A Database-driven Decision Support System: Customized Mortality Prediction

David Íñiguez & Jaume Sánchez
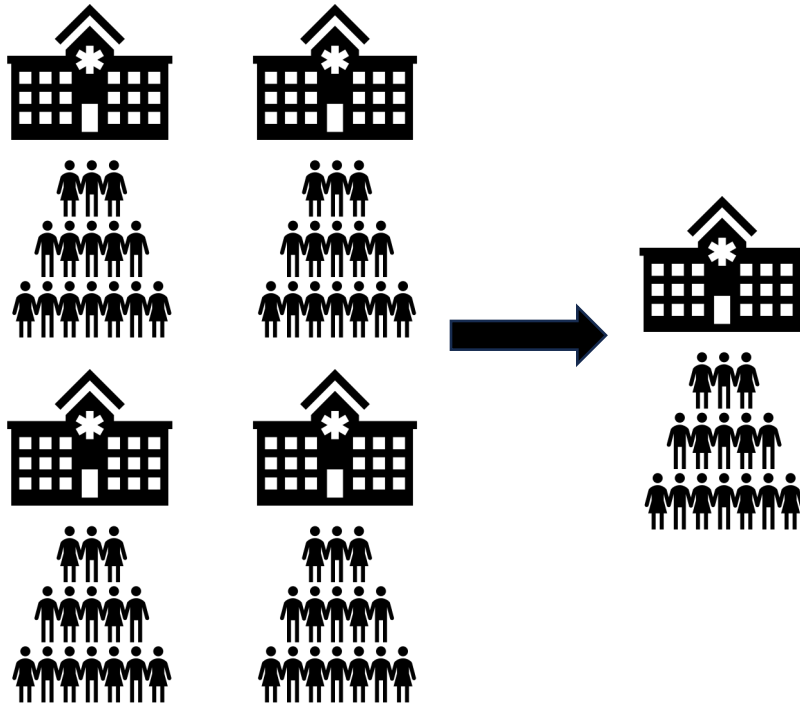
UNIVERSITAT DE BARCELONA

Table of contents

- Introduction and objective

- Data collection

- Methodology

- Validation results

- Discussion

- Conclusion

UNIVERSITAT DE BARCELONA

Introduction and objective

Mortality prediction

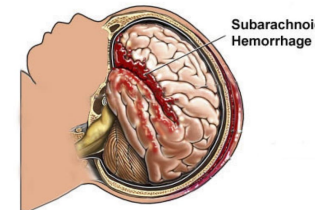**Scoring systems**   **VS**   **ML models**

- SAPS
- EuroSCORE

- Logistic regression
- Neural network
- Bayesian network

ICU patients

Acute kidney injury (AKI)   Subarachnoid hemorrhage (SAH)   Cardiac surgery

Subarachnoid Hemorrhage

Data

AKI & SAH patients

- Laboratory of Computational Physiology at MIT, MIMIC

- Demographic factors, SAPS, and physiologic variables measured during the first three days (AKI) in the ICU, and for the first day (SAH)

Cardiac surgery patients

- Dunedin Hospital, University of Otago, New Zealand

- Selected pre-defined variables entered by personnel with domain expertise: discrete pre-operative, intra-operative and post-operative variables.
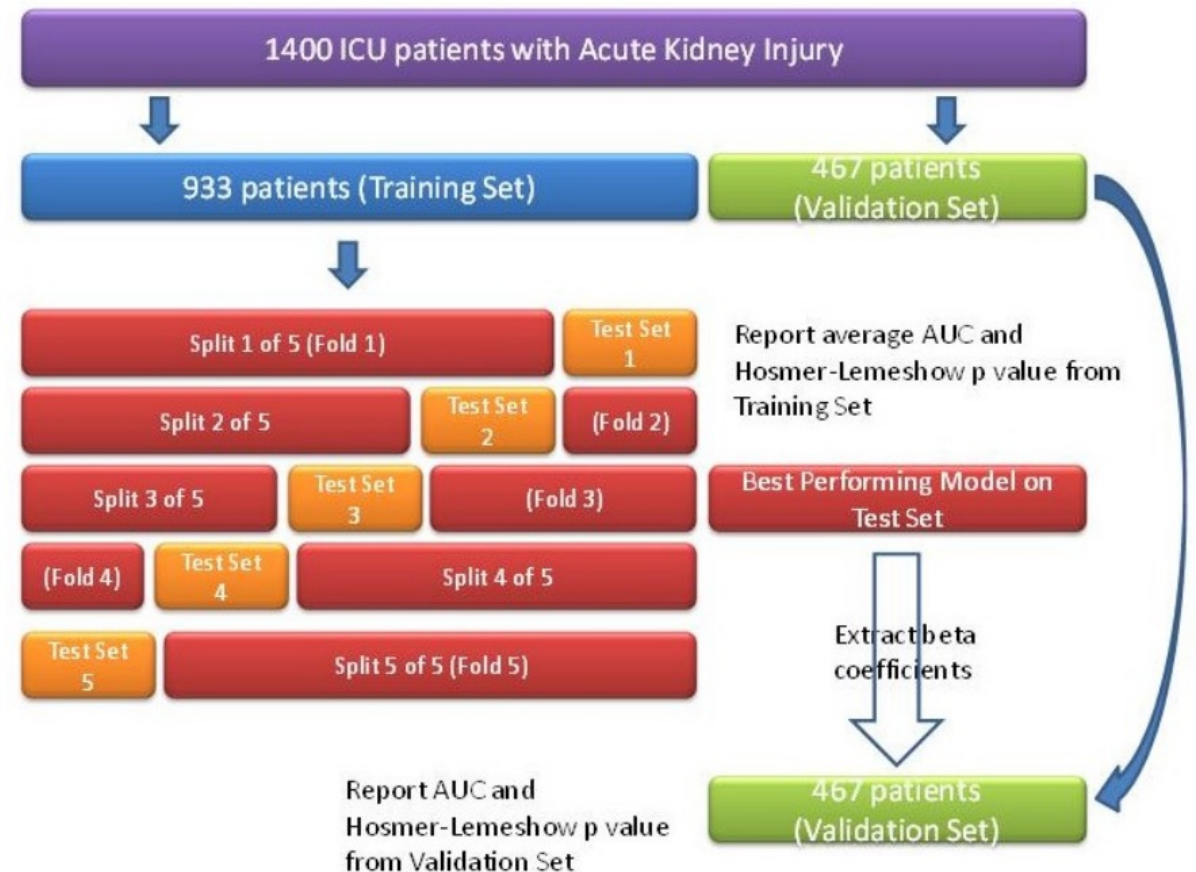
UNIVERSITAT DE BARCELONA

Methodology

Scoring systems based models

$$Predicted\ Death\ Rate = \frac{e^{Logit}}{1+e^{Logit}}$$

Logit is a function of the scoring system used

ML models

1400 ICU patients with Acute Kidney Injury

933 patients (Training Set)

467 patients (Validation Set)

Split 1 of 5 (Fold 1) — Test Set 1

Split 2 of 5 — Test Set 2 — (Fold 2)

Split 3 of 5 — Test Set 3 — (Fold 3)

(Fold 4) — Test Set 4 — Split 4 of 5

Test Set 5 — Split 5 of 5 (Fold 5)

Report average AUC and Hosmer-Lemeshow p value from Training Set

Best Performing Model on Test Set

Extract beta coefficients

467 patients (Validation Set)

Report AUC and Hosmer-Lemeshow p value from Validation Set

Validation results for AKI

	Accuracy	Mean Absolute Error	Area under the ROC Curve
Logistic Regression	72.9%	0.367	0.738
Bayesian Network	73.2%	0.306	0.761
Artificial Neural Network	81.9%	0.227	0.875

Table 2. Best fitted logistic regression model for ICU patients who developed AKI.

	Estimate	Standard Error	z value	Pr (>|z|)		Estimate	Standard Error	z value	Pr (>|z|)
Age	5.54e−03	2.30e−03	2.41	0.02	Standard deviation of the hematocrit (Day 2)	1.05e−01	5.05e−02	2.08	0.04
Maximum serum bilirubin (Day 2)	4.58e−02	1.46e−01	0.31	0.75	Minimum White Blood Cell count (Day 1)	−1.19e−03	1.98e−02	−0.06	0.95
Maximum serum bilirubin (Day 3)	1.66e−02	1.42e−01	0.12	0.91	Minimum White Blood Cell count (Day 2)	−7.07e−02	8.66e−02	−0.82	0.41
Minimum heart rate (Day 2)	3.64e−03	5.44e−03	0.67	0.50	Average White Blood Cell count (Day 2)	6.50e−02	8.59e−02	0.76	0.45
Average systolic blood pressure (Day 1)	−8.61e−03	5.71e−03	−1.51	0.13	Minimum White Blood Cell count (Day 3)	3.36e−02	2.28e−02	1.47	0.14
Minimum systolic blood pressure (Day 2)	−8.31e−04	6.25e−03	−0.13	0.89	Maximum BUN (Day 2)	−1.66e−02	8.29e−03	−2.00	0.05
Minimum systolic blood pressure (Day 3)	−2.18e−02	7.42e−03	−2.94	0.003	Maximum BUN (Day 3)	2.98e−02	8.39e−03	3.56	0.0004
Average systolic blood pressure (Day 3)	6.46e−03	7.71e−03	0.84	0.40	Glasgow coma score (Day 1)	−4.42e−02	1.71e−02	−2.59	0.01
Maximum respiratory rate (Day 3)	1.58e−02	1.14e−02	1.38	0.17	Maximum serum bicarbonate (Day 1)	6.20e−03	1.83e−02	0.34	0.73
Standard deviation of the hematocrit (Day 2)	1.05e−01	5.05e−02	2.08	0.04	Urine Output (Day 1)	−1.20e−04	8.44e−05	−1.43	0.15
					Urine Output (Day 2)	−6.60e−05	6.75e−05	−0.98	0.33
					Urine Output (Day 3)	−1.10e−04	7.44e−05	−1.48	0.14

Note: Hosmer-Lemeshow statistic = 6.472 ($p = 0.594$).

- SAPS RESULTS:
  - AUC = 0.6419
  - Hosmer-Lemeshow decile = 2941.64 (p=0)

UNIVERSITAT DE BARCELONA

Validation results for SAH

	Accuracy	Mean Absolute Error	Area under the ROC Curve
Logistic Regression	89.0%	0.158	0.945
Bayesian Network	87.7%	0.127	0.958
Artificial Neural Network	83.6%	0.168	0.868

Table 4. Best fitted logistic regression model for ICU patients who presented with SAH.

	Estimate	Standard Error	z value	Pr (>|z|)
Age	0.05	0.02	2.64	0.008
Average serum glucose	0.02	0.01	2.50	0.01
Maximum White Blood Cell count	0.01	0.05	0.10	0.92
Standard deviation of the serum glucose	0.13	0.32	0.41	0.68
Average prothrombin time INR	3.20	1.56	2.05	0.04
Minimum Glasgow coma score	−0.01	0.17	−0.06	0.95
Maximum Glasgow coma score	0.24	0.22	1.12	0.26

	Estimate	Standard Error	z value	Pr (>|z|)
Average Glasgow coma score	−0.60	0.33	−1.80	0.07
Minimum systolic blood pressure	−0.02	0.02	−0.94	0.34
Minimum serum sodium	0.03	0.32	0.10	0.92
Average serum sodium	−0.03	0.32	−0.10	0.92
Standard deviation of the serum sodium	0.02	0.44	0.04	0.97

Note: Hosmer-Lemeshow statistic = 7.196 ($p = 0.516$).

- SAPS RESULTS:
  - AUC = 0.84
  - Hosmer-Lemeshow decile = 887.95 ($p<0.001$)

Validation results for open heart surgery

	Accuracy	Mean Absolute Error	Area under the ROC Curve
Logistic Regression	80.0%	0.201	0.854
Bayesian Network	96.4%	0.129	0.931
Artificial Neural Network	96.4%	0.045	0.941

Table 6. Best fitted logistic regression model for elderly patients who underwent cardiac surgery.

	Estimate	Standard Error	z value	Pr (>|z|)
Ejection fraction	1.11	1.01	1.10	0.27
Use of an intra-aortic balloon pump	1.61	1.67	0.97	0.33
Chest Reopening	3.14	1.38	2.28	0.02
Development of atrial fibrillation	18.68	2.46	−0.01	0.99
Development of a post-operative infection	0.77	1.19	−0.65	0.52

Note: Hosmer-Lemeshow statistic = 5.671 ($p = 0.684$).

- SAPS RESULTS:
  - AUC = 0.648
  - Hosmer-Lemeshow decile = -181.8 (p=1.0)

UNIVERSITAT DE BARCELONA

Results discussion

	AKI	SAH	Cardiac Surgery
Scoring system used	SAPS	SAPS	EuroSCORE
Previous studies	Poor performance. Very dissimilar mortality predictions.	Few validation studies, although there exist numerous systems	Mortality in elderly patients is overestimated
Is SAPS/EuroSCORE based model calibrated?	No	No	Yes
Best ML model	ANN (calibration improved)	Logistic and Bayesian	Bayesian and ANN
Is ML model calibrated?	Yes	No	Yes

Conclusion

- Severity scoring systems perform well when dealing with ICU populations, but lack accuracy/calibration when predicting individual cases.

- Using different machine Learning algorithms for MIMIC II subsets and local databases outperformed the severity scoring systems in all cases, and even improved the calibration in the AKI subset.

UNIVERSITAT DE BARCELONA

Data Science for Health