

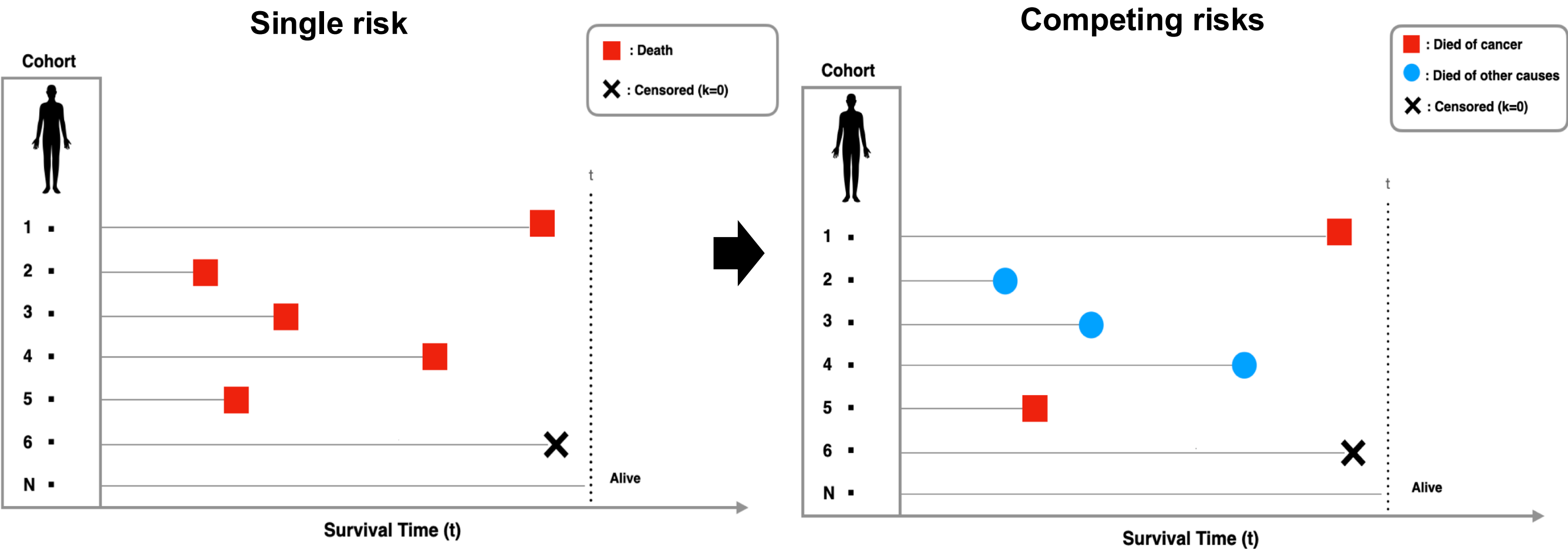
Advances in Survival Analysis with Competing risks: Benchmarking and Application to Recurrence Prediction in Colorectal Cancer

Begoña B. Sierra¹, Ariane Marandon-Carlhian², Colin McLean¹, Peter Hall¹, Joanne Edwards³, Catalina A. Vallejos^{4,5}

¹Edinburgh Cancer Centre, University of Edinburgh, ²Department of Statistical Science University College London, ³Glasgow Cancer Centre, University of Glasgow, ⁴MRC Human Genetics Unit University of Edinburgh, ⁵Health Data Research UK

Motivation

- Cancer outcomes are inherently **competing** such as **death from cancer recurrence** vs **death from other causes (e.g., cardiovascular disease)**.
- Problem:** Standard survival analysis often ignore this structure (e.g., using composite outcomes) or competing events are censored, which **biases event incidences** and mislead clinical interpretation.

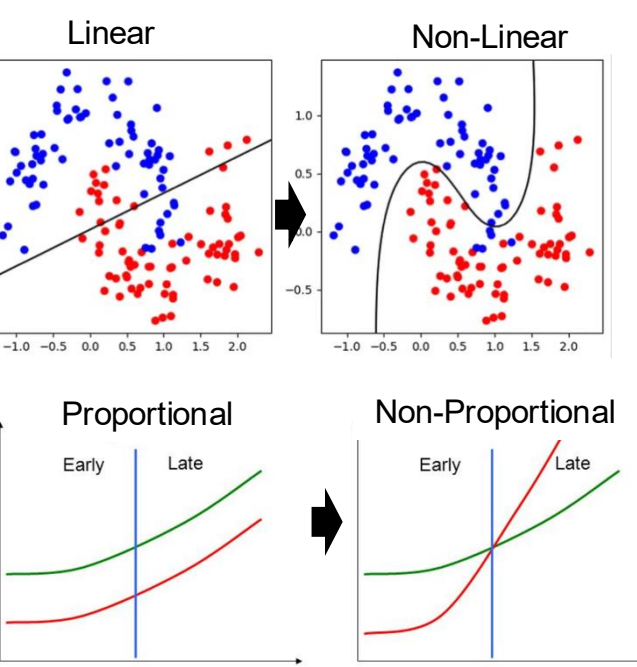


- Challenge:** Rapid growth in statistical and deep learning competing risk models has outpaced rigorous, comparable evaluation.
- Solution:** **Transparent and neutral benchmark** with a broad set of performance evaluation metrics.
- Application:** Benchmark is applied to a **Scottish colorectal cancer cohort** to identify which methods are reliable in practice.

Advances in survival models with competing risks

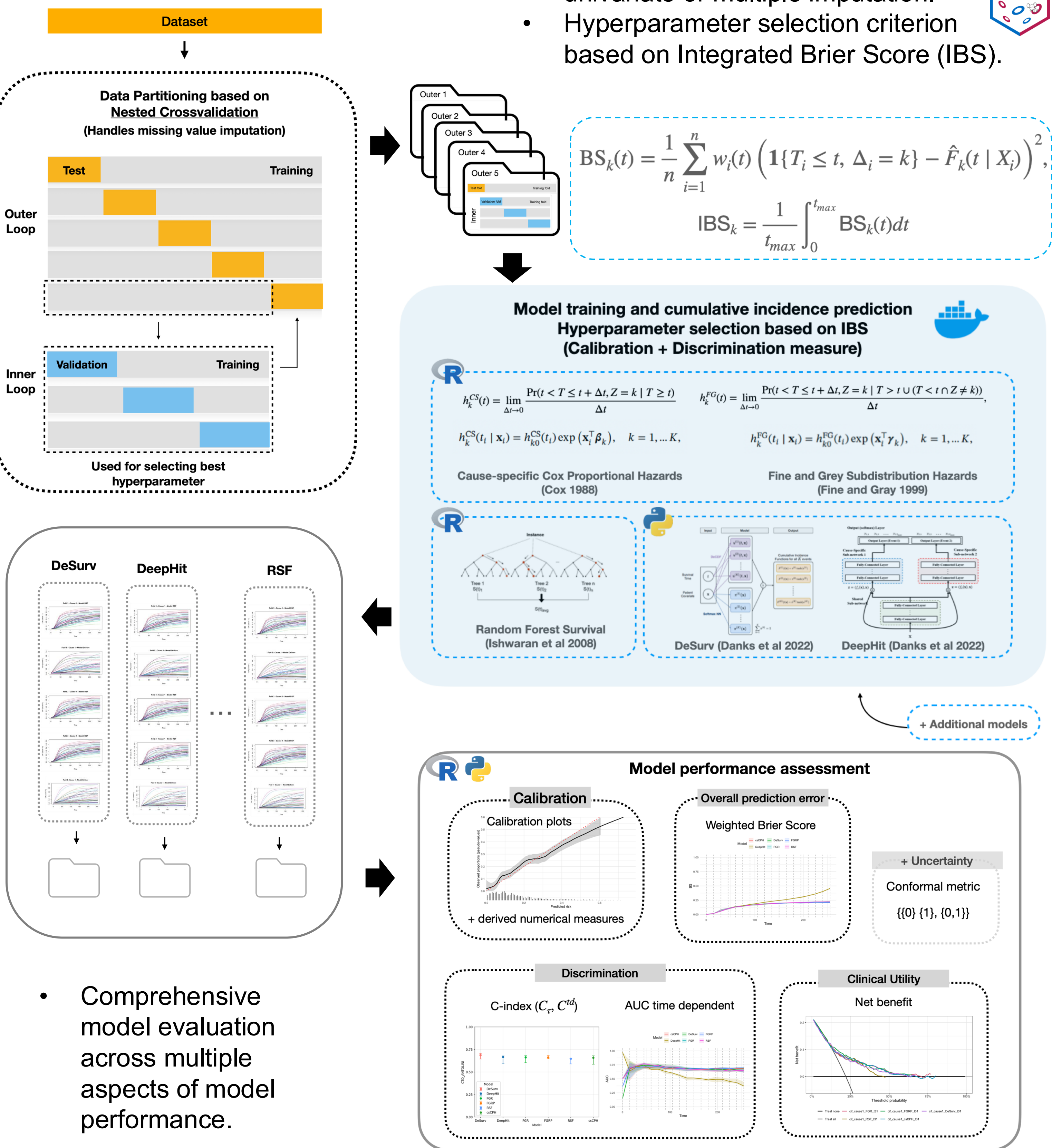
Survival models with competing risks used in the benchmark					
Model	Type	High dimensions (p)	Non-linear effects	Missing data	Covariate interpretability
Cause-specific hazard approaches					
Cause-specific Cox PH (csCPH)	Semiparametric	×	×	×	✓
Subdistribution / CIF-based approaches					
Fine & Gray regression (FGR)	Semiparametric	×	×	×	✓
Penalized subdistribution hazard (FGRP)	Semiparametric	✓	×	×	✓
Tree-based					
Random survival forests (RSF)	Nonparametric	✓	✓	✓	×
Deep learning models					
DeSurv	Nonparametric	×	✓	×	×
DeepHit	Nonparametric (discrete-time)	✓	✓	×	×

Moving towards **non-linear, non-parametric, no strong assumptions** (e.g. proportional hazards)

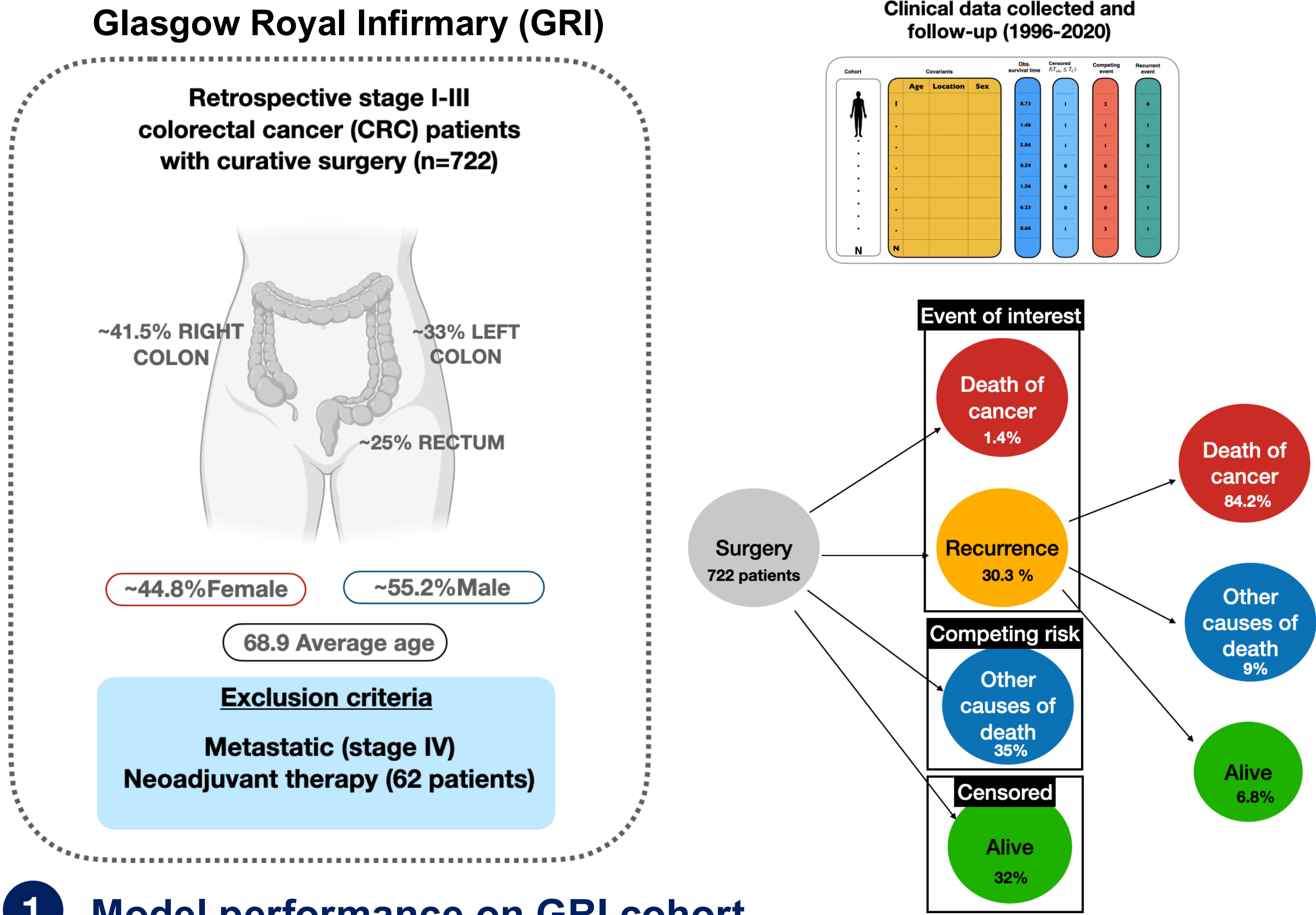


Benchmark overview

- Nested cross-validation set-up
- Easy to add new survival models with competing risks
- Decoupled python and R environments – Docker containers.
- Detects and imputes missing values, univariate or multiple imputation.
- Hyperparameter selection criterion based on Integrated Brier Score (IBS).

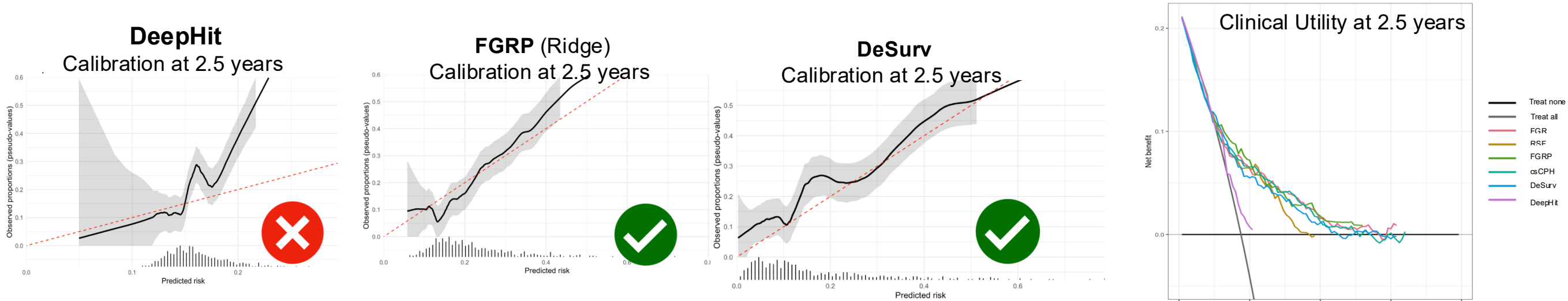


Real-world Application

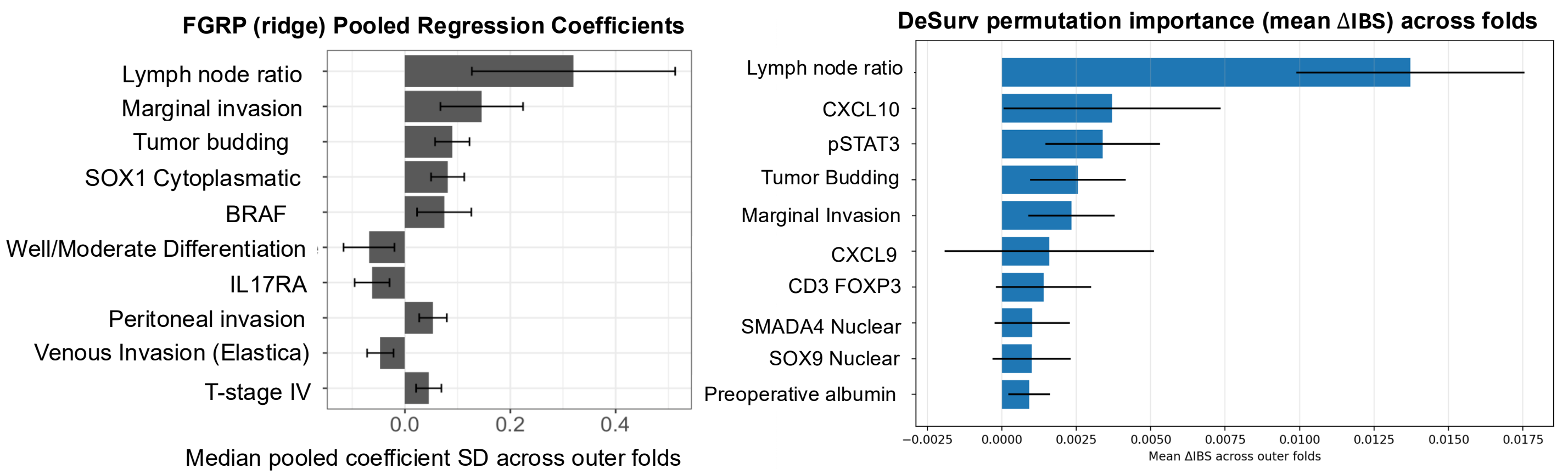


1 Model performance on GRI cohort

Integrated Brier Score (IBS) (lower the better)				C-index (Pec, RMLT input, IPCW) (higher the better)			
Model	Mean	Median	95% CI	Model	Mean	Median	95% CI
csCPH	0.1716	0.1702	(0.1588, 0.1839)	csCPH	0.6585	0.6492	(0.6381, 0.6883)
DeepHit	0.1959	0.1959	(0.1920, 0.1998)	DeepHit	0.6235	0.6441	(0.5475, 0.6703)
DeSurv	0.1688	0.1624	(0.1580, 0.1868)	DeSurv	0.6691	0.6707	(0.6199, 0.7053)
FGR	0.1709	0.1706	(0.1546, 0.1839)	FGR	0.6596	0.6589	(0.6414, 0.6894)
FGRP	0.1635	0.1676	(0.1538, 0.1711)	FGRP	0.6785	0.6629	(0.6523, 0.7224)
RSF	0.1704	0.1709	(0.1661, 0.1749)	RSF	0.6674	0.6808	(0.6223, 0.6925)

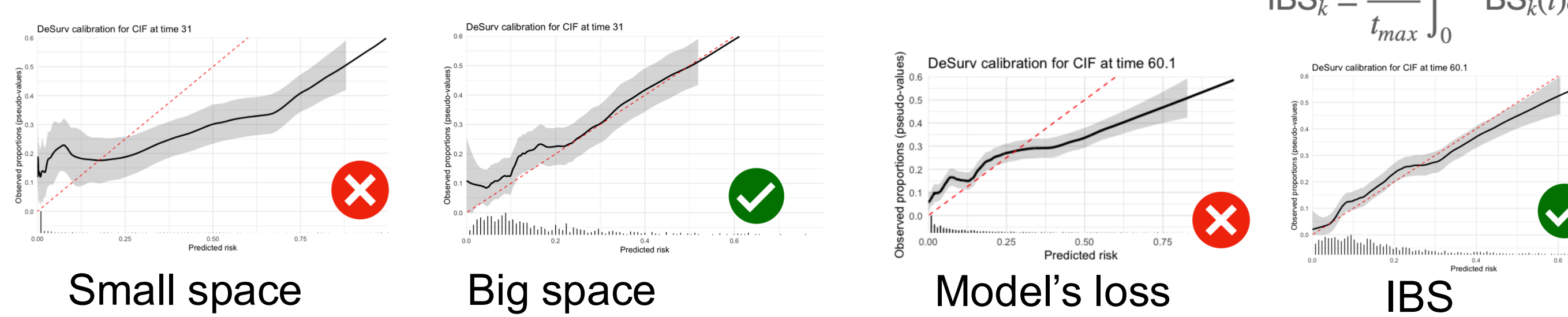


2 Predictors of death from cancer recurrence/CRC (TOP 10)



Main takeaways

- Clinical risk prediction models need to be evaluated based on different aspects of model performance.
- AI models need to be accurately compared to commonly used statistical approaches.
- Benchmarks offer a neutral and transparent model comparison.
- During cross-validation, Hyperparameters space is highly important.
- Hyperparameter selection criterion based on calibration and discrimination (IBS), results in better performance.



Future Work

- Benchmark**
 - Addition of alternative models
 - Uncertainty estimation with conformal
 - Private → Public GitHub
- Glasgow Royal Infirmary Cohort**
 - Agreement between models on risk predictors of recurrence
 - Explore histopathology as another source of potential risk factors

Acknowledgements

Translational Cancer Pathology, JE research group, for support on clinical data analysis. CRUK supporting BS PhD scholarship. CV and PH research group for continuous methodological support. AMG for new contributions on model assessment though conformal framework.