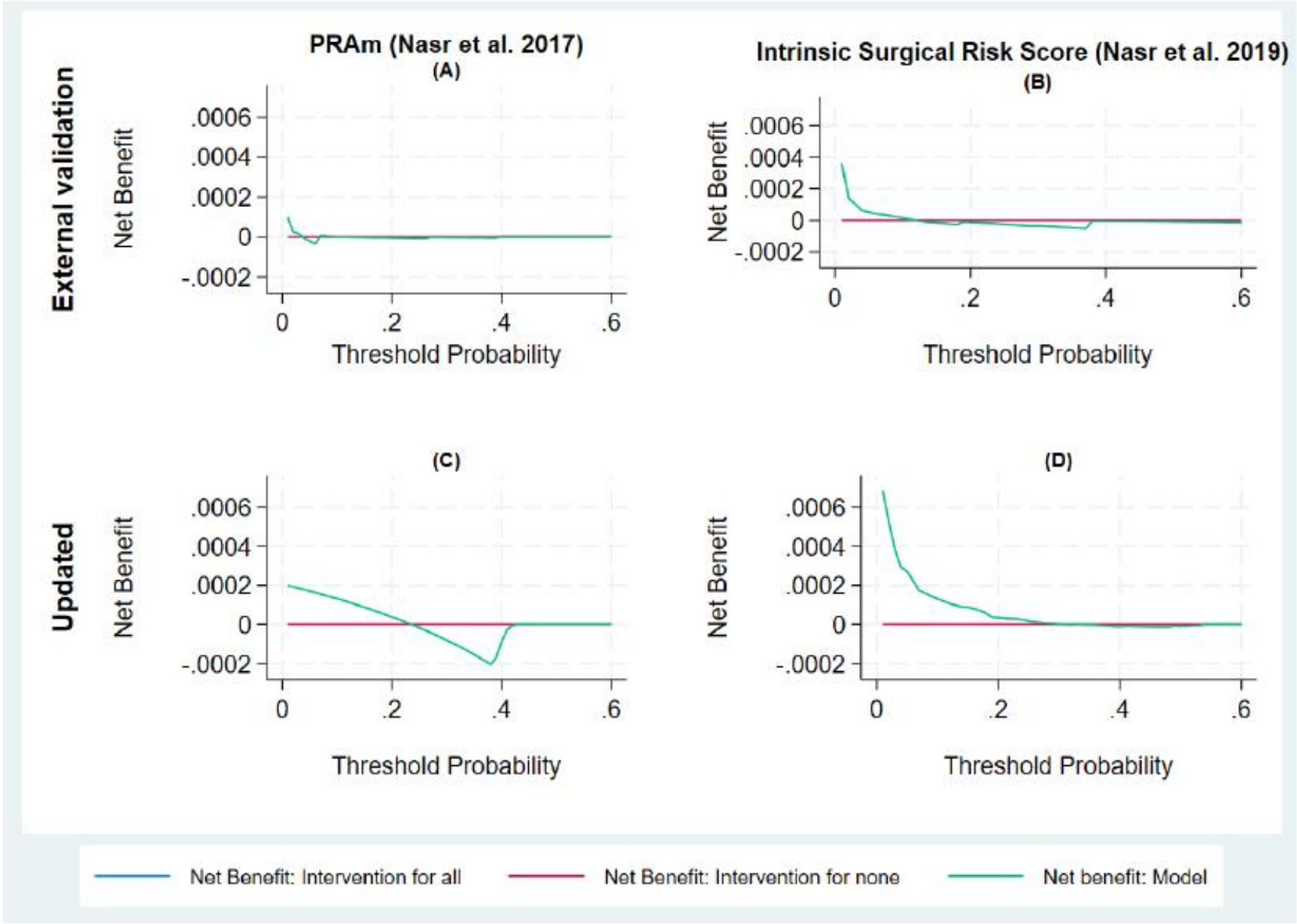


Assessing clinical utility

- Decision curve analysis



Vickers, Andrew J., Ben van Calster, and Ewout W. Steyerberg. "A simple, step-by-step guide to interpreting decision curve analysis." *Diagnostic and prognostic research* 3.1 (2019): 1-8.


External validation
Updated
PRAm (Nasr et al. 2017)
(A)
Intrinsic Surgical Risk Score (Nasr et al. 2019)
(B)
(C)
(D)
Net Benefit
.0006
.0004
.0002
0
-.0002
0
.2
.4
.6
Threshold Probability
Net Benefit: Intervention for all
Net Benefit: Intervention for none
Net benefit: Model

Summary of findings

- Performance inferior compared to original studies
- The **intrinsic surgical risk score** exhibited better discrimination than the **PRAm**
- Both scores resulted in large numbers of false positives
- 30-day mortality was over-estimated at higher probabilities in all models except for the recalibrated intrinsic surgical risk score
- Decision curve analyses showed limited benefit to using either score in clinical practice

Study conclusions

- Why the overall poor performance in this external validation compared to the original studies?
  - Accounting for differences between MPOG and NSQIP databases
  - Difficulties in defining measurements with MPOG data
- What is the utility of a risk score if it relies on another risk score as a predictor?
- Sicker patients have a great influence on model performance
  - aOR for ASA PS V in original study: 297; in updated model: 2471!
- Is using a risk score to predict the **rare outcome** of 30-day mortality in children clinically useful?

Study conclusions

- Why the overall poor performance in this external validation compared to the original studies?
  - Accounting for differences between MPOG and NSQIP databases
  - Difficulties in defining measurements with MPOG data
- What is the utility of a risk score if it relies on another risk score as a predictor?
- Sicker patients have a great influence on model performance
  - aOR for ASA PS V in original study: 297; in updated model: 2471!
- Is using a risk score to predict the **rare outcome** of 30-day mortality in children clinically useful?

LESSONS LEARNED: USING THE MPOG DATA FOR EXTERNAL VALIDATION

Types of external validation studies

- **Temporal validation:** more recent patients
- **Geographic validation:** multisite testing
- **Fully independent validation:** other investigators at other sites, different definitions of predictors, study patients differently selected

Types of external validation studies

- **Temporal validation:** more recent patients
- **Geographic validation:** multisite testing
- **Fully independent validation:** other investigators at other sites, different definitions of predictors, study patients differently selected

Lessons learned: differences in databases

- Conducting an external validation using an entirely different database from the original study
- Comparison between NSQIP and MPOG databases:

	NSQIP	MPOG
Available measures	No 1:1 concordance with MPOG variables	No 1:1 concordance with NSQIP variables
Population	Sample of high-risk surgical procedures	All procedures with anesthesia
Data collection method	Manual data collected by trained research nurse	Data collected directly from EMR
Outcome	Comprehensive outcome data available	Less comprehensive outcome data available
Specific to this validation study	*Has data on preop CPR & preop meds (with associated timing)*	*Less comprehensive data if not documented on anesthesia record*

MPOG

Lessons learned: differences in databases

- Conducting an external validation using an entirely different database from the original study
- Comparison between NSQIP and MPOG databases:

	NSQIP	MPOG
Available measures	No 1:1 concordance with MPOG variables	No 1:1 concordance with NSQIP variables
Population	Sample of high-risk surgical procedures	All procedures with anesthesia
Data collection method	Manual data collected by trained research nurse	Data collected directly from EMR
Outcome	Comprehensive outcome data available	Less comprehensive outcome data available
Specific to this validation study	*Has data on preop CPR & preop meds (with associated timing)*	*Less comprehensive data if not documented on anesthesia record*

MPOG

Lessons learned: Differences in procedure text documentation

MPOG:

@laparoscopy, appendicocecostomy, robot assist (wrvu 19.91) (n/a flank); modifier robot,davinci xi (n/a)
cecostomy laparoscopic [phi] tube with 5mm instruments;exam under anesthesia (eua) with injection of botox;removal rectal obstruction;appendectomy laparoscopic;dental extractions
right laparoscopic appendectom
intestine, laparoscopy creation cecostomy;skin level -peds (n/a); appendix, laparoscopy appendectomy-peds (n/a); rectum, remove fecal impaction-peds (n/a)
intestine, laparoscopy creation cecostomy;skin level -peds (n/a); appendix, laparoscopy, unlisted procedure; appendix (n/a)
intestine, laparoscopy creation cecostomy;skin level -peds (n/a); rectum, remove fecal impaction-peds (n/a); appendix, laparoscopy, unlisted procedure; appendix (n/a)
laparoscopic assisted appendicostom

Nasr et al. (2019):

Risk Quartile 3

1. Arthrodesis
2. Arthrotomy (hip infection)
3. Brain tumor resection/open or endoscopy
4. Bronchoscopy (foreign body removal)
5. Colectomy for congenital megacolon
6. Craniectomy with cervical laminectomy
7. Craniotomy: electrode placement for seizure monitoring
8. Cystoscopy and ureteroscopy with stent placement
9. Cystostomy with drainage
10. Diagnostic thoracoscopy (mediastinum)
11. Diverticulum
12. Enterostomy closure
13. Enterotomy/foreign body
14. Enterolysis
15. Excision of submandibular gland

Lessons learned: Takeaways

- Why conduct an external validation with the MPOG data?
  - Comprehensive and representative: Census of anesthesia records
  - Data directly from EMR
- Roadblocks to validation with the MPOG data
  - Not all desired data points are captured in the anesthesia record
  - Measures not necessarily equivalent (or absent in some cases)

Thank you/ Dank aan:

Weill Cornell:

Jurgen de Graaff
Hugh Hemmings
Kane Pryor
Zachary Turnbull
Silis Jiang
Samson Obembe

Erasmus MC:

Sanne Hoeks
Robert Jan Stolker

Michigan:

Sydney Brown
Sachin Kheterpal
Mike Mathis
David Clark

MPOG