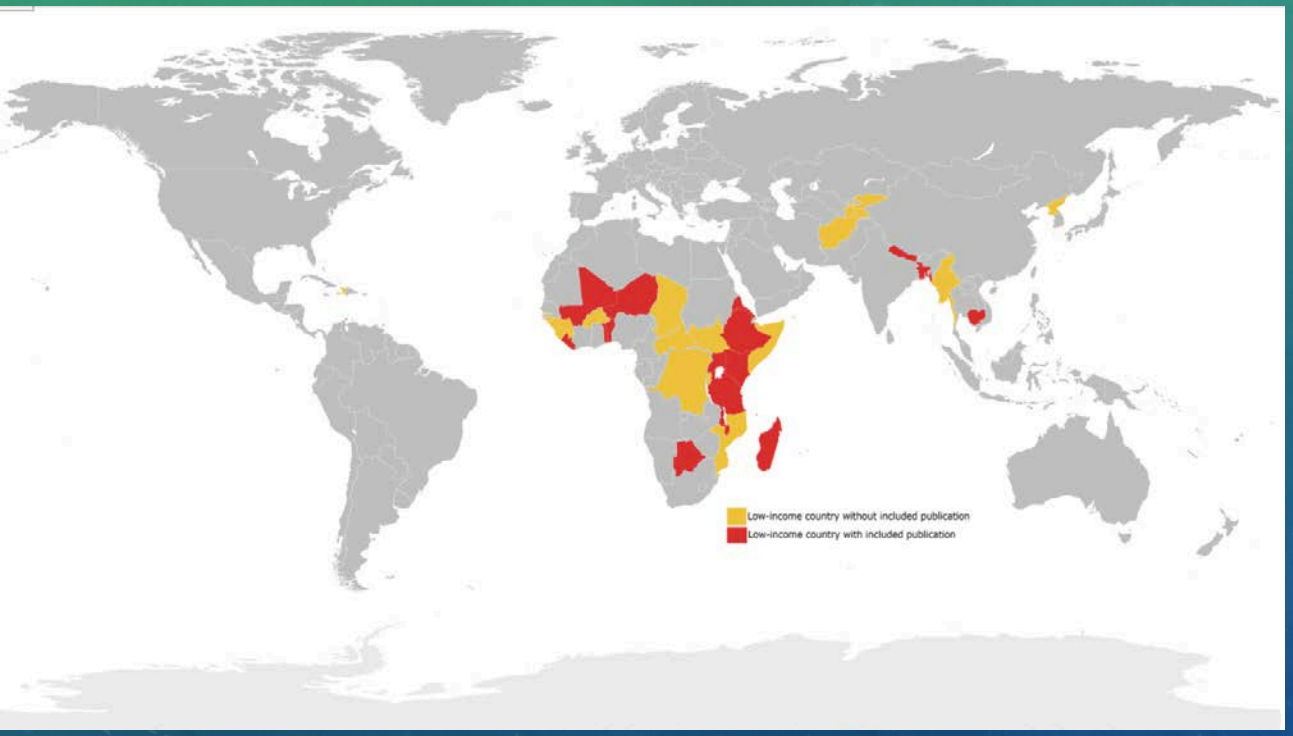
THE VALIDITY OF ICU SCORING SYSTEMS IN LOW AND MIDDLE INCOME COUNTRIES: A STRUCTURED REVIEW

DILIP JAYASIMHAN

GLOBAL BURDEN OF CRITICAL ILLNESS

- Difficult to quantify
- Generally under-appreciated
- Low- and middle-income countries (LMICs) lacking capacity to deal with burden

CAPACITY IN LOW-INCOME COUNTRIES



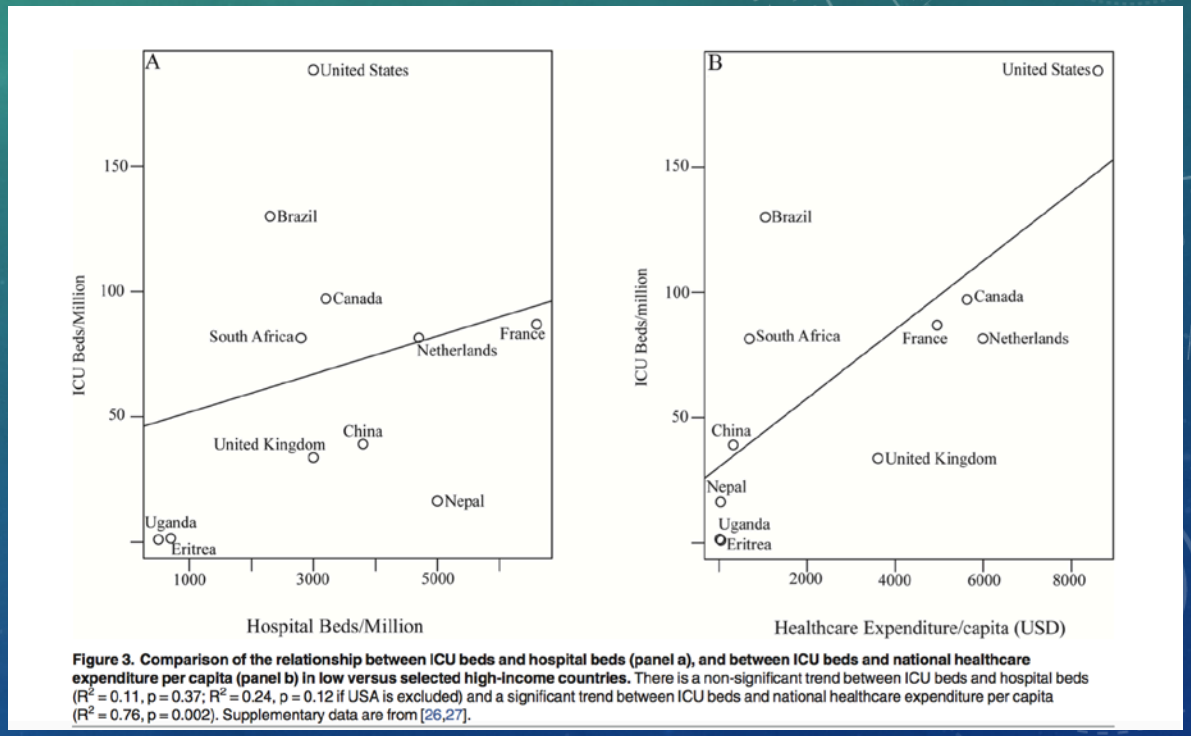
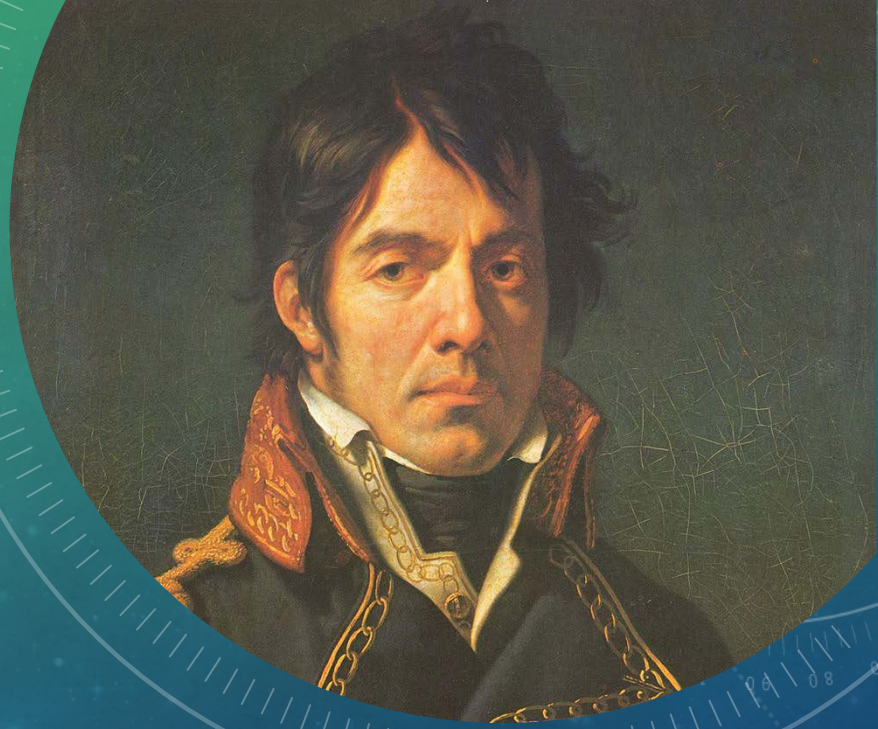


Figure 3. Comparison of the relationship between ICU beds and hospital beds (panel a), and between ICU beds and national healthcare expenditure per capita (panel b) in low versus selected high-income countries. There is a non-significant trend between ICU beds and hospital beds ($R^2 = 0.11$, $p = 0.37$; $R^2 = 0.24$, $p = 0.12$ if USA is excluded) and a significant trend between ICU beds and national healthcare expenditure per capita ($R^2 = 0.76$, $p = 0.002$). Supplementary data are from [26,27].

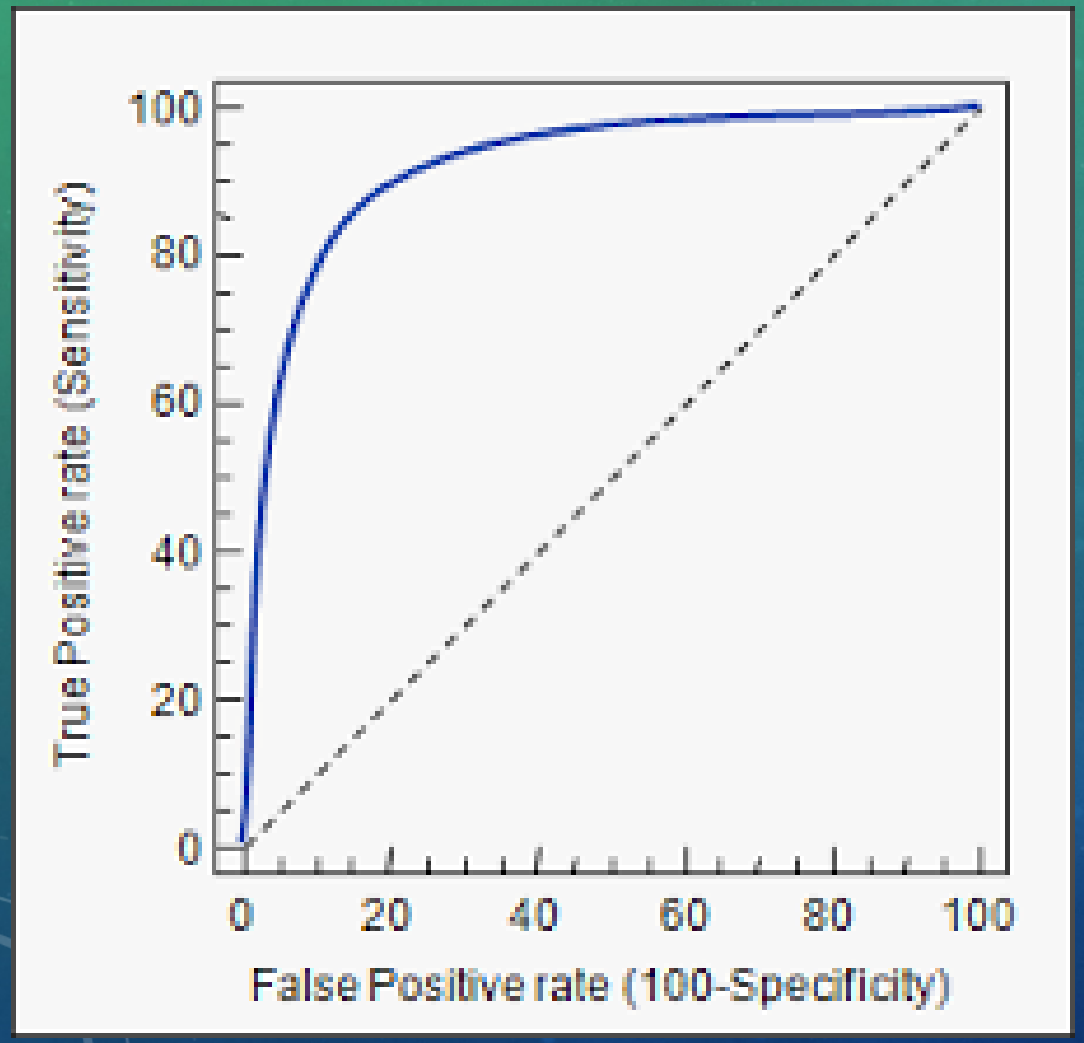
Murthy et al. (2015) Intensive Care Unit Capacity in Low-Income Countries: A Systematic Review

ICU SCORING SYSTEMS

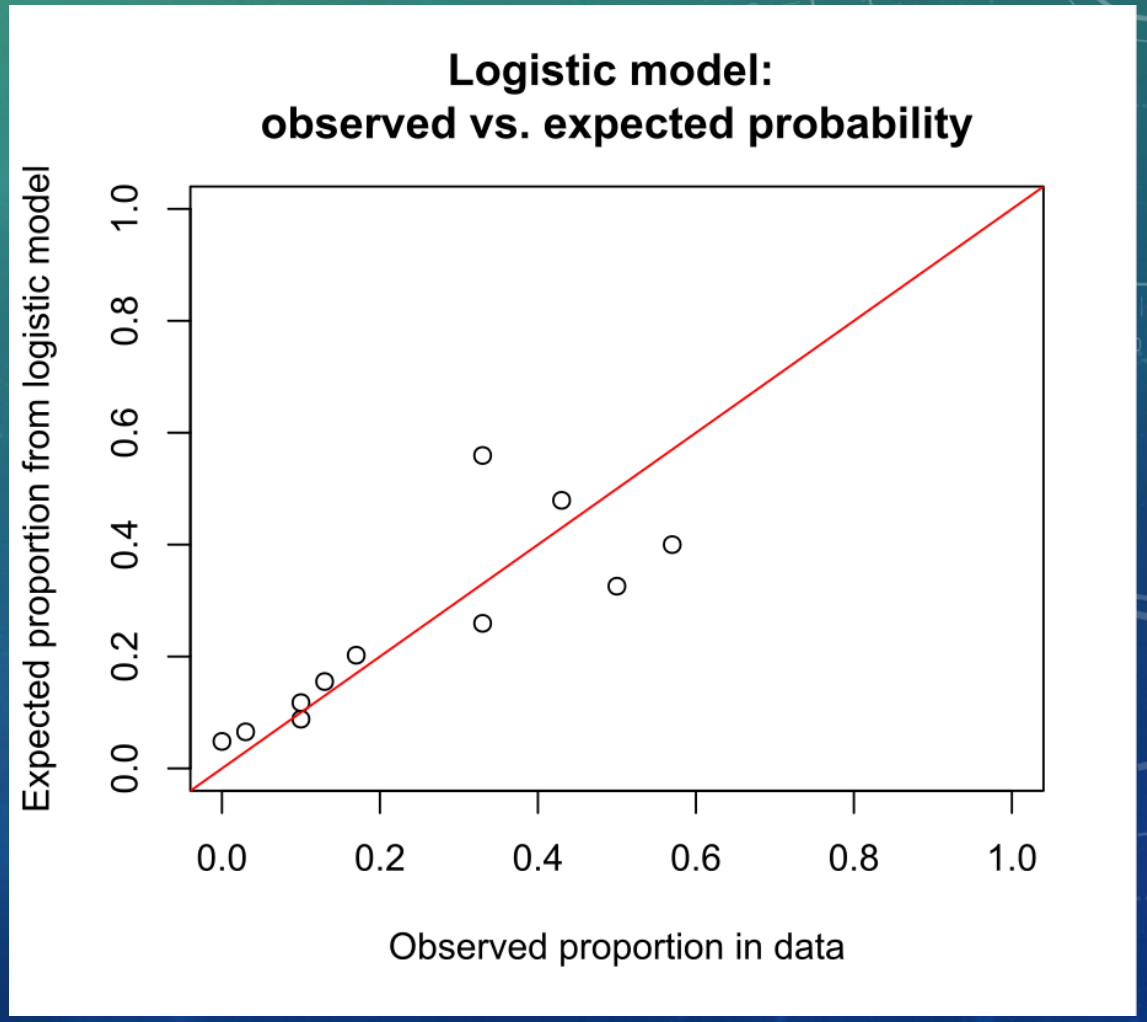
- Triage admissions.
- Benchmarking
- Estimating disease burden
- Morbidity quantification
- Risk stratification & randomisation
- Developed with large population divided to 2 cohorts



DISCRIMINATION



CALIBRATION



BARRIERS IN THE LMIC SETTING

- Significant resources to develop scoring systems
- Most established scoring systems developed with high income country cohorts.
- Widely used in LMICs, ***but are they still applicable?***

METHODS – SEARCH STRATEGY

Terms used		
ICU	Illness severity score	Low and middle income countries
Intensive care	APACHE	Developing countries
Intensive care unit	Mortality prediction model	LMIC
Critical care	SAPS	
	SOFA	
	MPM	
	MODS	
	LODS	
	Mortality prediction	
	Illness severity index	
	Organ failure score	

STUDY INCLUSION & EXCLUSION

Inclusion

- Age >16
- Articles available in the English language.
- Primary/secondary objective: validate the use of a scoring system in adult ICU populations in an LMIC to predict mortality.

Exclusion

- Trauma scoring systems/other disease specific scoring systems without inclusion of general scoring systems.

RESULTS

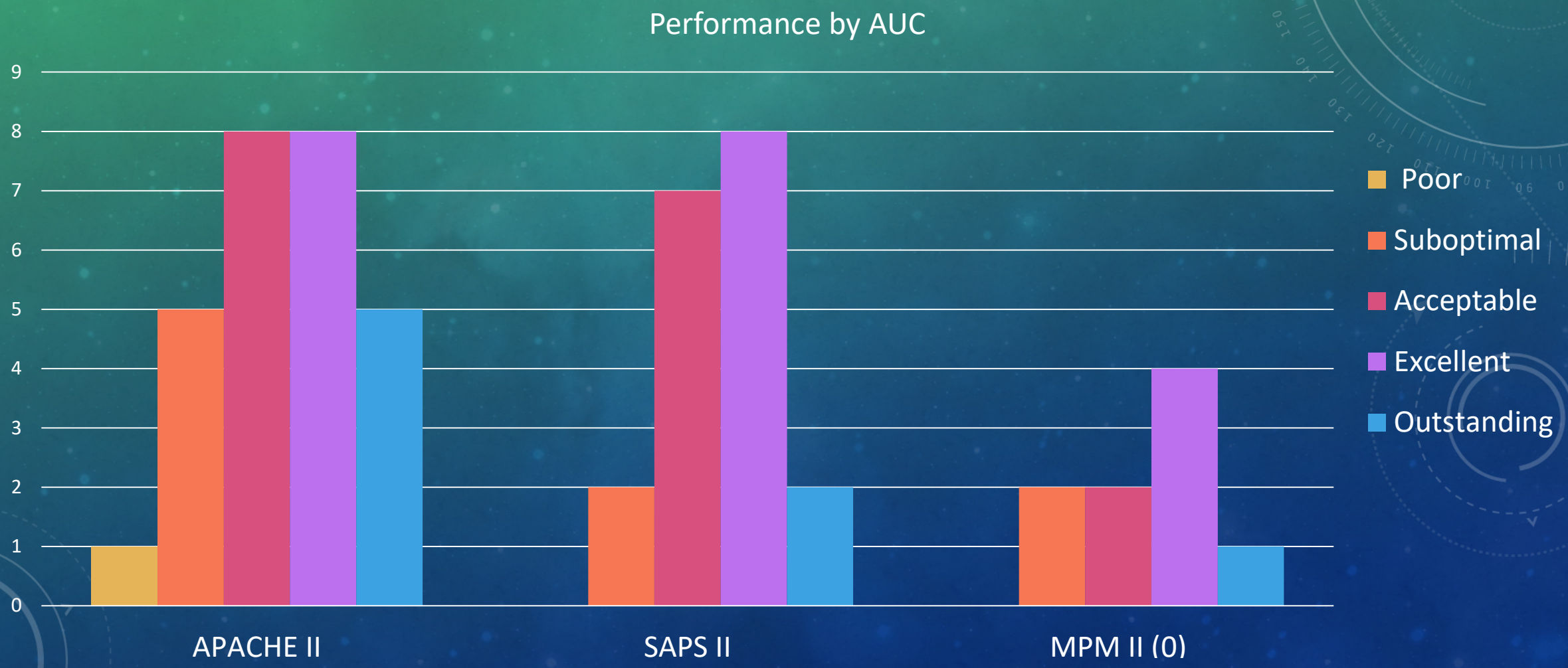
- 42 studies identified
- 26 prospective single centre cohorts , 10 retrospective single centre cohorts, and 6 prospective multicentre cohorts.
- Median population size: 273
- Older generation scoring systems more commonly assessed.

DISCRIMINATION - AUC

Scoring system	Poor (AUC ≤0.5)	Suboptimal (AUC 0.51-0.7)	Acceptable (AUC 0.71-0.8)	Excellent (AUC 0.81-0.9)	Outstanding (AUC >0.9)
APACHE II	1	5	8	8	5
SAPS II	0	2	7	8	2
MPM II (0)	0	2	2	4	1
APACHE III	0	1	1	1	1
SAPS 3	0	0	1	3	2
MPM III (0)	0	0	2	1	1
APACHE IV	0	0	2	0	3
All scoring systems	1	14	30	38	18

DISCRIMINATION

DISCRIMINATION OF MOST COMMON SCORING SYSTEMS



CALIBRATION

Scoring system	Good Calibration (P>0.05)	Poor Calibration (P≤0.05)
APACHE II	9	8
SAPS II	6	10
MPM II (0)	3	4
APACHE III	1	1
SAPS 3	5	4
MPM III (0)	2	2
APACHE IV	2	1
All Scoring systems	32	35

SUMMARY OF FINDINGS

- Most scores showed good discrimination
- Calibration was suboptimal with variation in findings
- Most studies are single centre cohorts
- Older generation scoring systems more commonly assessed.

GOOD DISCRIMINATION

- Physiological variables
- Correlation with adverse prognosis
- Scores developed for general populations regardless of underlying disease.

SUBOPTIMAL CALIBRATION

- Different case mix , unit specific factors, and data collection
- Similar issues observed in HICs. Scores require periodic recalibration.
- Utility from clinical perspective may be minimal.
- Potential use in better understanding differences between HICs and LMICs

SINGLE CENTRE COHORTS

- Small sample sizes make type 2 error more likely
- Significant differences in performance of scoring systems and actual outcomes may not be detected.
- Original validation cohorts of major scores had 4000-6000 events/admissions.

MOVING FORWARD

- Utility of research networks and registries in validation and potentially developing scores
- E.g. BRICNET (Brazil), ANZICS-CORE, ICNARC (UK), NICST (Sri Lanka)
- Better collaboration between HICs and LMICs.