Developing Trustworthy Artificial Intelligence for Organ Transplantation

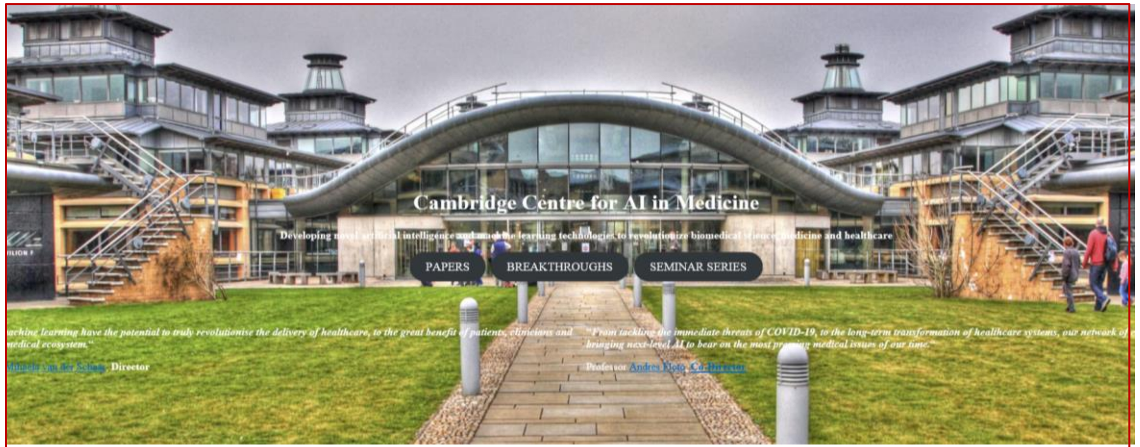
mpj1001@cam.ac.uk

Dr Alexander Gimson

CCAIM

UNIVERSITY OF CAMBRIDGE

Acknowledgements



Professor Mihaela van der Schaar
Dr Ahmed Alaa
Dr James Jordan
Jeroen Berrevoets
Zhaozhi Qian
Dr Fergus Imrie
Alihan Huyuk
Bogdan Cerebrere

CCAIM

NHS Blood & Transplant
Statistics Bristol UK

Dr Kerrie Barber
Dr Alisa Allen
Dr Rhiannon Taylor

Members of the UK Liver
Selection and Allocation
Working Parties

Developing Trustworthy Artificial Intelligence and Organ Transplantation

SUMMARY

Components of clinical trustworthiness

Complexities of liver transplantation

ML applications for organ allocation

Interpretability of ML allocation decisions

Interpreting clinical decision making

Some caveats

Avoiding an AI winter

Majority of studies remain in testing environment *Kim et al Korean J Radiol 2019; 20; 405-410*

Some have not met their clinical aims *Wikinson J et al. Time to reality check the promise of machine learning powered precision medicine Lancer Digit Health 2020; 2; e677-80*

Improve trustworthiness by regulation

Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback. Food and Drug Administration, 2019.14 FaD A. *Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)*. Action Plan: Food and Drug Administration, 2021.

Commission E. *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain Union legislative acts*. Brussels: European Commission, 2021.

CCAIM

Components of trustworthiness for ML platforms

- Improvement on previous methodologies
- Interpretable results
- Clinically relevant problem
- Open process of development – formulation to implementation
- Public/Patient involvement
- Multiple simulation processes – synthetic data
- Transferable across jurisdictions
- Regulatory authorities
- Laws of Tort

CCAIM

Developing, implementing and governing artificial intelligence in medicine:

Preparation prior to AI development

Define clinical problem *Wiens et al Nat Med 2019; 25; 1627*

Evaluate deficiencies in previous models

Consider data biases *Wolff et al PROBAST Ann Int Med 2019; 170; 51-8*

Data privacy

AI model development

Applicable regulatory requirements- *FDA; harmonised rules on AI (EU)*

Prepare data

Train and validate

Evaluate, report results – *TRIPOD-ML Collins et al Lancet 2019; 393; 1577-9*

van de Sande D, Van Genderen ME, Smit JM, et al. Developing, implementing and governing artificial intelligence in medicine: a step-by-step approach to prevent an artificial intelligence winter. BMJ Health Care Inform 2022;29:e100495. doi:10.1136/bmjhci-2021-100495

CCAIM

Developing, implementing and governing artificial intelligence in medicine:

Assess performance and reliability

Externally validate – *Futoma et al Lancet Digit Health 2020; 2; e489-92*

Clinical papers – *DECIDE-AI New reporting guideline Nat Med 2021; 27; 186-187*

Clinical testing

Design an clinical study –*CONSORT-AI extension. Lancet Digit Health 2020; 2020; e537-48*

Implementation

Legal/regulatory – *Muehlematter et al Lancet digit Health 2021; 3; e195-203*

Model outcome governance – FDA, MDFR

van de Sande D, Van Genderen ME, Smit JM, et al. Developing, implementing and governing artificial intelligence in medicine: a step-by-step approach to prevent an artificial intelligence winter. BMJ Health Care Inform 2022;29:e100495. doi:10.1136/bmjhci-2021-100495

CCAIM

Why is this an important area?

- Supply and Demand

  Demand for transplantation increases

  Limited increase in supply of donor organs

  Mortality waiting for a liver transplant - 5% (UK) - 20% (US)

  Quality of organ donors deteriorating

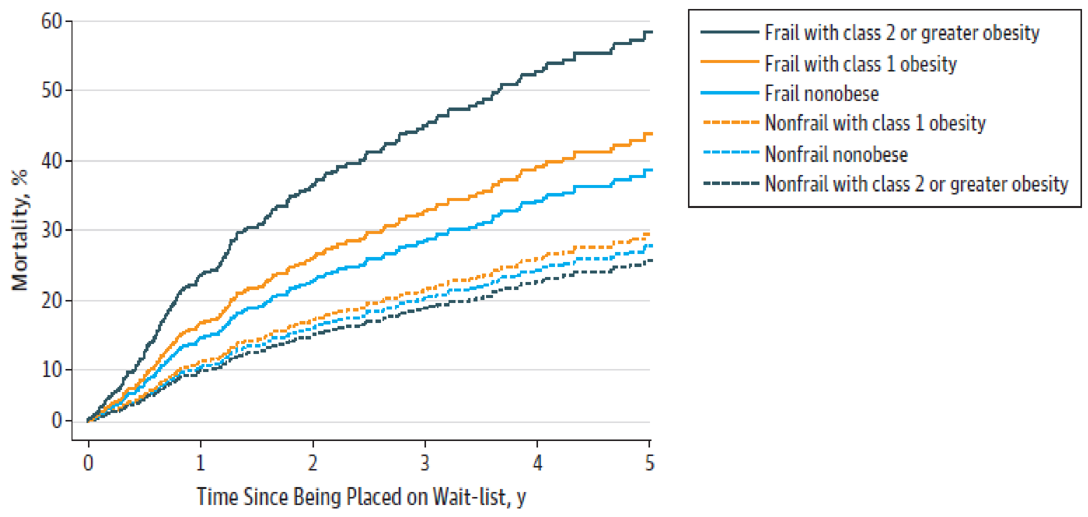
  - older, obese, ‘marginal’ donors

- A paradigm for scarce healthcare resource

Why is this a complex and interesting area?

- High quality databases
- Multi-dimensional donor and recipient space
  - Up to 17 donor/recipient factors impact outcome
- Non-linear interactions
  - Na, K, urea/creatinine, BMI
- Counterfactuals
  - impact of not receiving a transplant
- Assignment bias
- Informative censoring

Figure 2. Cumulative Incidence of Wait-list Mortality by Frailty Status in Nonobese Candidates, Candidates With Class 1 Obesity, and Candidates With Class 2 or Greater Obesity



CCAIM

Liver Transplantation – some basics

Only solution for end-stage chronic liver disease

Multiple causes for end stage liver disease

3 year survival without a transplant - 5%

Good outcome – 94% survival at one year, 75% at 5 years

Highly technical, costly intervention

UK - 776 (2021); US - 8372 (2019)

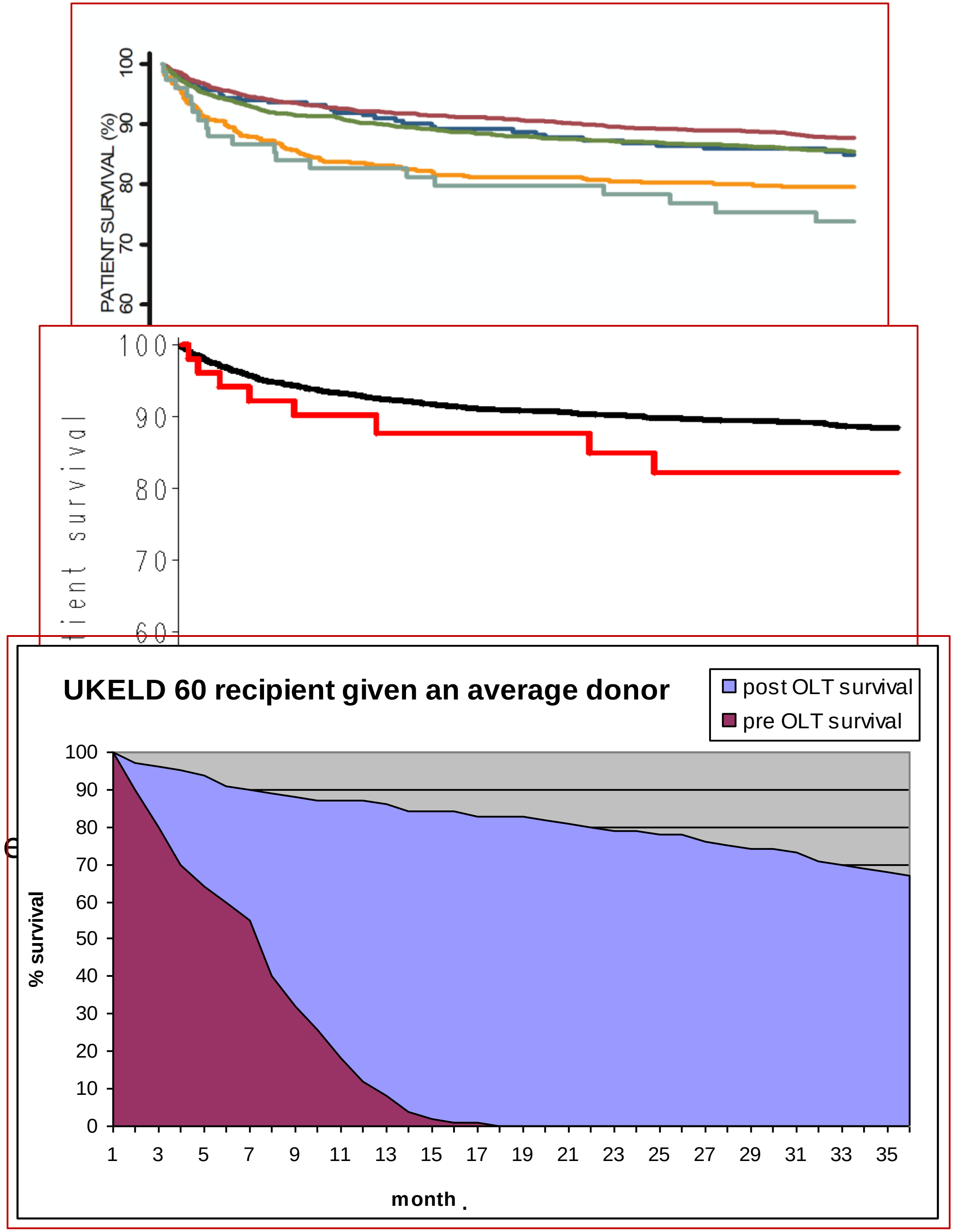
Both Recipient Disease Severity and Donor Quality and impact outcome

Allocation principles varies

Need - sickest patient first; US

Utility – best outcome

Benefit - net life years gained; UK March 2018



CCAIM

Organ transplantation and machine learning

1. Wait list entry criteria
2. Optimal donor organ allocation
3. Clinical variation in offer acceptance rates – quantitative epistemology
4. Predicting graft failure rates
5. Individualised immunosuppression regimens
6. Temporal phenotyping donor-recipient pairs
7. Time dependant monitoring policies

CCAIM

Briceño J, Cruz-Ramírez M, Prieto M, Navasa M, Ortiz de Urbina J, Orti R, et al. Use of artificial intelligence **(ANN)** as an innovative donor-recipient matching model for liver transplantation: results from a multicenter Spanish study. J Hepatol 2014;61:1020-1028.

Cruz-Ramírez M, Hervás-Martínez C, Fernández JC, Briceño J, de la Mata M. Predicting patient survival after liver transplantation using evolutionary multi-objective **artificial neural networks**. Artif Intell Med 2013;58:37-49.

Haydon GH, Hiltunen Y, Lucey MR, Collett D, Gunson B, Murphy N, et al. **Self-organizing maps** can determine outcome and match recipients and donors at orthotopic liver transplantation. Transplantation 2005;79:213-218.

Pérez-Ortiz M, Gutiérrez PA, Ayllón-Terán MD, Heaton N, Ciria R, Briceño J, Hervás-Martínez C. **Synthetic semi-supervised learning** in imbalanced domains: constructing a model for donor-recipient matching in liver transplantation. Knowledge-Based Syst 2017;2017:75-87.

Yoon J, Zame WR, Banerjee A, Cadeiras M, Alaa AM, van der Schaar M. Personalized survival predictions via **Trees of Predictors**: An application to cardiac transplantation. PLoS One. 2018 Mar 28;13(3):e0194985.

Organ allocation; principles and considerations

Outcome without a donor

Outcome with a specific donor

Time till another “better / optimal” donor appears

Mortality waiting for the “better / optimal” donor

Impact of any deterioration in clinical status whilst waiting

Impact of new potential recipients on the transplant list

Interpretable results

Need – sickest patient first

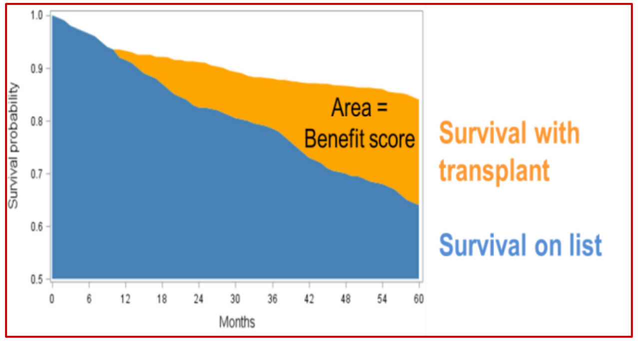
Risks increasing post transplant mortality

Utility – best match for outcome

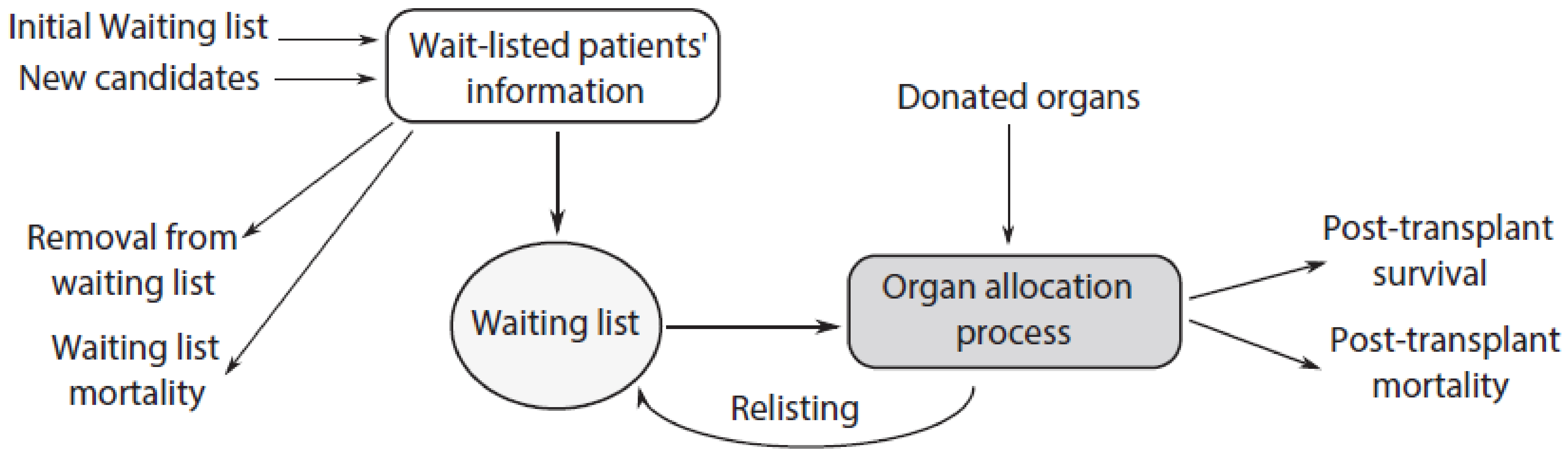
Risks increasing pre-transplant mortality

Benefit – incremental gain in survival

Net life years gained
Population life years
Complex



CCAIM


Initial Waiting list
New candidates
Wait-listed patients' information
Donated organs
Removal from waiting list
Waiting list mortality
Waiting list
Organ allocation process
Post-transplant survival
Post-transplant mortality
Relisting

Population life years

Transplantation judged from point of registration (minimum entry criteria)

Death or removal from transplant waiting list
Death after transplantation
Removal from post transplant list due to graft failure
Survival to end time point – 5 years

Societal aim of organ transplantation is to maximise population life years on an intention to treat basis

CCAIM

OrganITE: Optimal transplant donor organ offering using an individual treatment effect

Jeroen Berrevoets
University of Cambridge,
Vrije Universiteit Brussel
jeroen.berrevoets@maths.cam.ac.uk

James Jordon
University of Oxford
james.jordon@wolfson.ox.ac.uk

Ioana Bica
University of Oxford,
The Alan Turing Institute
ioana.bica@eng.ox.ac.uk

Alexander Gimson
Cambridge University Hospitals
alexander.gimson@nhs.net

Mihaela van der Schaar
University of Cambridge,
University of California, Los Angeles,
The Alan Turing Institute
mv472@cam.ac.uk

Simulation of outcomes between real time allocation compared to allocation by other methodologies

A balanced score composed of;

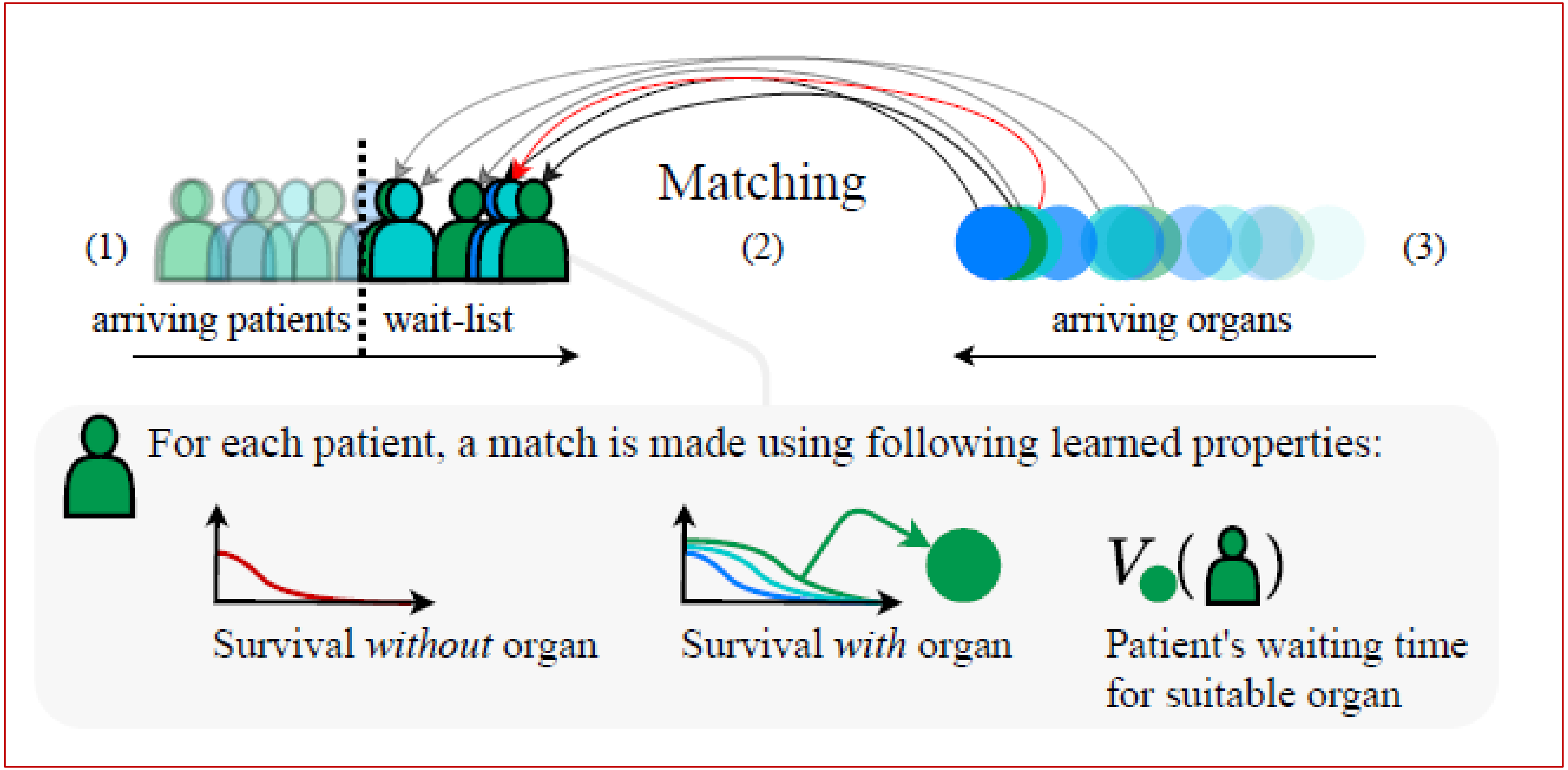
- Transplant benefit using Individual Treatment Effects

Bica, I., Alaa, A. M., Lambert, C., & Van Der Schaar, M. (2021). From real-world patient data to individualized treatment effects using machine learning: current and future methods to address underlying challenges. Clinical Pharmacology & Therapeutics.

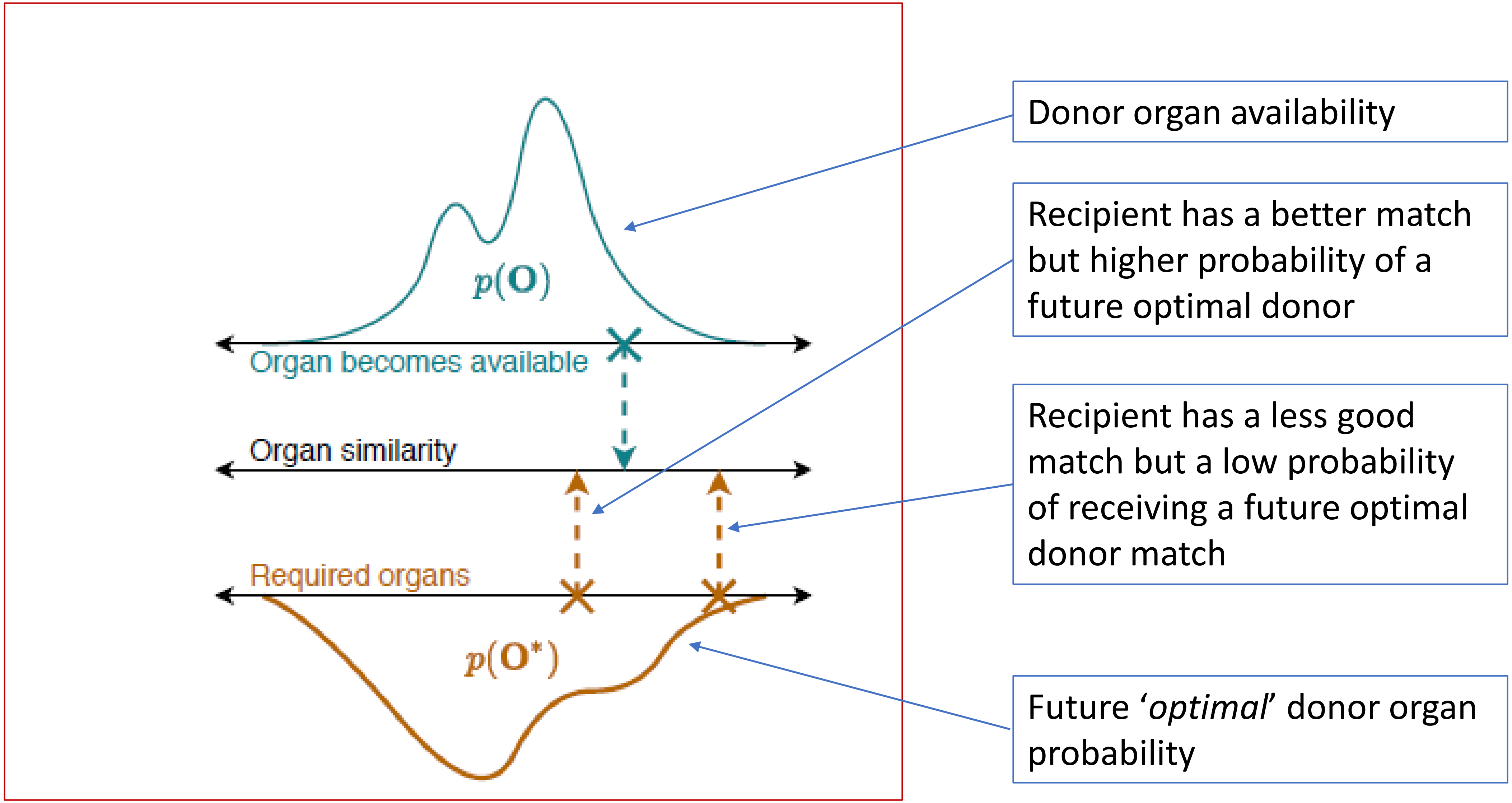
- Estimation of 'optimal' donor for each case on the list

- Future probability of the optimal donor arriving

https://proceedings.neurips.cc/paper/2020/hash/e7c573c14a09b84f6b7782ce3965f335-Abstract.html



UK Transplant Database; 18,048 recipients; 14,168 donors with clinical and laboratory data;



$p(\mathbf{O})$
Organ becomes available
Organ similarity
Required organs
$p(\mathbf{O}^*)$
Donor organ availability
Recipient has a better match but higher probability of a future optimal donor
Recipient has a less good match but a low probability of receiving a future optimal donor match
Future '*optimal*' donor organ probability

Using our ITE model	FIFO	SPF	BM	IS	CM	OrganITE
Population life years	83509	92153	104889	111228	110129	**112359**
Deaths in $\mathcal{X}_Q$	0.2646	0.2309	0.2357	0.2067	0.2038	**0.1926**
Deaths before 5 years in $\mathcal{X}_M$	0.1683	0.1869	0.1702	0.1593	0.1891	**0.1472**
Avg. days alive in $\mathcal{X}_Q$	32.49	32.38	32.81	32.65	33.12	**37.19**
Avg. years alive in $\mathcal{X}_M$	4.347	4.138	5.088	5.057	5.165	**5.905**

FIFO – first in - first out

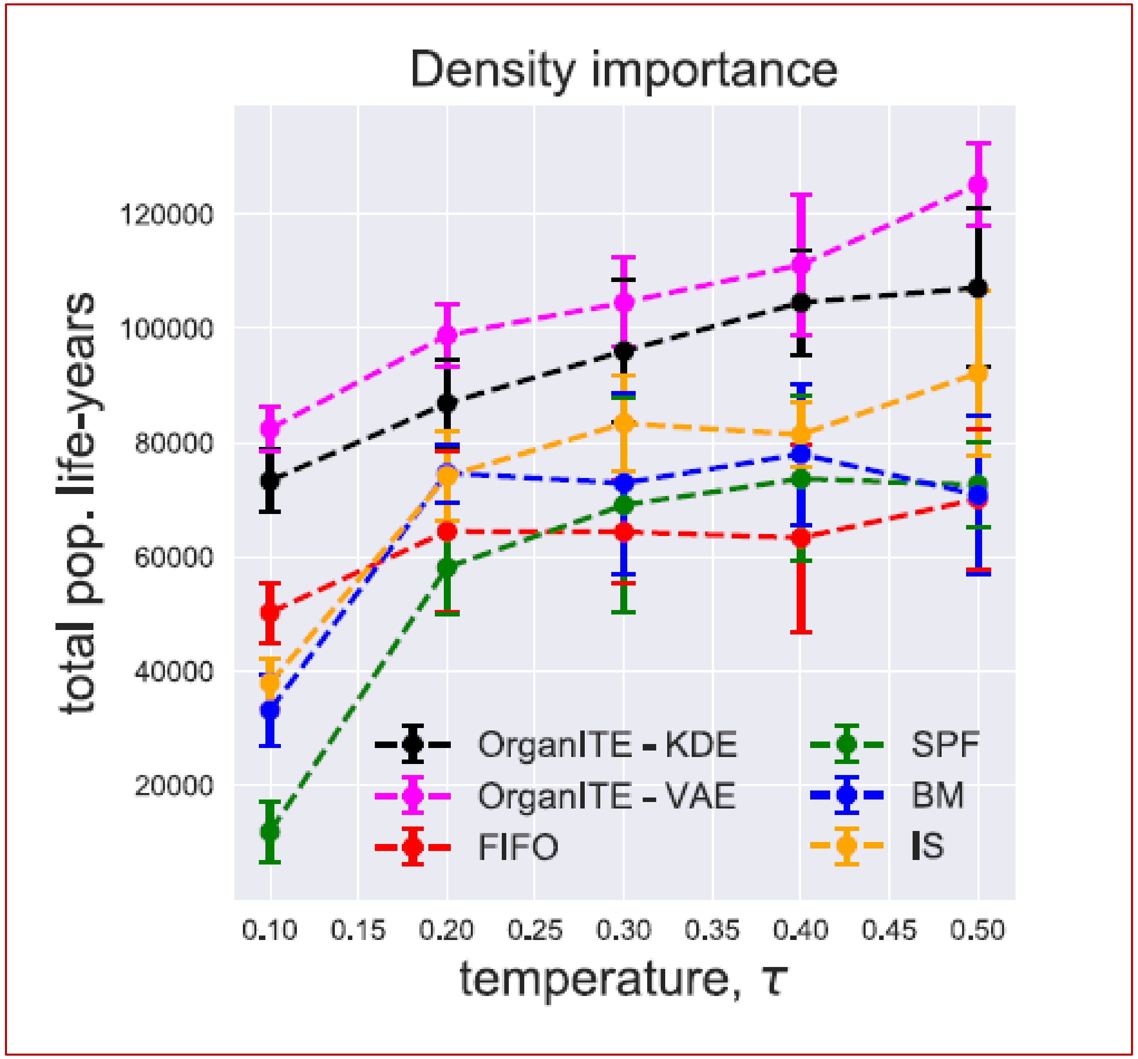
SPF – sickest patient first

BM – best match for post transplant survival

IS – incremental survival, transplant benefit without considering organ density

X_Q- waiting list mortality

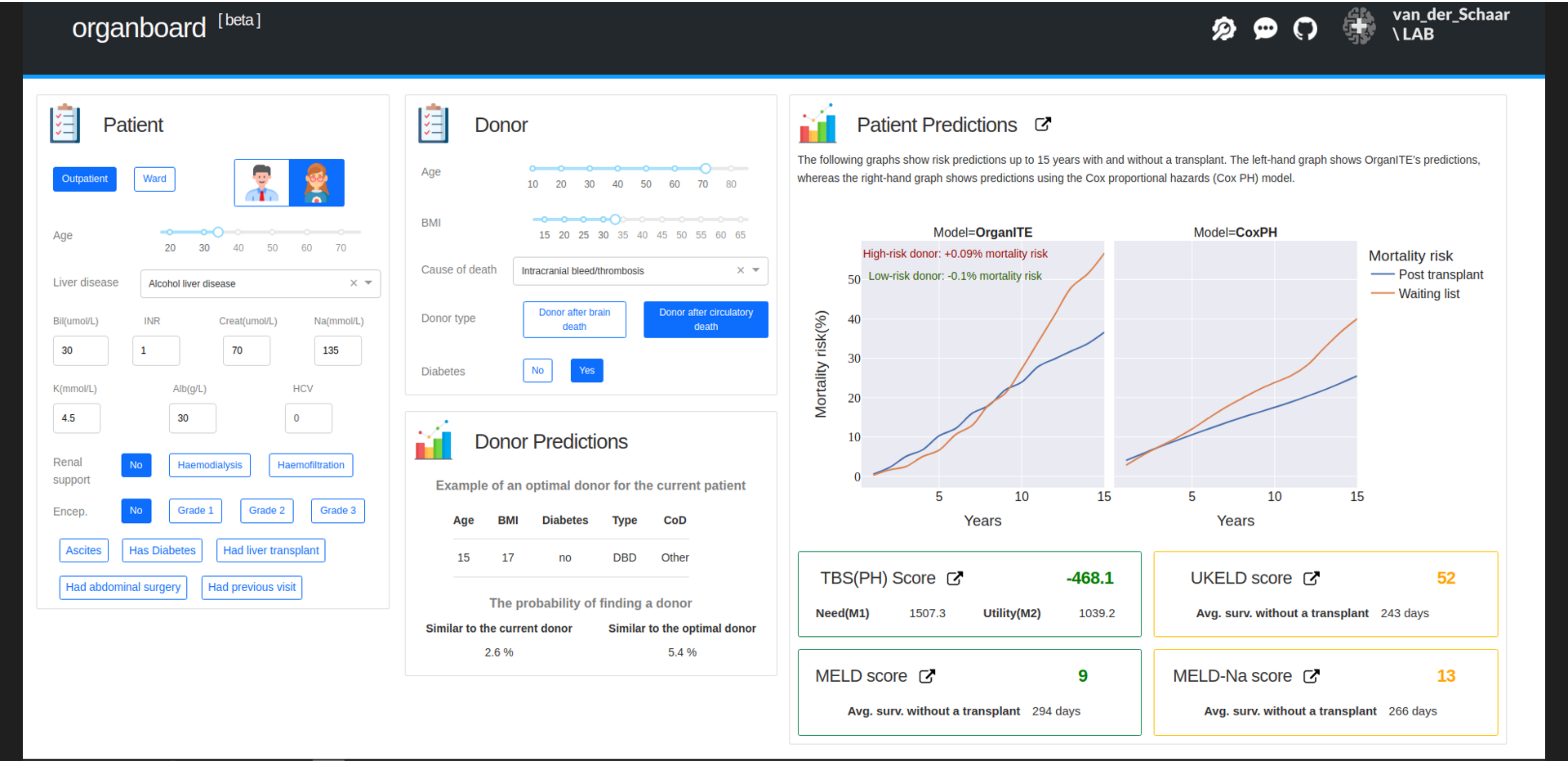
X_M- post transplant mortality



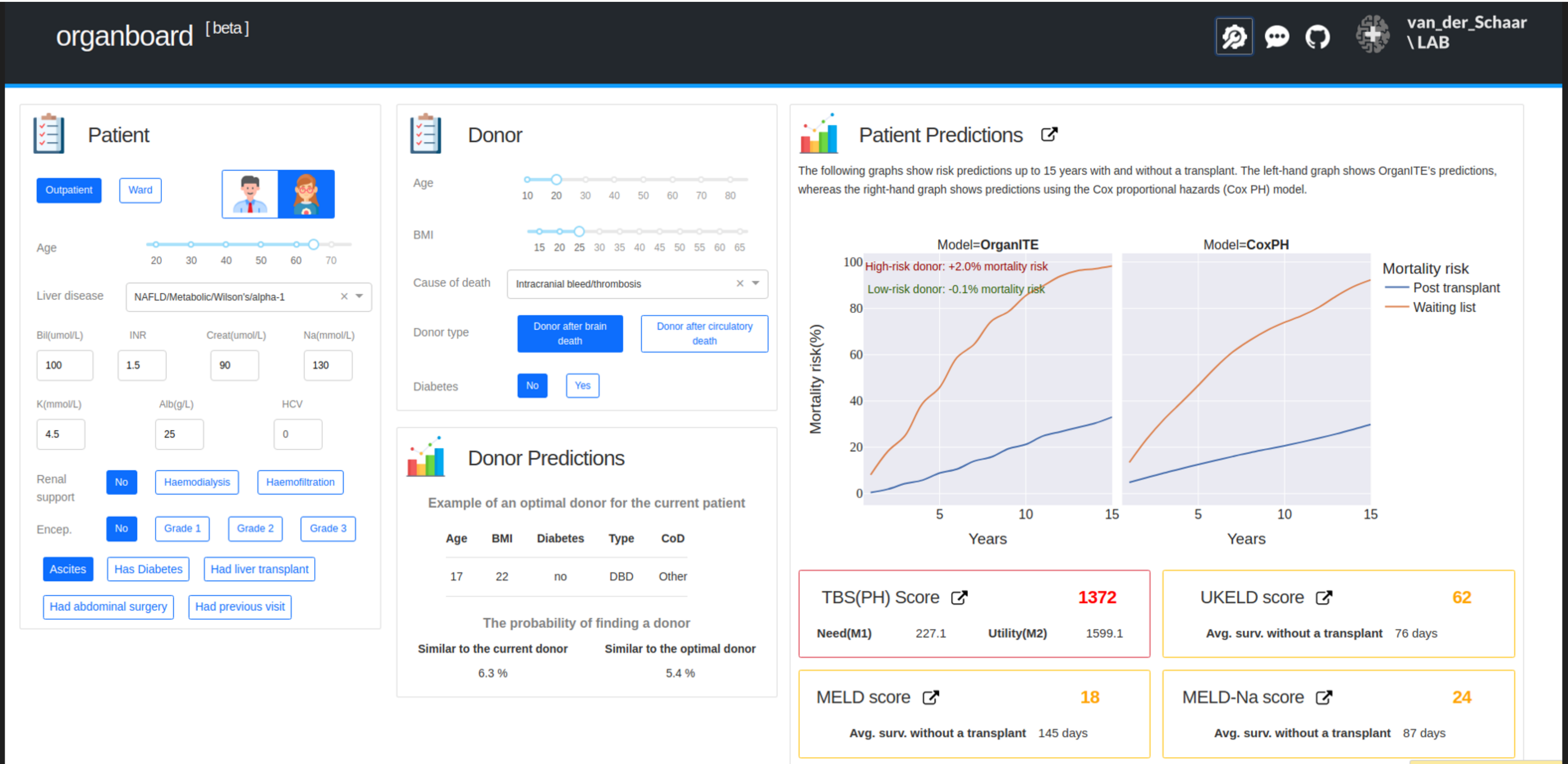
Density importance
total pop. life-years
120000
100000
80000
60000
40000
20000
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
temperature, τ
OrganITE - KDE
OrganITE - VAE
FIFO
SPF
BM
IS

CCAIM

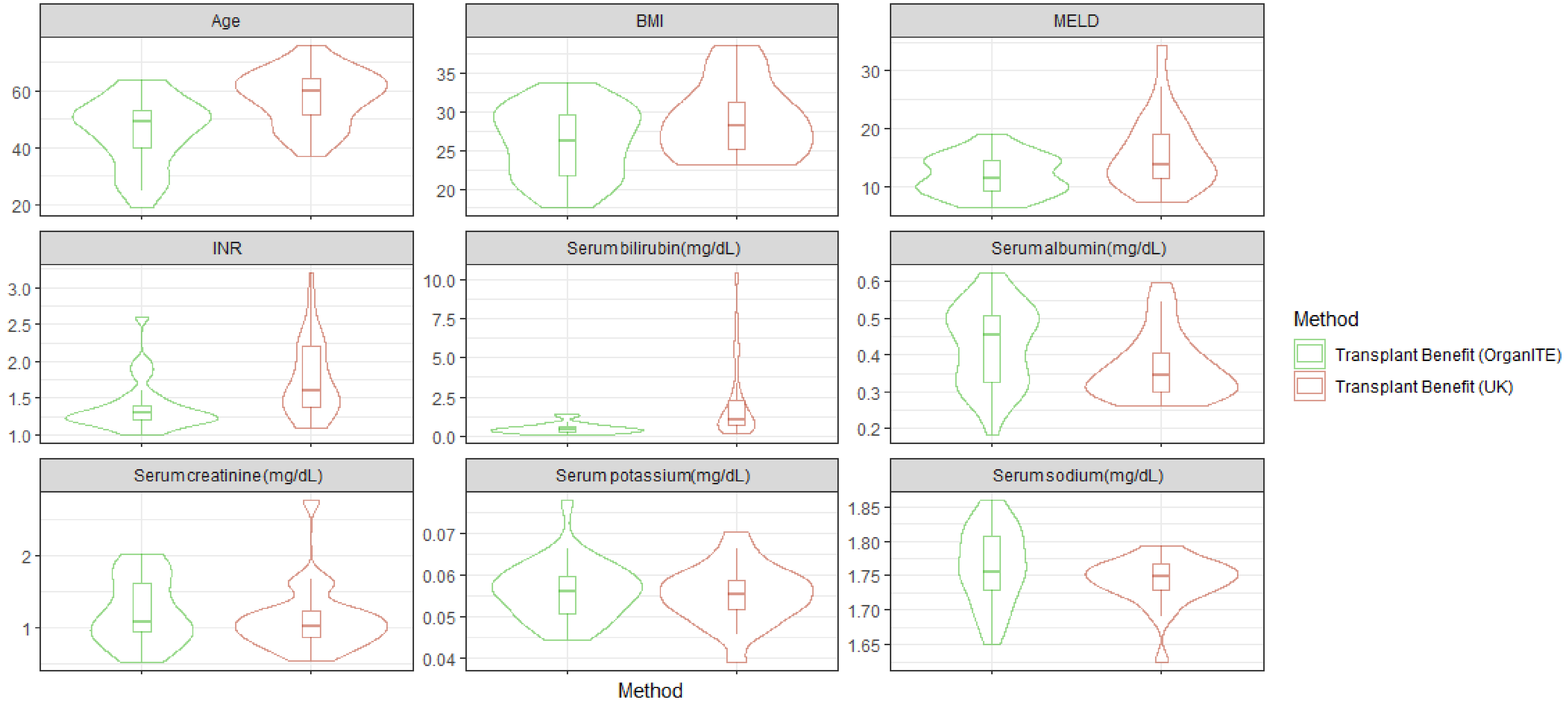
Low Risk recipient with a High Risk donor



High Risk recipient with a Low Risk donor



Characteristics of patients transplanted by OrganITE and CoxPH



CCAIM

Learning Queueing Policies for Organ Transplantation Allocation using *Interpretable* Counterfactual Survival Analysis

Jeroen Berrevoets[1] Ahmed M. Alaa[2] Zhaozhi Qian[1]
James Jordon[3] Alexander Gimson[4] Mihaela van der Schaar[1 2 5]

Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 2021.

OrganSync

1. ITE survival estimation, with organ density
2. An interpretable high-dimensional potential outcomes estimator
3. An new queueing-theoretic framework

Learning queueing policies using interpretable counterfactual survival analysis
OrganSync

Modelling the future arrival distribution of the high-dimensional donor organ space is difficult.

Group donors into a queue with similar ‘outcomes’

Reduce the problem of estimating the complete future organ arrival process, to estimating the arrival process of k distinct “types” of organs. (cohorts, groups, classes)

When a patient enters the transplant system

1. Placed in one of the clusters on basis of their optimal outcome from both survival with that organ class and survival in the time before organs in that cluster are expected to arrive.
3. Within each organ cluster class we use the patient’s survival without an organ to prioritise them in their cluster’s ranking.

When a new donor organ arrives

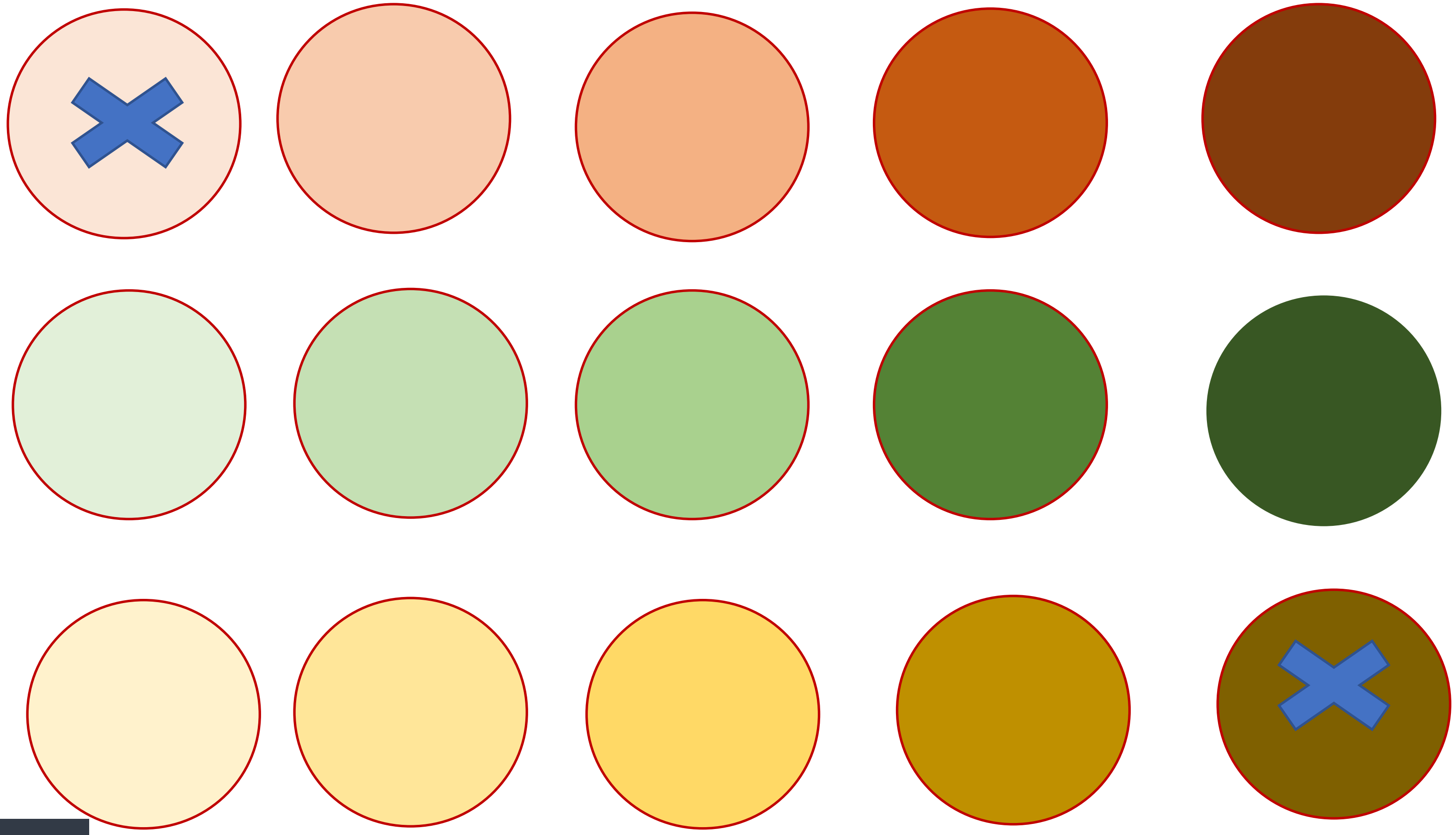
1. Placed in the cluster class more closely resembling it
2. Offered to the first ranked in the organ cluster class queue.

CCAIM

Table 2. **Results on organ allocation.** For each dataset we report the allocation performance of the benchmarks outlined in Table 1, in terms of added life-years (ALY), as well as total deaths over the course of one year. We set MELD as the baseline, to compare against. All results are reported in percentages ("%" is dropped for brevity) and ran over ten data-folds, standard deviation in brackets.

	UNOS		UKReg	
Method	Deaths	ALY	Deaths	ALY
MELD	compared against			
FIFO	-0.9 (.01)	-2.0 (.11)	-1.1 (.16)	-5 (.01)
M-na	-0.3 (.13)	+1.2 (.10)	-2.1 (.18)	+6 (.01)
TB	+7.0 (.19)	+2.4 (.21)	+0.9 (.11)	+8 (.03)
CM	-0.01 (.09)	+12.8 (.31)	+0.1 (.11)	+7 (.02)
O-ITE	**-3.6 (.18)**	+11.1 (.28)	-3.3 (.12)	+11 (.15)
OS	-3.5 (.15)	**+13.1 (.19)**	**-4.1 (.21)**	**+ 13 (.03)**

Each cluster class will have specific donor and recipient features that are different to other classes, but are associated with similar outcomes, allowing interpretation of reasons why allocation of a specific donor to a particular recipient was made



Each cluster will differ with respect to donor features (e.g. age, DM, BMI, cause of death, DCD,DBD…..)

and recipient parameters (eg Na, bilirubin, albumen, INR, age, clinical characteristics…)

Optimal organ allocation processes

Outcome without a donor

Outcome with a specific donor

Time till another "better / optimal" donor appears

Mortality waiting for the "better / optimal donor"

Impact of any deterioration in clinical status whilst waiting

Impact of new potential recipients on the transplant list

Interpretable results

CCAIM

Clinical variation has a crucially important impact on patient care and outcomes

Liver transplant center variability in accepting organ offers and its impact on patient survival

David S. Goldberg, MD, MSCE[1,2,3], Benjamin French, PhD[2,3], James D. Lewis, MD, MSCE[1,2,3], Frank I Scott, MD, MSCE[1,2], Ronac Mamtani, MD, MSCE[4], Richard Gilroy, MD[5], Scott D. Halpern, MD, PhD[2,3,6], and Peter L Abt, MD[7]

J Hepatol. 2016 April ; 64(4): 843–851. doi:10.1016/j.jhep.2015.11.015.

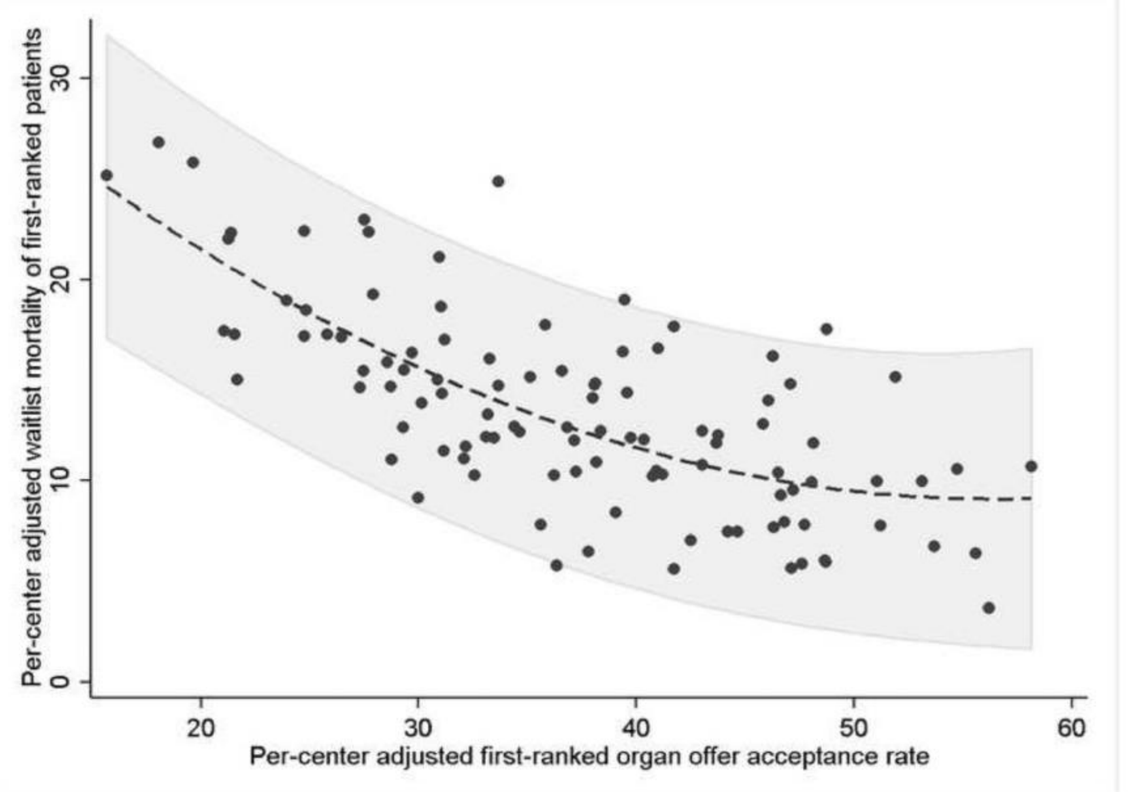
23,740 organ offers, 8,882 (37.4%) accepted for the first-ranked patient.

Adjusted center-specific organ acceptance rates (OAR) ranged from 15.7% to 58.1%.

For every 5% decrease in OAR 27% increased odds of waitlist mortality

4% absolute difference in median 5-year graft survival

Variance in clinical decisions have important consequences



Donor offer acceptance rates for donors after brain death and donors after cardiac death. UK 2018-2020

Figure 3.11 Named adult elective liver offer decline rates that resulted in a liver only first transplant from DBD donors, 1 April 2018 and 31 March 2020

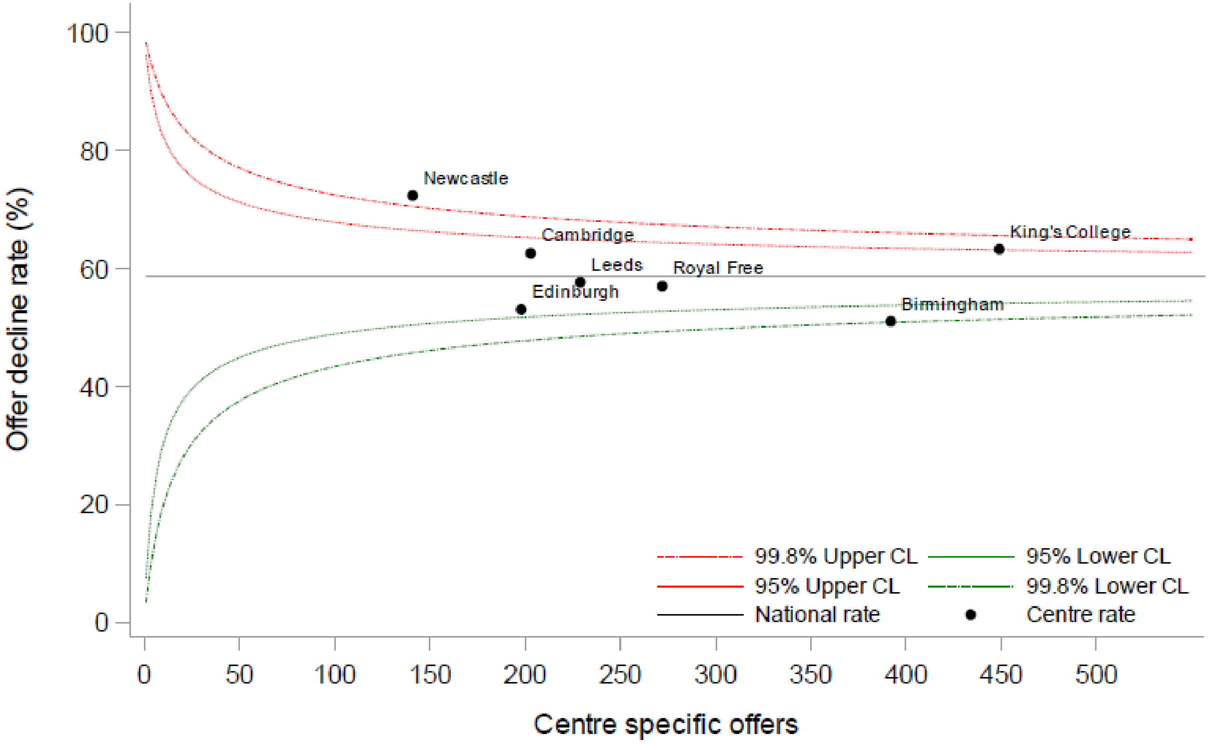
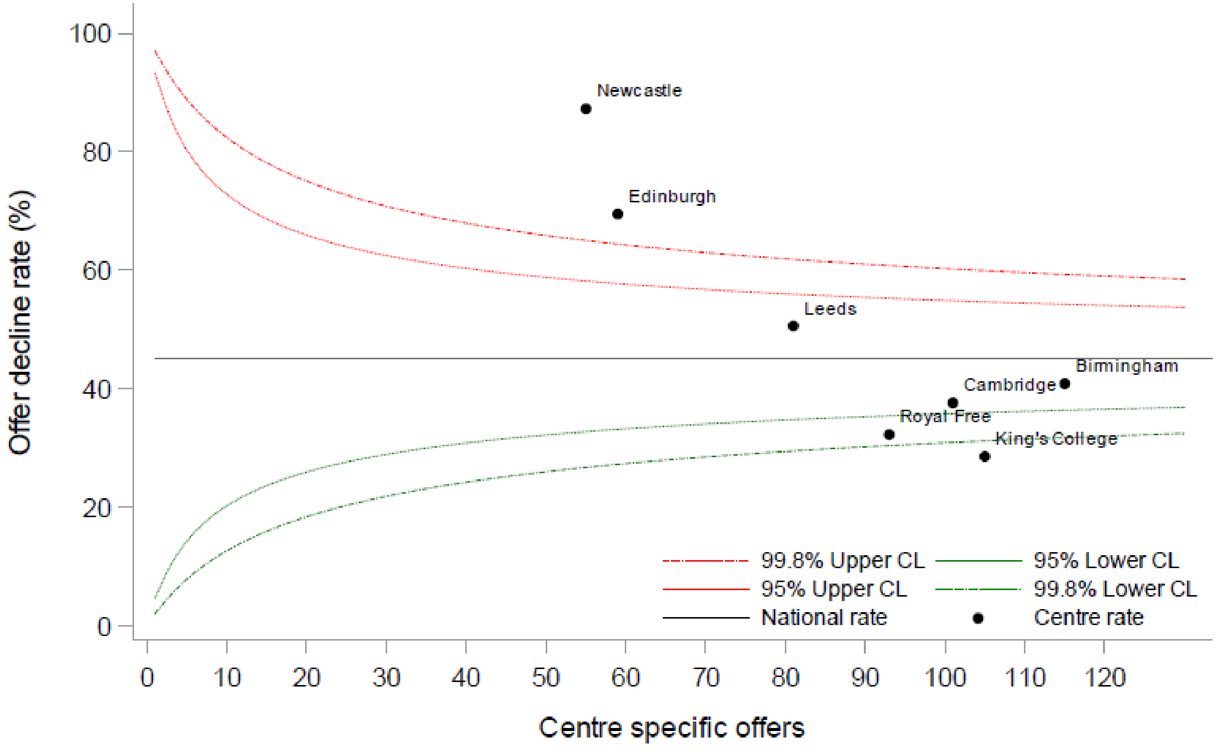


Figure 3.12 Adult elective liver offer decline rates that resulted in a liver only first transplant from DCD donors, 1 April 2018 and 31 March 2020



Addressing Clinical Variation – quantitative epistemology

Can we identify the drivers of clinical decisions

- at a population level?

- at a instance-wise level ?

- how such drivers have changed with time?

- national allocation guidelines/policies ?

Closing the loop in medical decision support by understanding clinical decision-making: A case study on organ transplantation

Yuchao Qin
University of Cambridge
yq257@cam.ac.uk

Fergus Imrie
University of California, Los Angeles
imrie@g.ucla.edu

Alihan Hüyük
University of Cambridge
ah2075@cam.ac.uk

Daniel Jarrett
University of Cambridge
daniel.jarrett@maths.cam.ac.uk

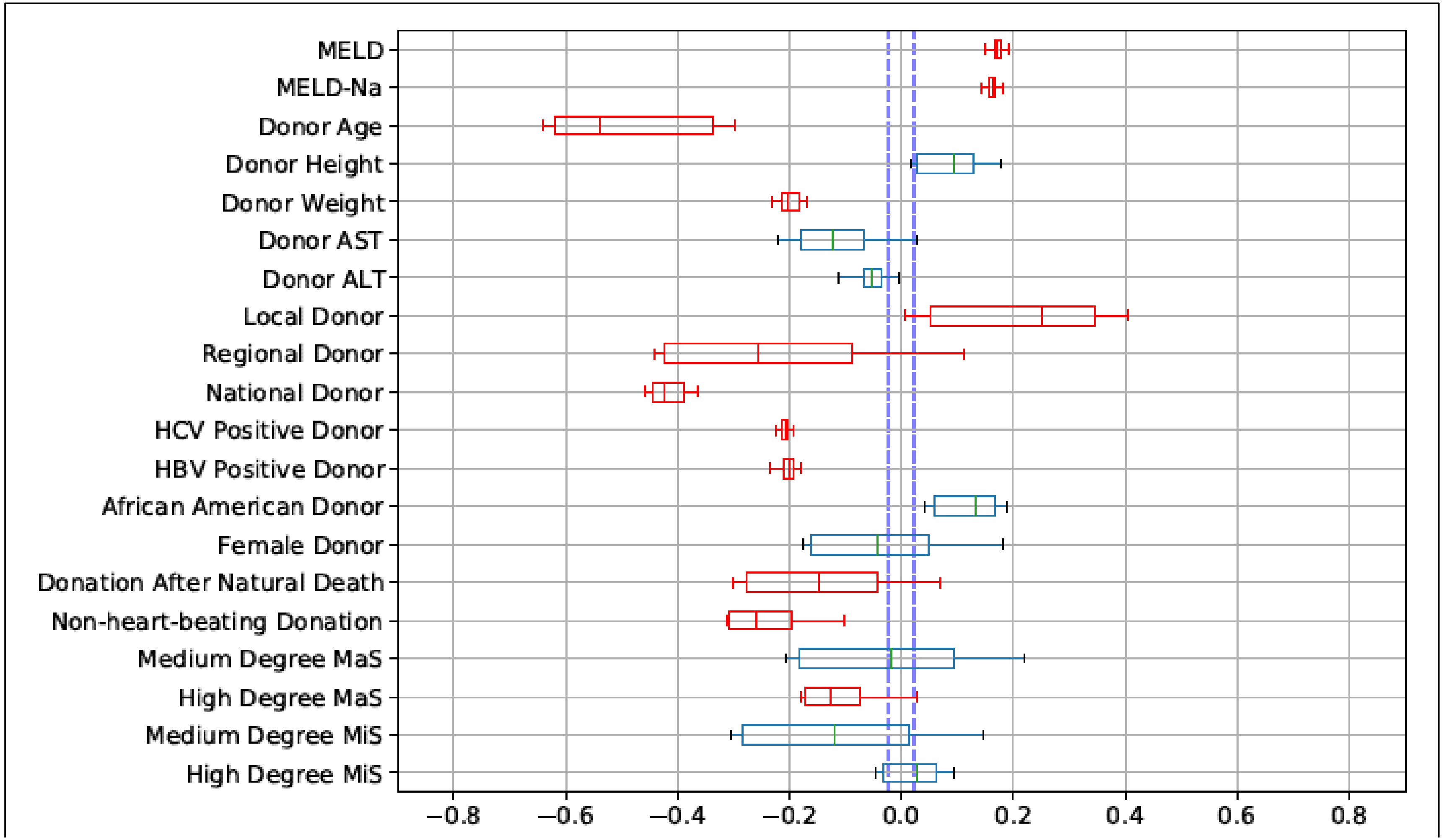
Alexander Edward Gimson
University of Cambridge
alexander.gimson@nhs.net

Mihaela van der Schaar
University of Cambridge
mv472@cam.ac.uk

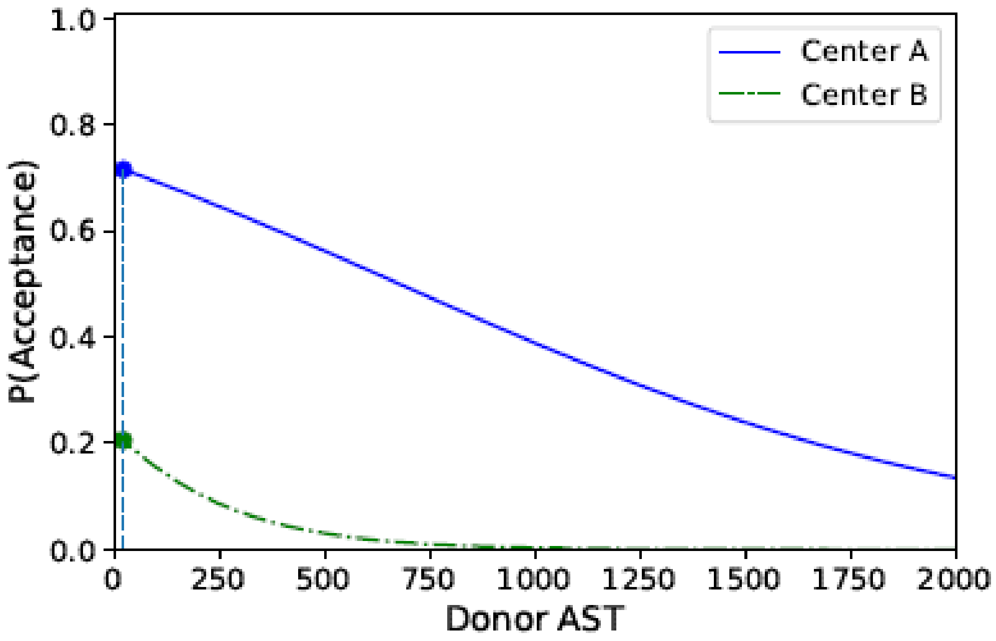
- Discover which criteria are most important to clinicians for organ offer acceptance;
- Identify patient-specific organ preferences of centres

Explore variations in transplantation practices between different transplant centres.

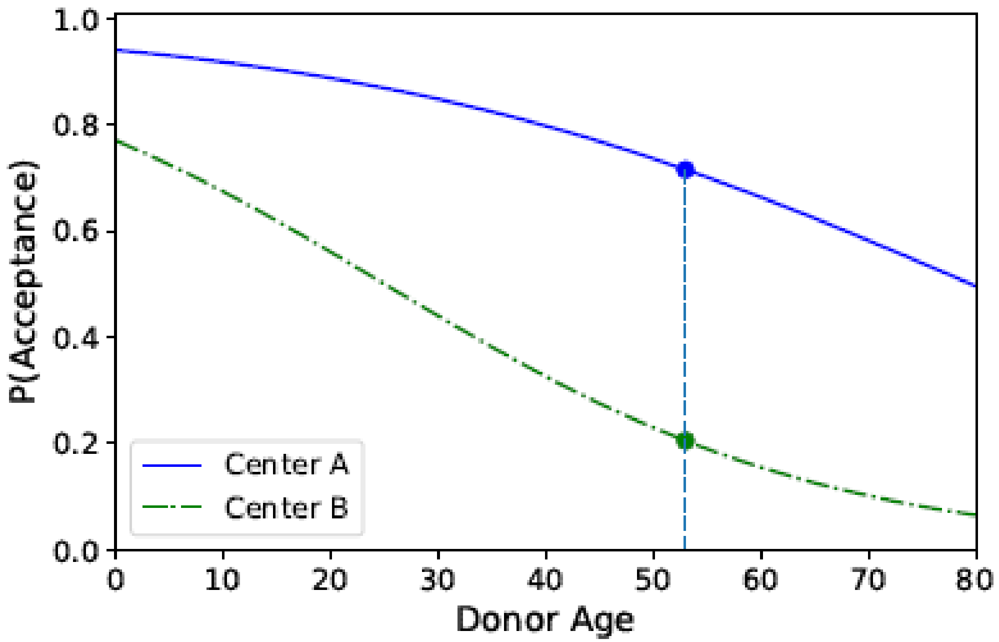
We achieve this by training a neural network-based policy selector to identify individualized policies for patients from different cohorts. These policies act on the space of known match criteria using a white-box function, ensuring interpretability with respect to the match criteria.



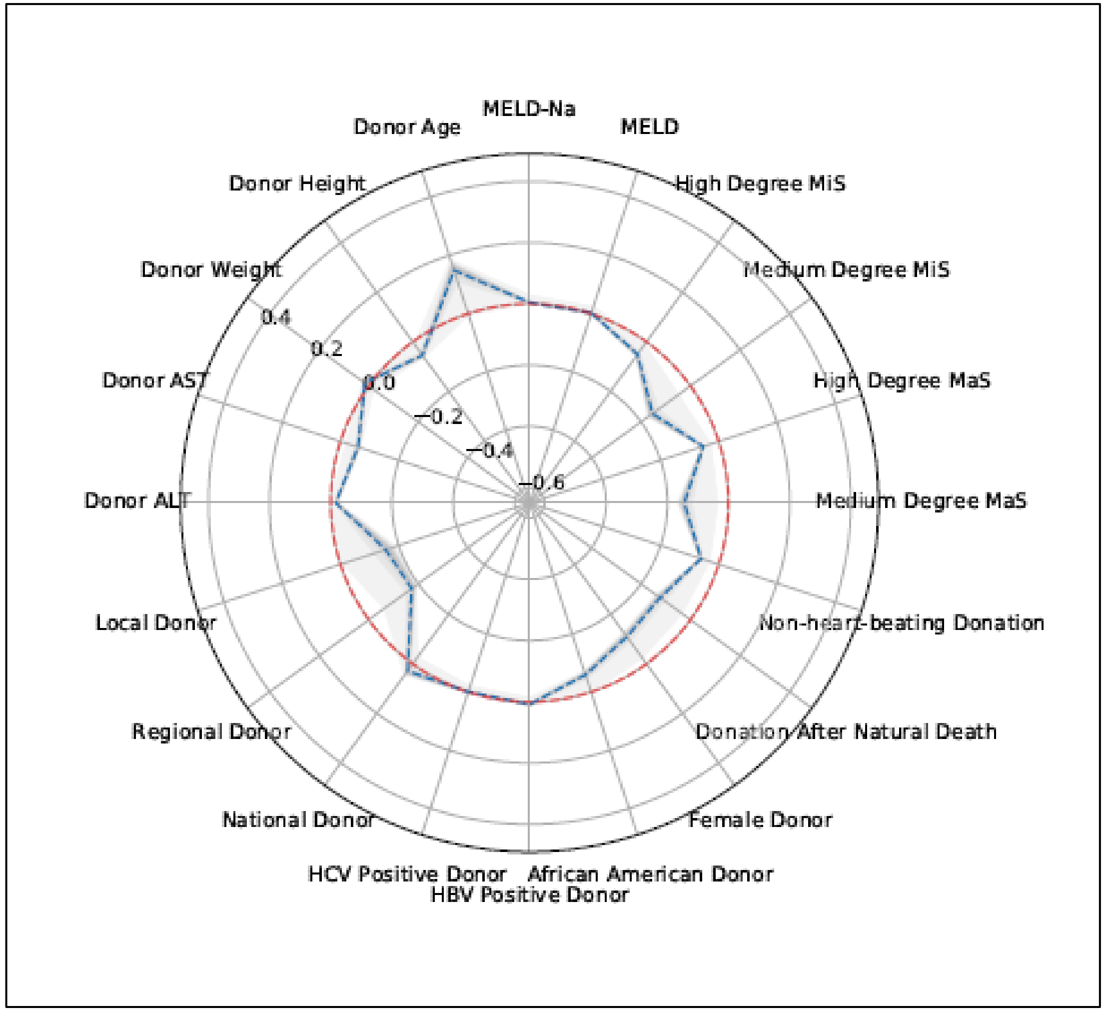
METHOD	AUC-ROC	AUC-PRC	LL
LOGISTIC REGRESSION (LR)	0.794±0.054	0.341±0.061	-0.538±0.051
PER-CLUSTER LR	0.803±0.049	0.352±0.063	-0.527±0.048
PER-CLUSTER LR (WITH INTERACTION TERMS)	0.824±0.054	0.371±0.073	-0.490±0.063
ITRANSPLANT (OURS)	0.898±0.048	0.508±0.064	-0.385±0.076

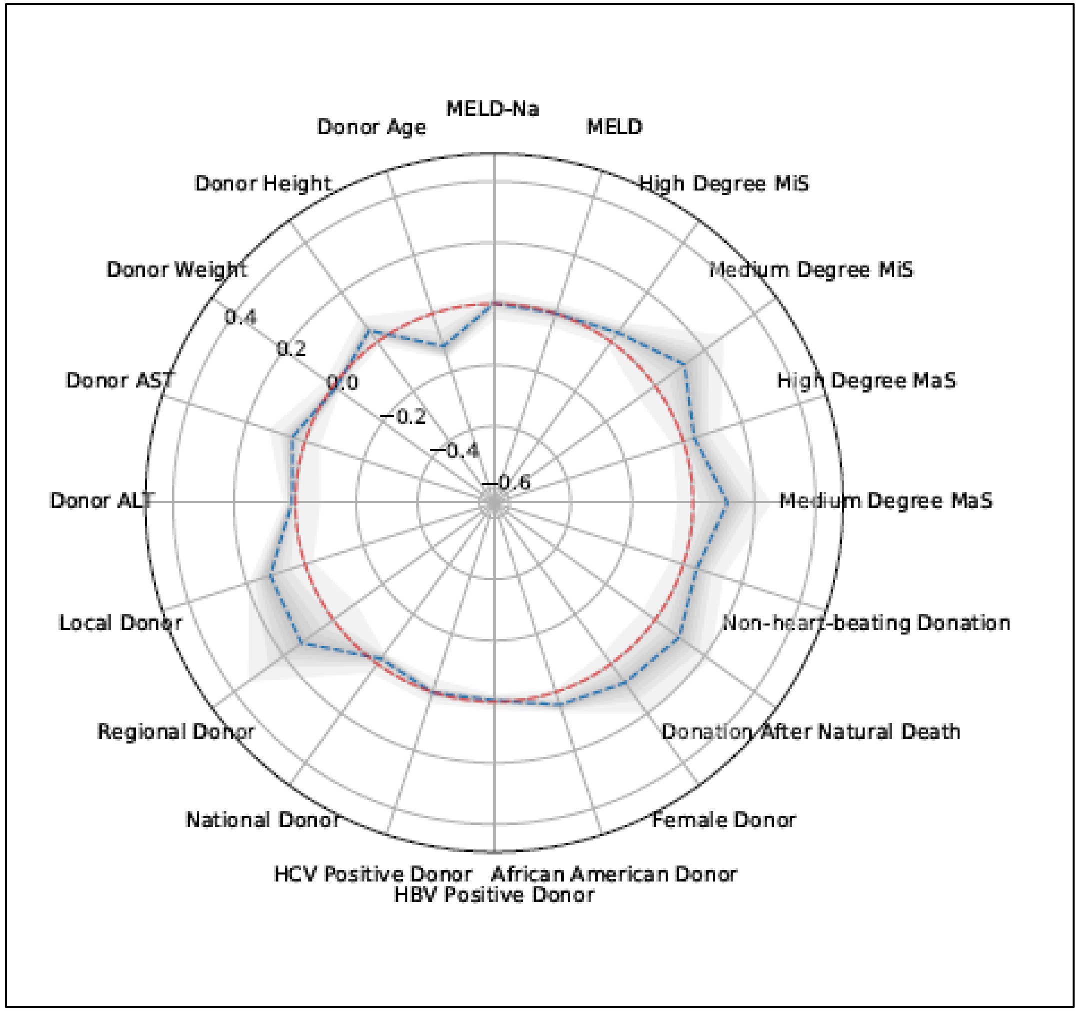


(a) Counterfactual impact of donor AST test value.



(b) Counterfactual impact of donor age.





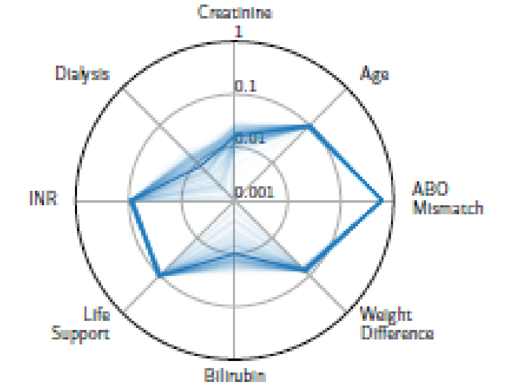
Cluster	MELD Score					HCC Status	
	$[0, 10)$	$[10, 20)$	$[20, 30)$	$[30, 40)$	$[40, +\infty)$	Positive	Negative
1	14.8 %	76.36%	8.67%	0.17%	0.00%	1.53%	98.47 %
2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3	0.97%	65.52%	27.59%	4.41%	1.52%	0.00%	100.00%

Inverse Contextual Bandits: Learning How Behavior Evolves over Time

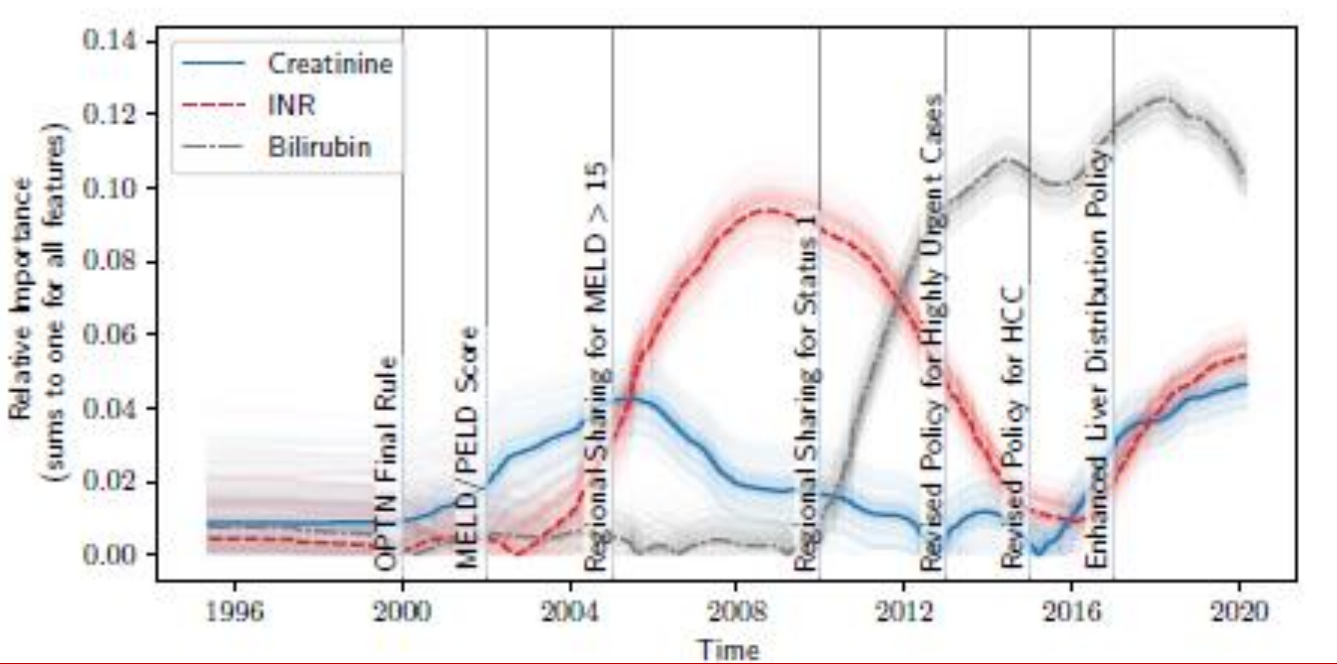
Alihan Hüyük * [1] Daniel Jarrett * [1] Mihaela van der Schaar [1 2]

https://arxiv.org/abs/2107.06317

(a) Feature Importances in 2000



(b) Feature Importances in 2010



Future ecosystems of ML applications in chronic liver disease

Chronic liver disease

Transplant waiting list

Post transplant

Ascites

Renal failure

Variceal haemorrhage

Autoprognosis 2.0

OrganITE
OrganSync

i-Transplant, ICBs

Temporal phenotyping
for treatment selection

Individualised immunosuppressive treatment regimens

Predicting graft failure

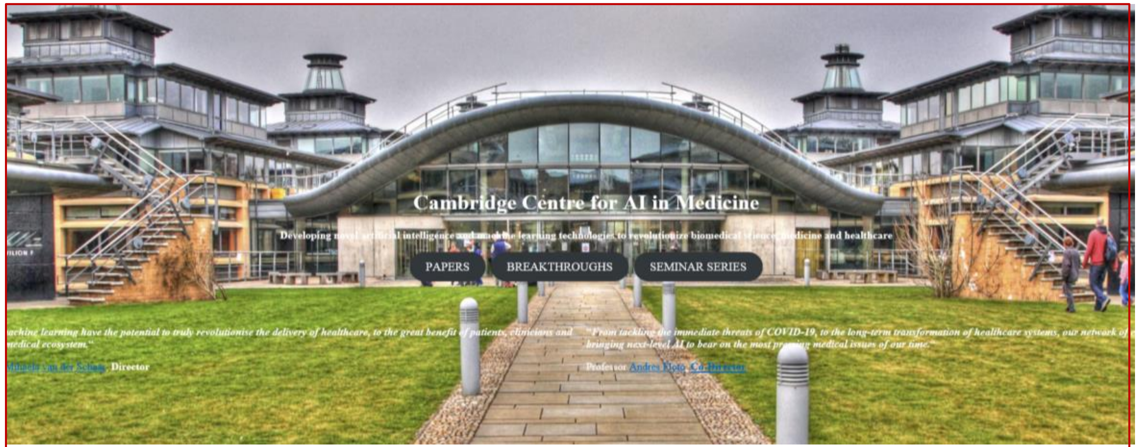
Monitoring regimens

CCAIM

Some caveats

- Implementation of change within Medicine is often slow
- Interpretability and predictive accuracy are two main components of trust in new AI methodologies
- Public involvement
- Changes in waiting list therapies
- Changes in treatments for specific diagnoses
- New indications for transplantation
- Waiting list entry criteria at an individual versus a population level

Acknowledgements



Professor Mihaela van der Schaar
Dr Ahmed Alaa
Dr James Jordan
Jeroen Berrevoets
Zhaozhi Qian
Dr Fergus Imrie
Alihan Huyuk
Bogdan Cerebrere

CCAIM

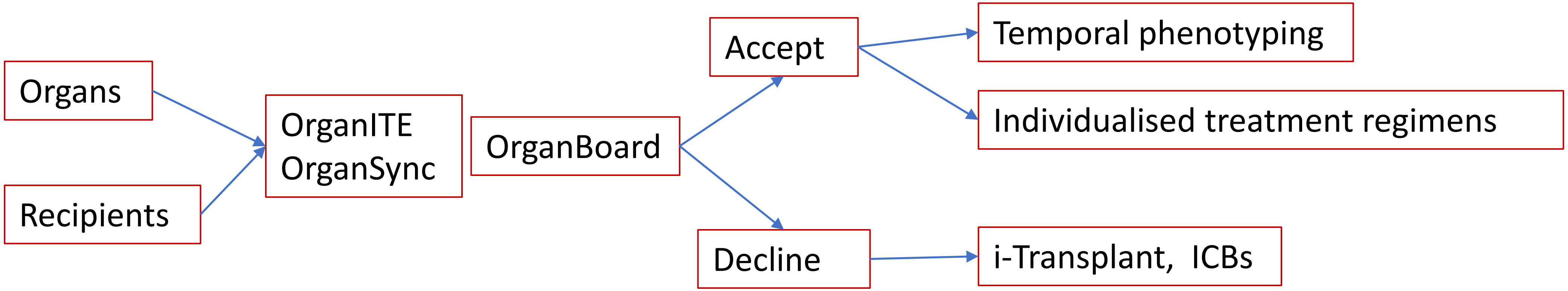
NHS Blood & Transplant Statistics
Bristol UK

Dr Kerrie Barber
Dr Alisa Allen
Dr Rhiannon Taylor

Members of the UK Liver Selection and Allocation Working Parties

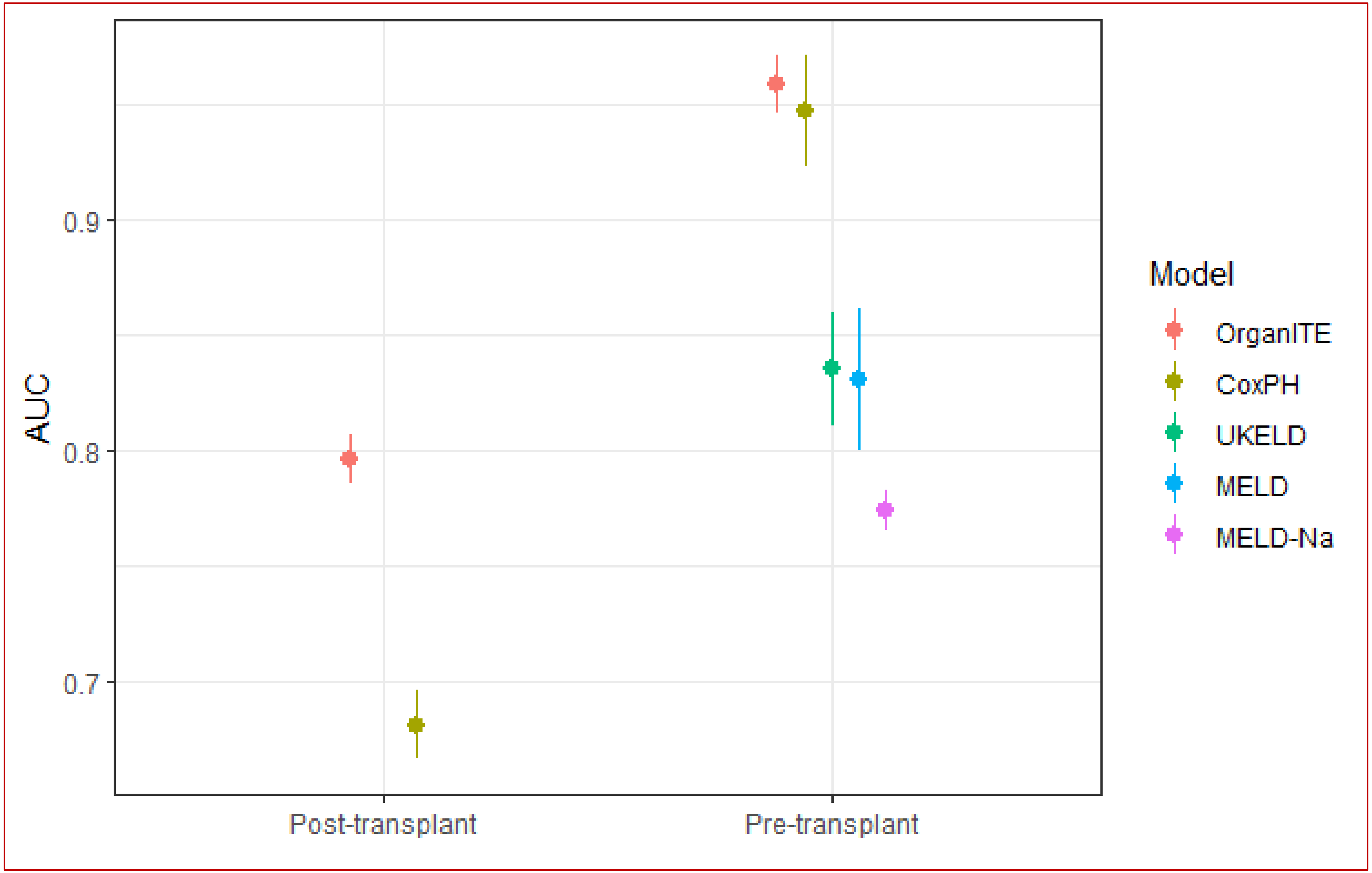
CCAIM

Future ecosystems of ML applications in organ transplantation



CCAIM

Predictive performance of OrganITE



CCAIM