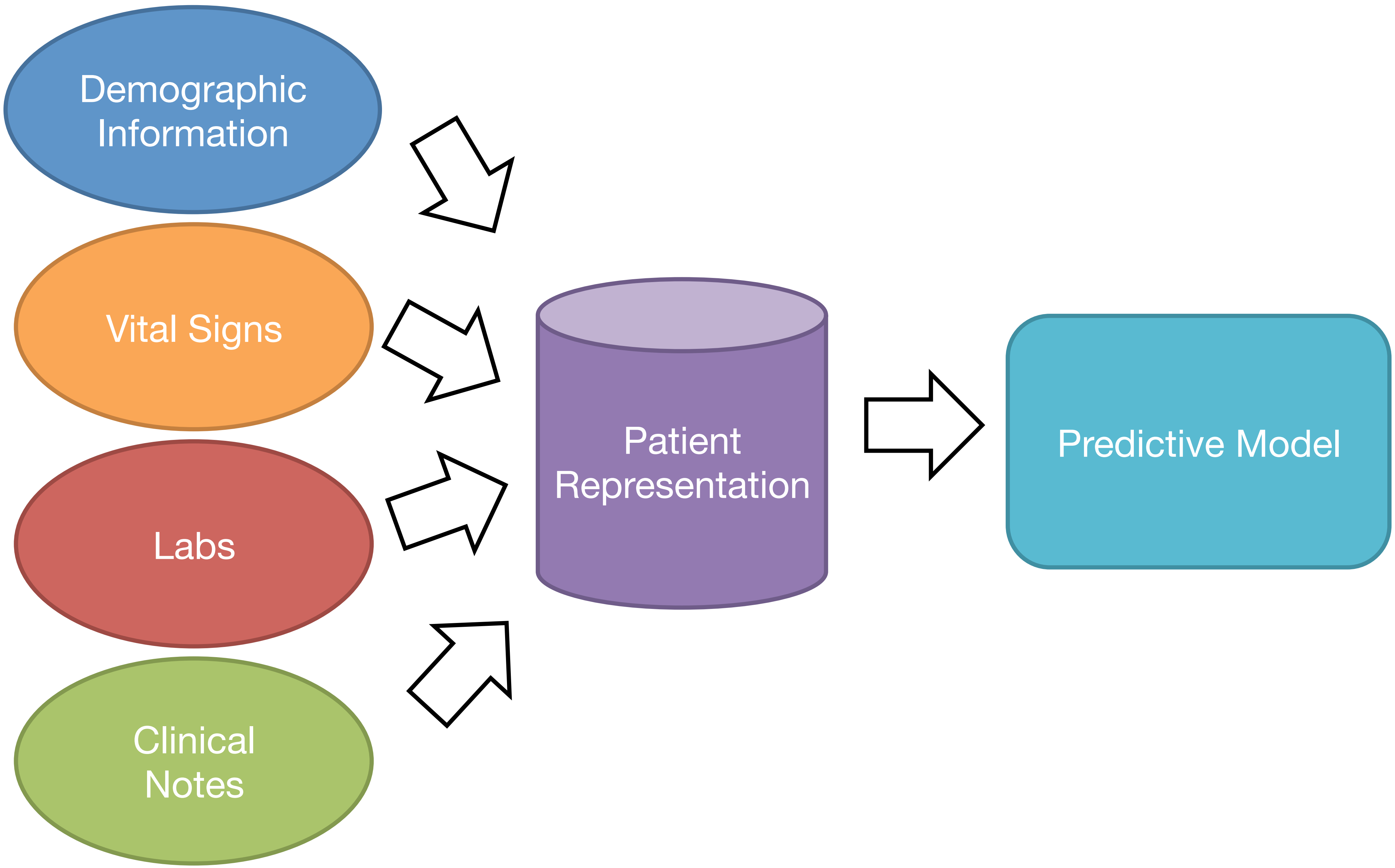
SEPTIC SHOCK PREDICTION FOR PATIENTS WITH MISSING DATA

Joyce C Ho, Cheng Lee, Joydeep Ghosh
University of Texas at Austin

What is Sepsis and Septic Shock?

- **Sepsis** is a systemic inflammatory response to infection
- 11th leading cause of death in 2010
- Estimated $14.6 billion spent on sepsis in 2008
- **Septic shock** (sepsis-induced hypotension) has a mortality rate of 45.7%

In-Hospital Detection



Missing Data Problem

- Clinical studies must deal with large amounts of missing data
- Measurements are noisy and irregularly sampled
- Highly accurate measurements require invasive techniques (may not be medically necessary)

Typical Approach

- Ignore subjects with missing observations
- Ignore features without complete data
- Result: Highly curated datasets with limited features and small samples

OUR SEPTIC SHOCK MODEL

Problem: Given a patient has sepsis, can we predict complications at least one hour prior to onset of septic shock?

- Generalization to patients with partially missing observations
- Simple and accessible approaches
- Focus on commonly observed, non-invasive measurements

Clinical Features

- Summary statistics (last measurement, min, mean, and max) in 8 hour window
  - Cardiac: non-invasive blood pressure, heart rate, pulse pressure
  - Other: respiratory rate, SpO_2, temperature
- Last measurement only (less observations)
  - White blood cell count
  - Index scores: SOFA, SAPS-I, Shock index

Imputation Approaches

- Mean / median imputation
- Matrix factorization techniques
  - Singular value based imputation (SVD)
  - Probabilistic principal component analysis (PPCA)
- K-nearest neighbors (KNN)

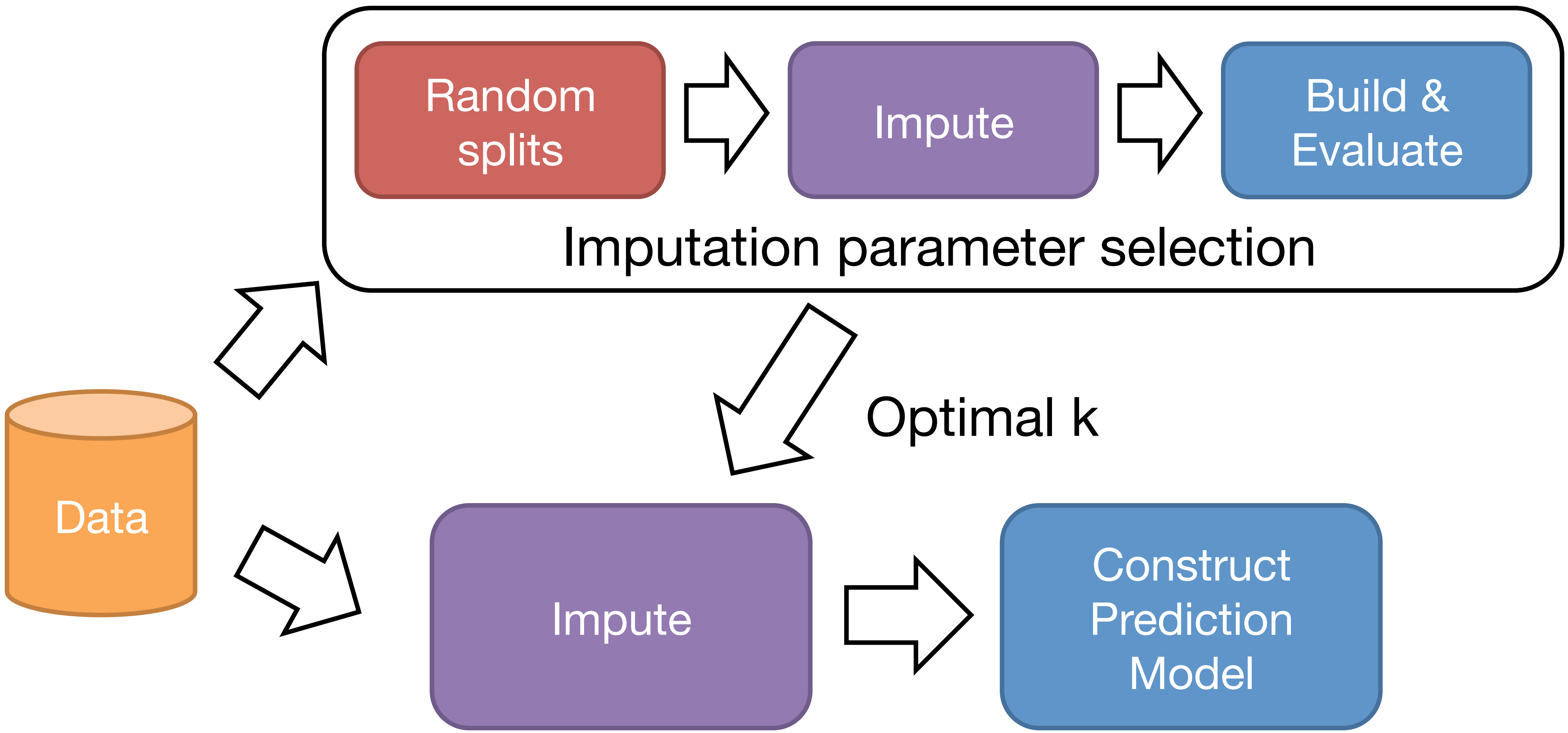
Imputation Selection Criteria

- Matrix factorization and neighborhood techniques have parameter to control resolution or locality of imputation

- Evaluation metric typically involves randomly removing observations and comparing fit using root mean square error (RMSE) or mean absolute error (MAE)

- RMSE / MAE may not necessarily translate to improved predictive performance

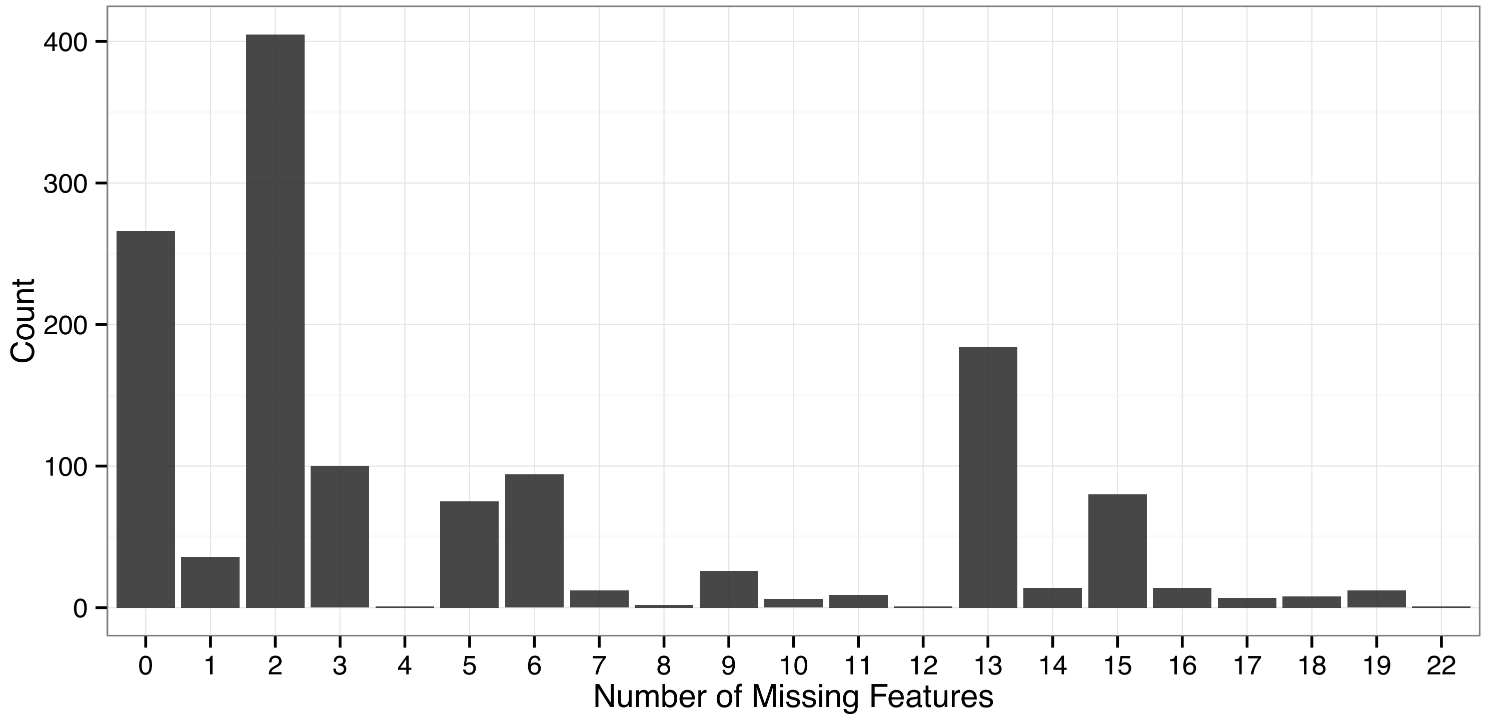
Performance-Oriented Imputation (POI)



MIMIC-II Database

- Extensive, publicly available ICU data resource
- Data between 2001 and 2007 from Boston's Beth Israel Deaconess Medical Center ICUs
- Over 40,000 ICU stays from 30,000+ patients
- Clinical records with physiological measures, medication records, laboratory tests, free-form text notes, etc.

IMPORTANCE OF IMPUTATION



Less than 22% of the 1,353 patients have complete data

Feature	30 mins	60 mins
Respiratory rate	0.67%	0.68%
Temperature	1.70%	2.05%
White blood cells	15.30%	14.69%
Blood pressure	23.28%	23.44%

Non-invasive BP is not always available

Differences in Population

	Missing patients		Complete only		
Time	**Sepsis (only)**	**Shock**	**Sepsis (only)**	**Shock**	**P-value**
30 mins	749	79	199	110	4.56E-26
60 mins	723	79	196	106	6.99E-24
90 mins	705	79	196	103	4.63E-22
120 mins	685	74	193	103	7.06E-23

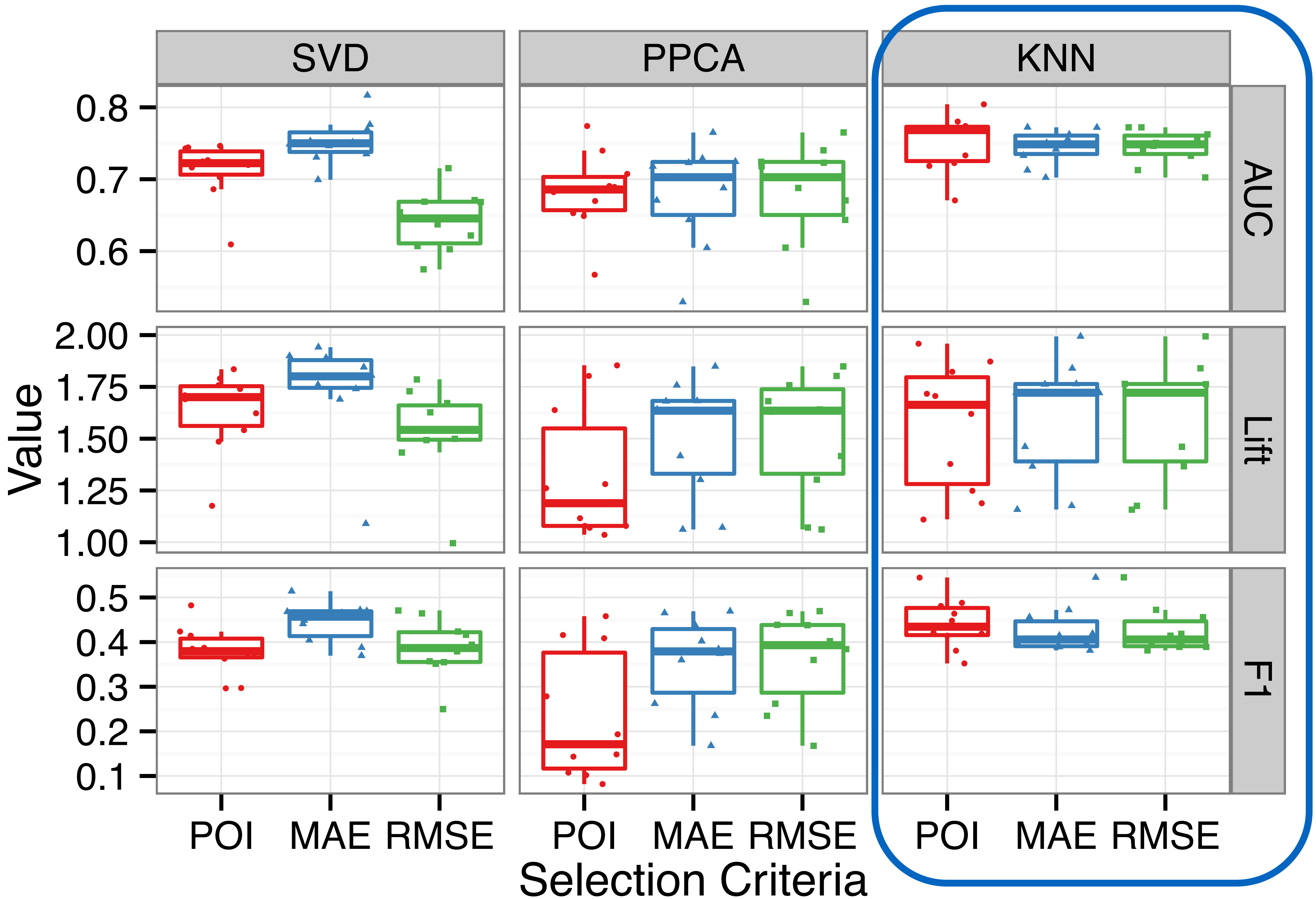
Statistically significantly higher ratio of shock patients if you ignore patients with missing data

Predictive Power of Mean Imputed Model

Train Data	Test Data	30 minutes before (AUC)	60 minutes before (AUC)
Complete	Complete	0.796±0.065	0.777±0.050
Complete	Imputed	0.815±0.033	0.800±0.053
Imputed	Imputed	0.834±0.025	0.829±0.030
Imputed	Complete	0.839±0.044	0.828±0.047

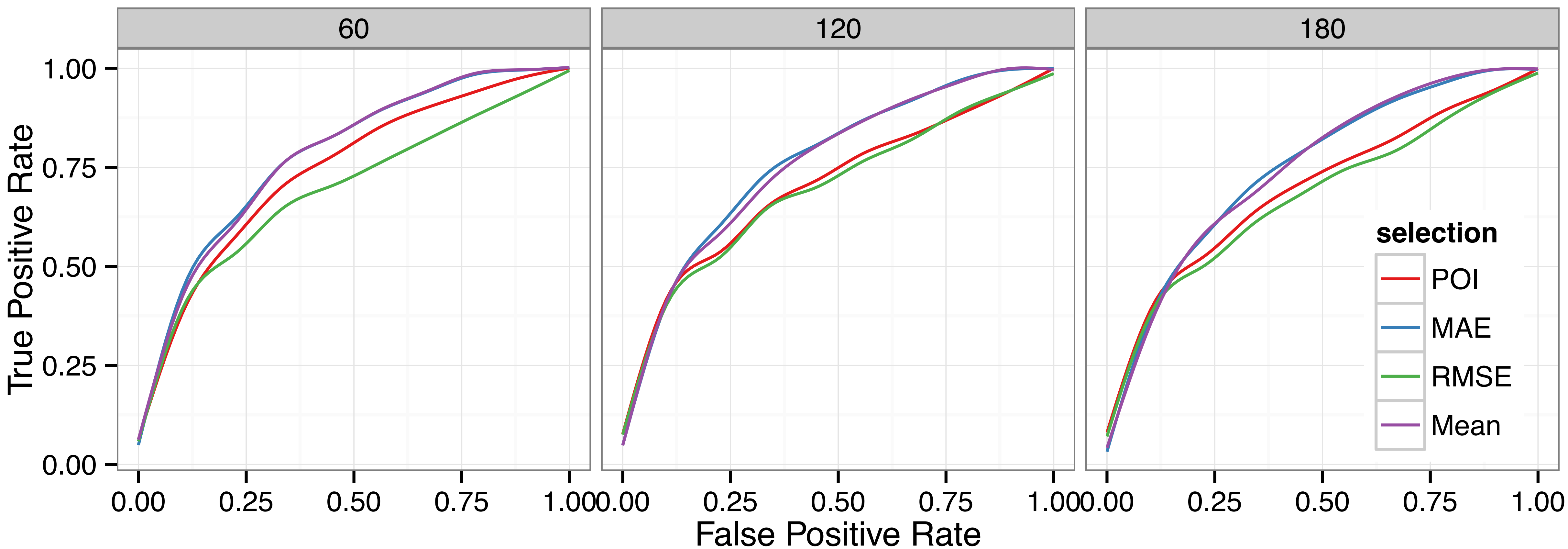
Model generalizes to broader population with slightly better predictive performance

Comparison of Selection Criteria (SVM)



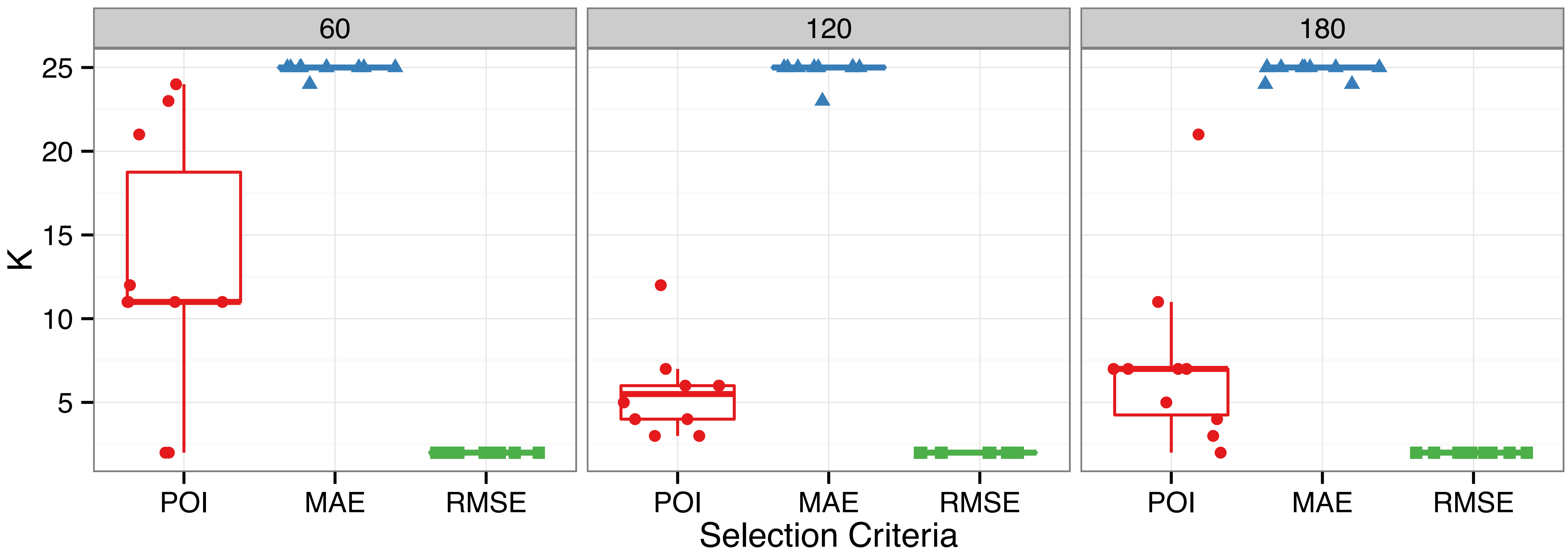
POI is generally better for AUC + F-measure

Comparing Imputation Approaches (SVD + LogR)



POI outperforms RMSE, but mean and MAE are generally the best

Comparing Imputation Approaches (SVD + LogR)



RMSE favors the simplest model (k=1),
MAE favors most complex (k=25),
POI lies in between the two

Comparing Imputation Approaches (Feature Rank)

Feature	Mean	AUC	F1	RMSE
Systolic BP	1.50	1.70	1.70	2.40
SpO2	2.22	3.00	3.22	2.56
Shock Index	4.40	4.40	4.60	3.30
Temp	5.00	5.00	7.50	
Diastolic BP	11.00	8.00	8.25	5.00

Selection criteria influences feature ranking within the same imputation method

CONCLUSION

- Generalizes to all ICU patients
- Focuses on commonly observed, non-invasive clinical measurements
- Uses simple and accessible approaches for missing data problem

References

Joyce C Ho, Cheng H Lee, and Joydeep Ghosh. Imputation-enhanced prediction of septic shock in ICU patients. In *2012 ACM SIGKDD Workshop on Health Informatics (HI-KDD)*, 2012.

Joyce C Ho, Cheng H Lee, and Joydeep Ghosh. Septic shock prediction for patients with missing data. *ACM Transactions on Management Information Systems (TMIS)*, 5(1):1:1–1:15, 2014.