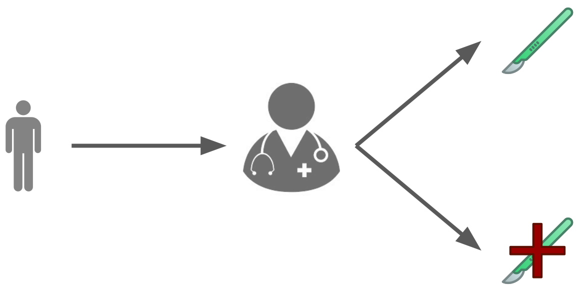
Clinical Presence: Impact on Algorithmic Fairness

Vincent Jeanselme

2024.11.21

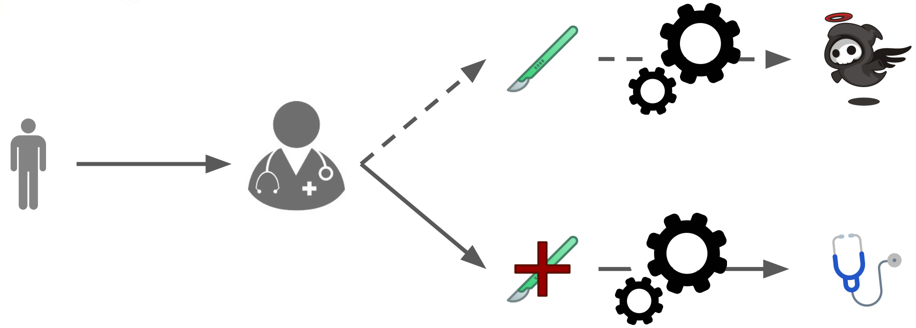
Medical data can improve care



Medical data can improve care

The New York Times

How Artificial Intelligence Could Transform Medicine



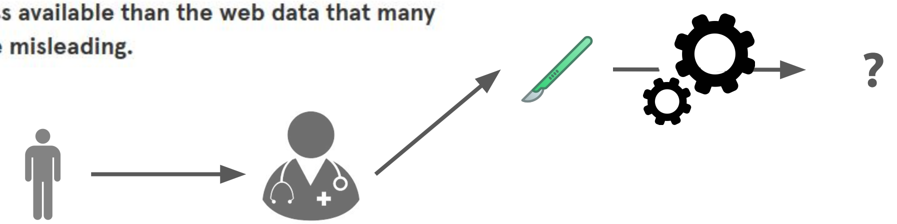
Predictive models can inform decision-making

Medical data present modelling challenges

WIRED

When It Comes to Health Care, AI Has a Long Way to Go

Medical information is more complex and less available than the web data that many algorithms were trained on, so results can be misleading.

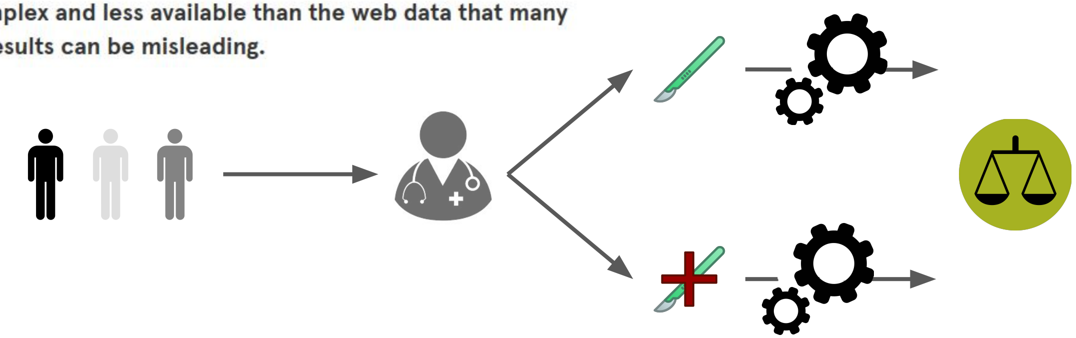


Medical data embed disparities

WIRED

When It Comes to Health Care, AI Has a Long Way to Go

Medical information is more complex and less available than the web data that many algorithms were trained on, so results can be misleading.



UNEQUAL
TREATMENT

CONFRONTING RACIAL AND ETHNIC
DISPARITIES IN HEALTH CARE

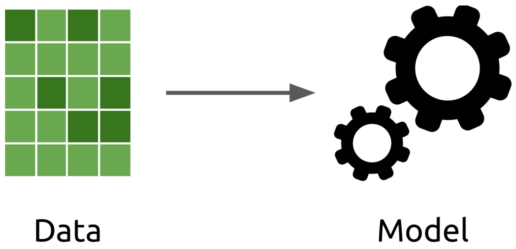
The New York Times

A.I. Could Worsen Health Disparities

In a health system riddled with inequity, we risk making dangerous biases automated and invisible.

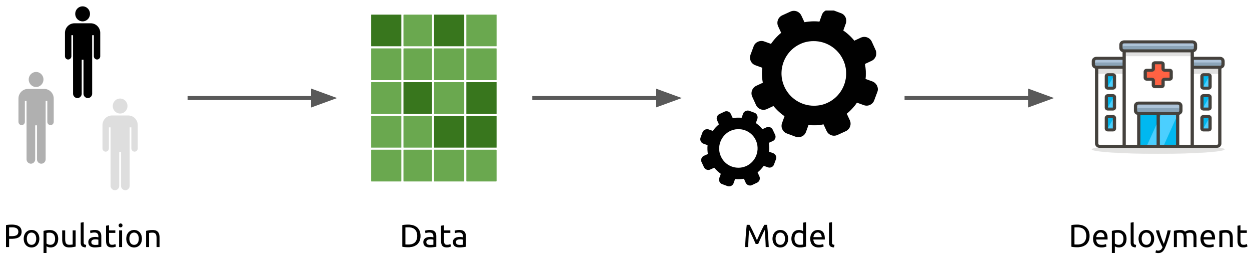
Research

*Develop predictive models for **medical decision-making** and **addressing socio-medical disparities** present in medical data.*

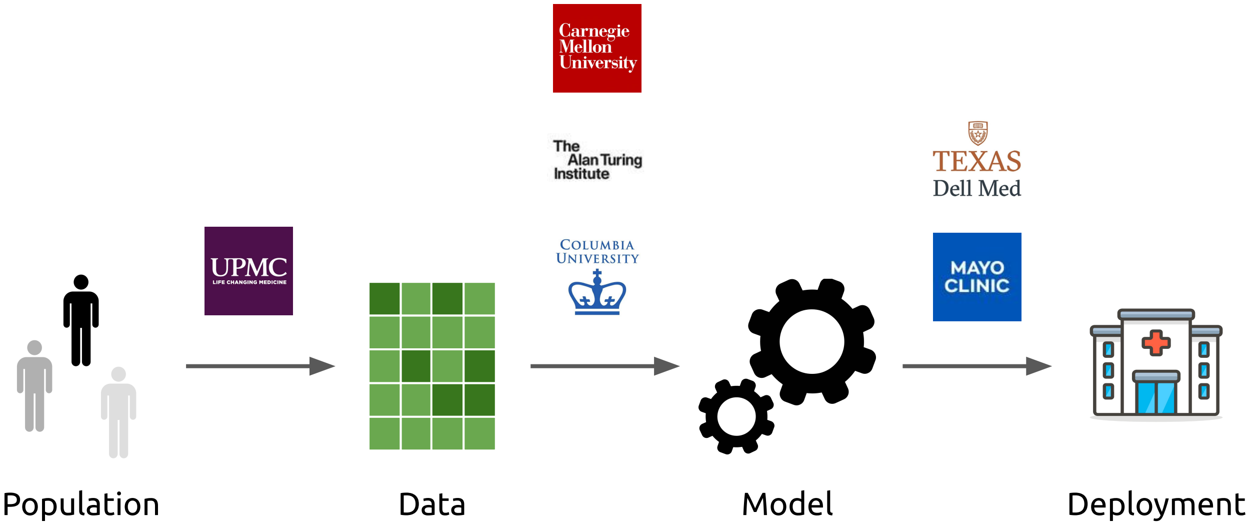


Research

Develop predictive models for ***medical decision-making*** *and* ***addressing socio-medical disparities*** *present in medical data.*



Research



Res + 2021

ML4H 2022

MNSC 2024

AAAI 2021
CHIL 2022
CHIL 2023
AAAI 2024
Under Review 2024

Working 2024

ATS 2020
ISICM 2020
Critical Care 2020
Under Review 2024

NeurIPS 2022

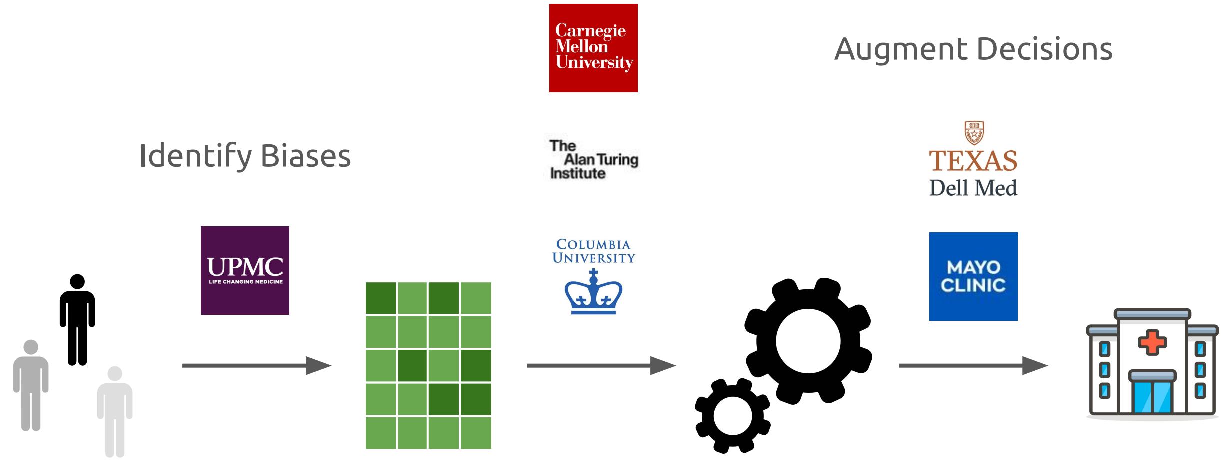
Medical Venues
ML Venues
Business Venues

Research

Mitigate Biases

Augment Decisions

Identify Biases



Population

Data

Model

Deployment

Res + 2021

ML4H 2022

MNSC 2024

AAAI 2021
CHIL 2022
CHIL 2023
AAAI 2024
Under Review 2024

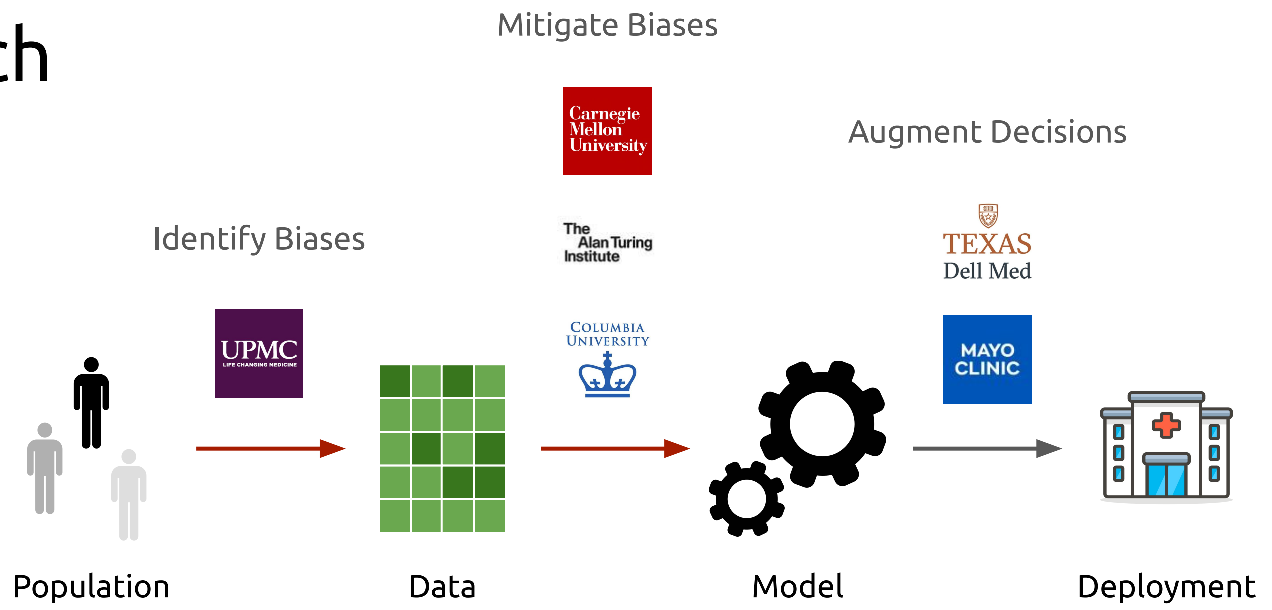
Working 2024

ATS 2020
ISICM 2020
Critical Care 2020
Under Review 2024

NeurIPS 2022

Medical Venues
ML Venues
Business Venues

Research



Res + 2021

ML4H 2022

MNSC 2024

AAAI 2021
CHIL 2022
CHIL 2023
AAAI 2024
Under Review 2024

Working 2024

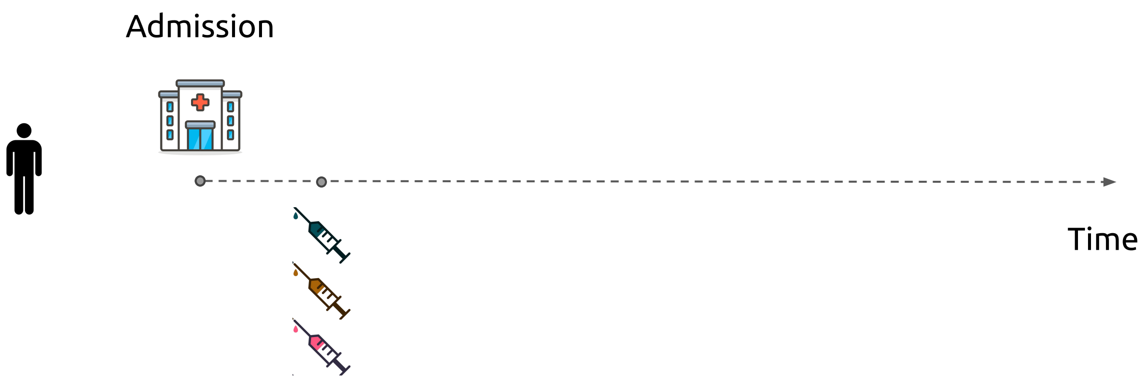
ATS 2020
ISICM 2020
Critical Care 2020
Under Review 2024

NeurIPS 2022

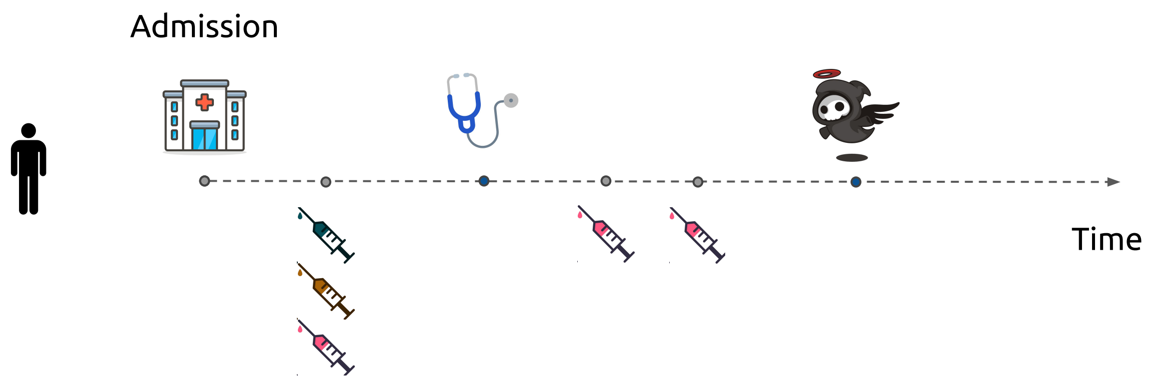
Medical Venues
ML Venues
Business Venues

Clinical Presence

Clinical Presence

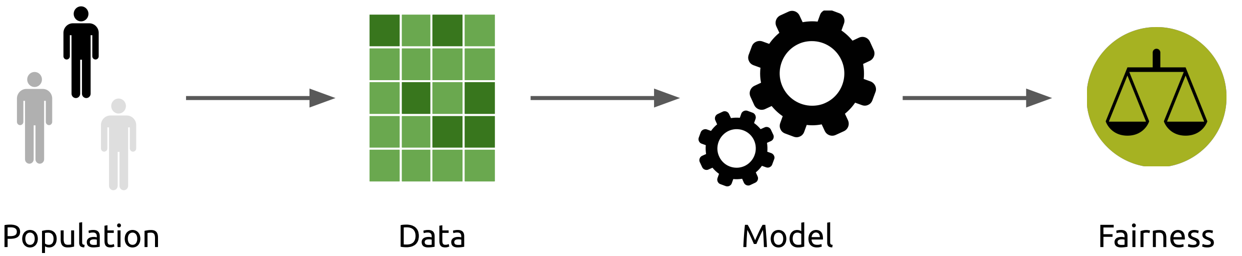


Clinical Presence



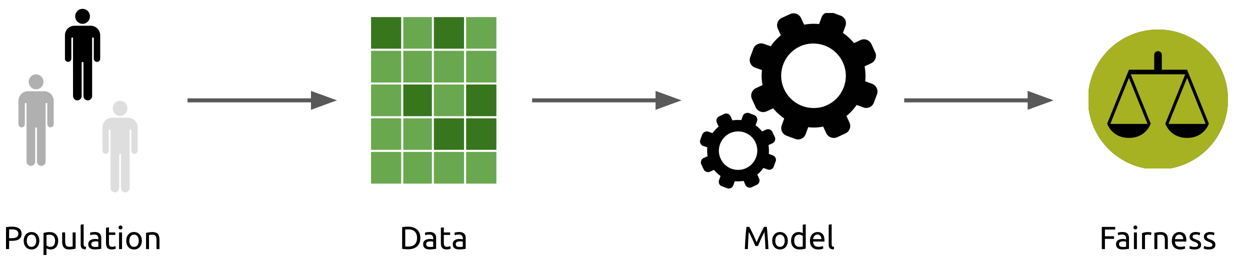
Observations are the result of the interaction between ***patients*** *and the* ***healthcare system****.*

Talk structure



What is algorithmic fairness ?

Talk structure

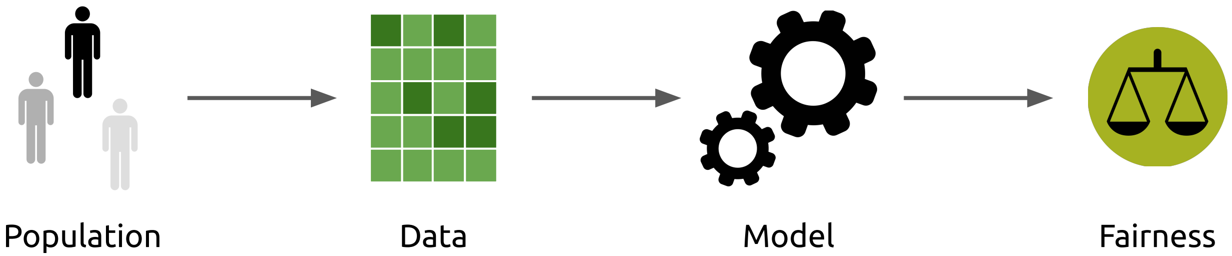


How to inform imputation choice?

What is algorithmic fairness ?

Talk structure

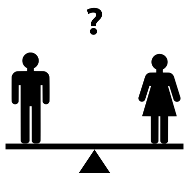
How to model
unobserved outcomes?



How to inform
imputation choice?

What is algorithmic
fairness ?

What is algorithmic fairness?



*This talk focuses on **group fairness**, measured through **equal performance across groups**, i.e. a pipeline is fairer than another with regard to a group if its performance gap is the smallest.*

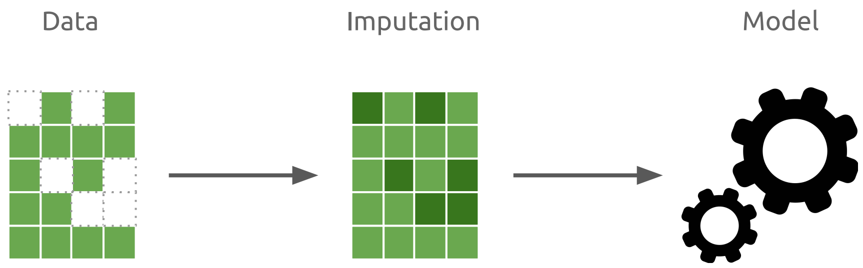
TEXAS McCombs
The University of Texas at Austin
McCombs School of Business

UC San Diego
RADY SCHOOL OF MANAGEMENT

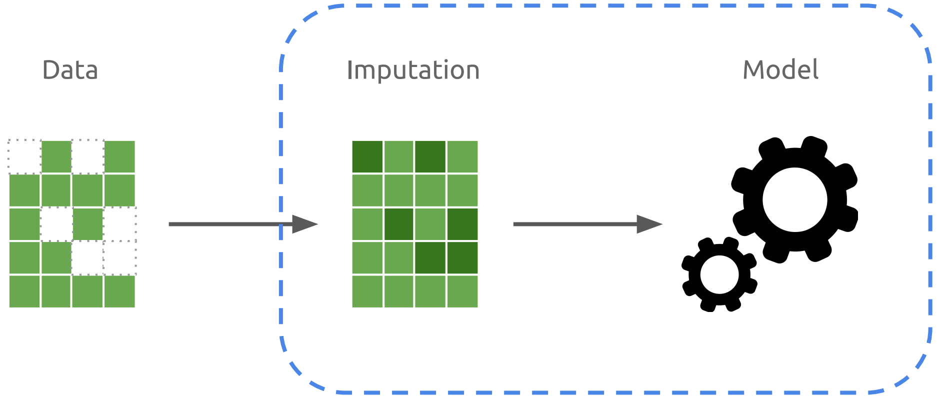
Imputation Strategies Under Clinical Presence: Impact on Algorithmic Fairness

V. Jeanselme, M. De-Arteaga, Z. Zhang, J. Barrett and B. Tom

Canonical pipeline



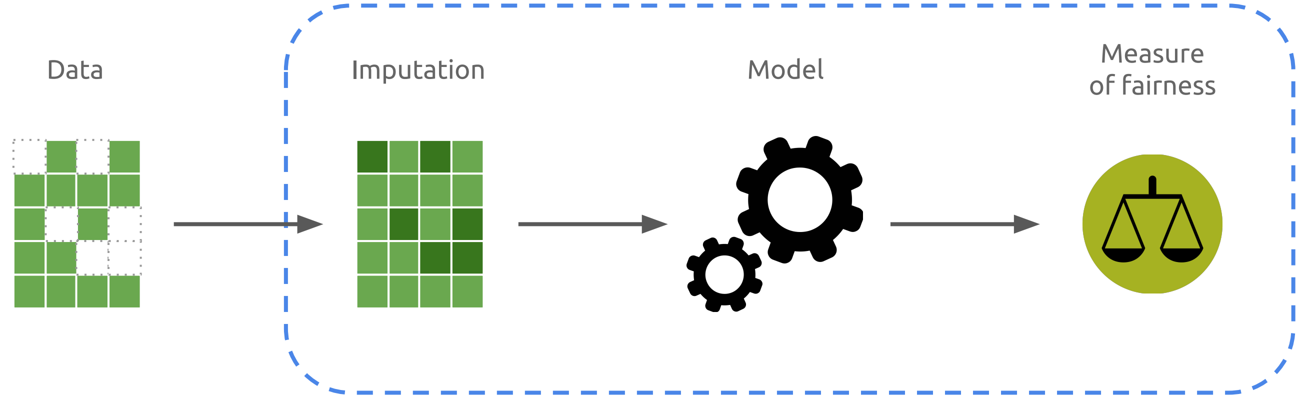
Development of a machine learning pipeline



***65%** of ML for healthcare papers have*
***missingness**, but **<10% report** handling*

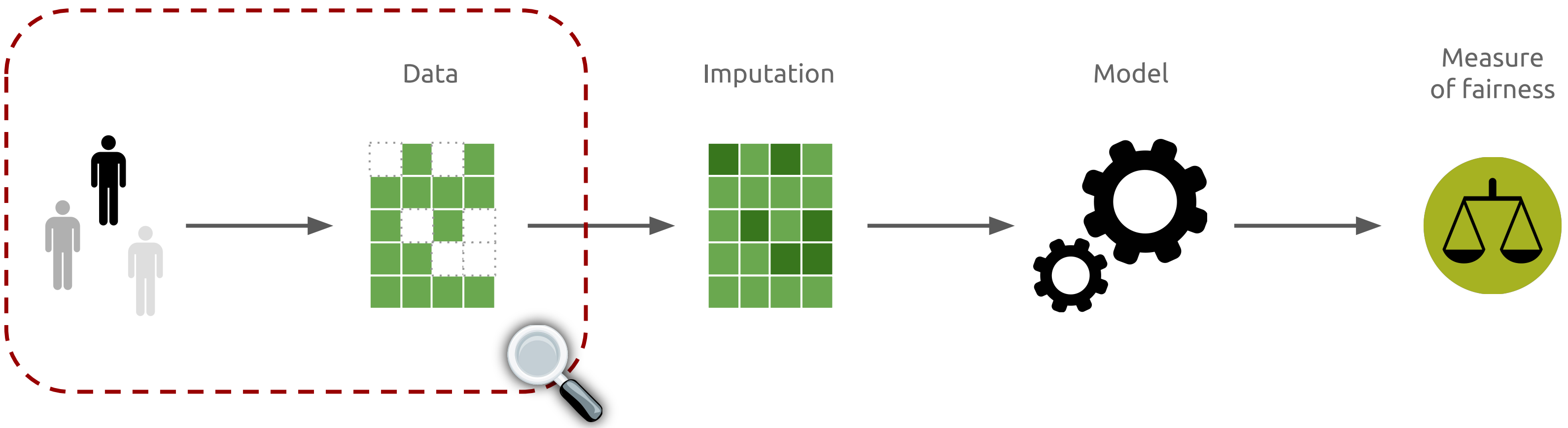
Missing data is poorly handled and reported in prediction model studies using machine learning: a literature review by Nijman et al., 2022

Fairness literature focuses on modelling



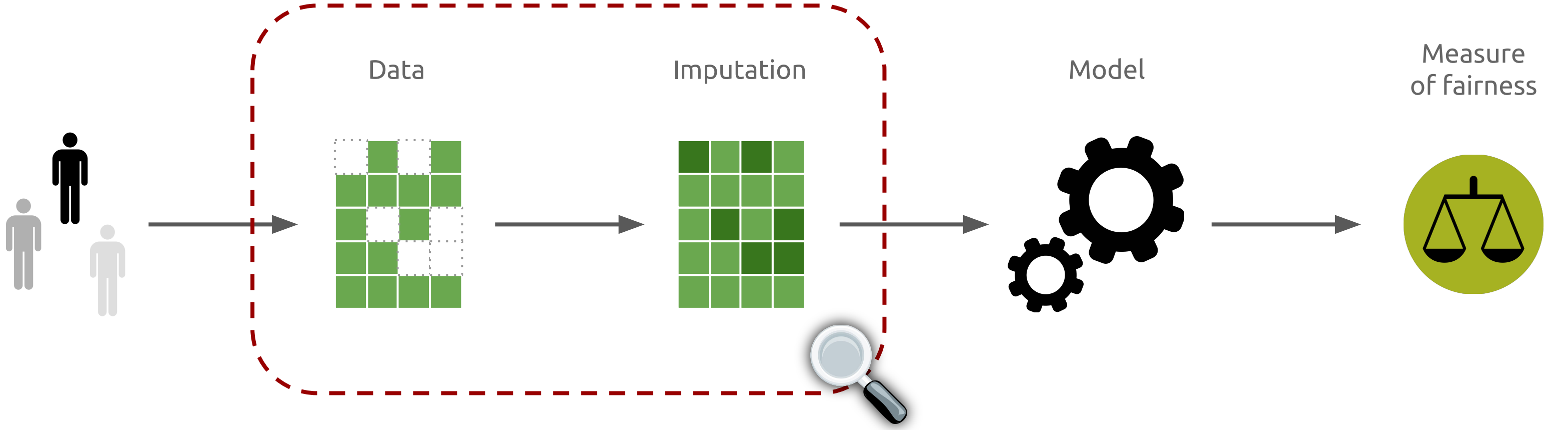
*The fairness literature studies how to **detect** and **mitigate** **biases present in the data**. Current focus has been on **modelling** choices' consequences on algorithmic fairness.*

Missingness patterns reflect disparities



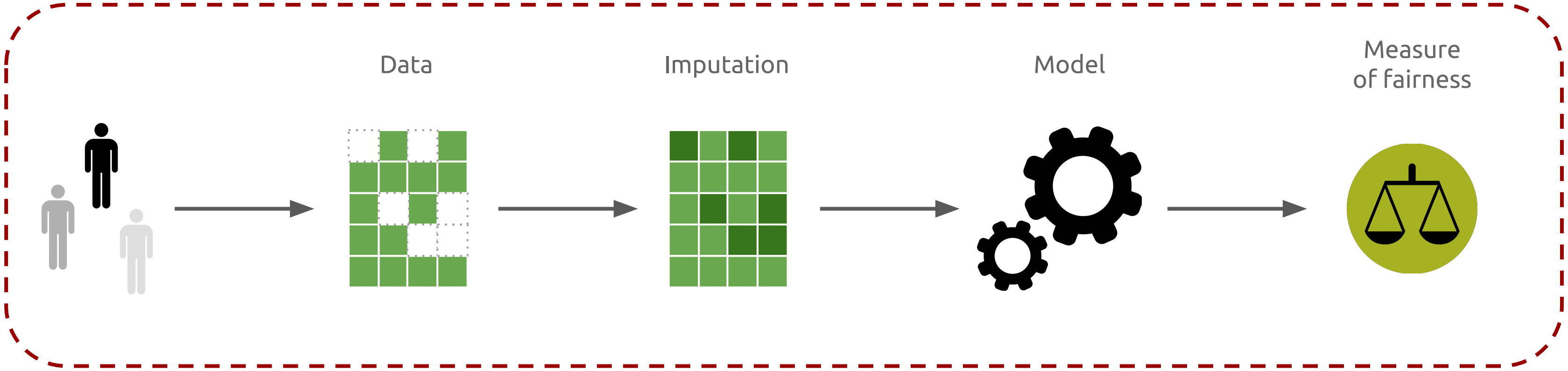
This paper focuses on ***biases in what is absent*** *from the data*

Imputation impacts algorithmic fairness



*How do **current imputation practices** impact algorithmic fairness ?*

Proposed path forward

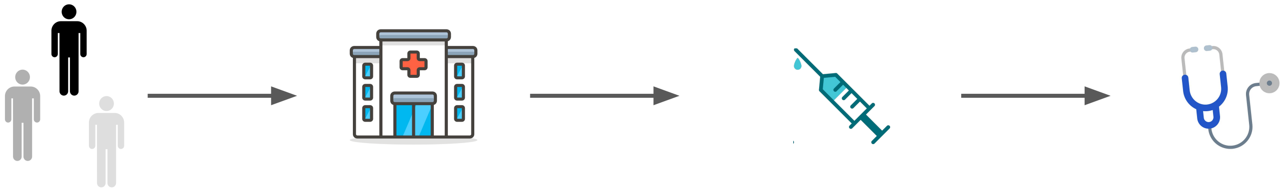


*We introduce a path forward to better **inform imputation choice** when concerned with algorithmic fairness*

Group-specific Missingness Patterns

Missingness can reflect disparities

(Mis)-informed collection
Concentration of missing data for patients who do not present "standard" symptoms that trigger questions or laboratory tests.

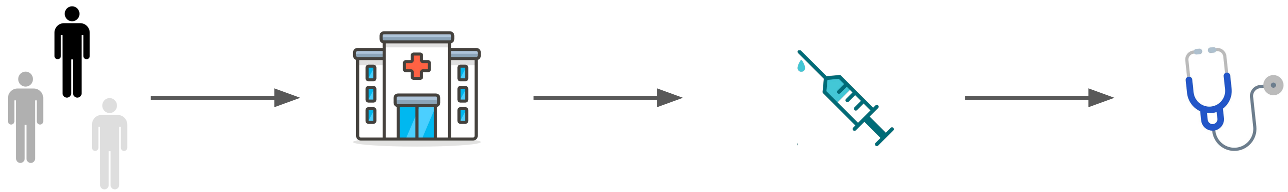


Limited access to quality care
Resulting in more missing data for some patients, e.g. as a result of less follow-up visits.

Confirmation bias
Information is only collected when a condition is expected.

Missingness can reflect disparities

(Mis)-informed collection
Concentration of missing data for patients who do not present "standard" symptoms that trigger questions or laboratory tests.

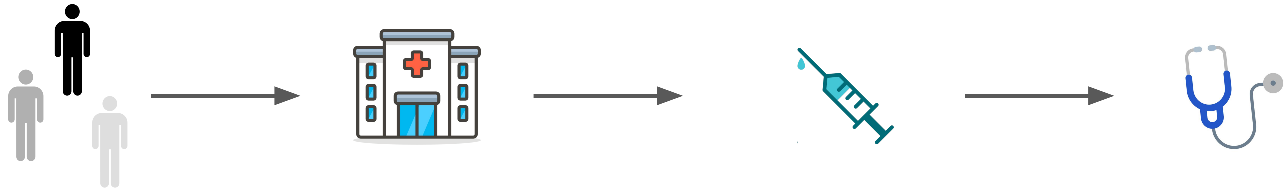


Limited access to quality care
Resulting in more missing data for some patients, e.g. as a result of less follow-up visits.

Confirmation bias
Information is only collected when a condition is expected.

Missingness can reflect disparities

(Mis)-informed collection
Concentration of missing data for patients who do not present “standard” symptoms that trigger questions or laboratory tests.

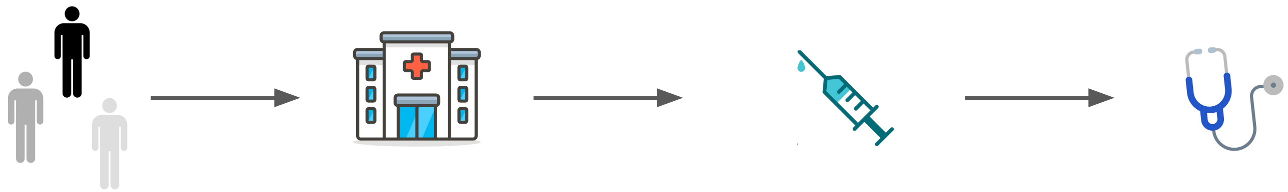


Limited access to quality care
Resulting in more missing data for some patients, e.g. as a result of less follow-up visits.

Confirmation bias
Information is only collected when a condition is expected.

Missingness can reflect disparities

(Mis)-informed collection
Concentration of missing data for patients who do not present "standard" symptoms that trigger questions or laboratory tests.



Limited access to quality care
Resulting in more missing data for some patients, e.g. as a result of less follow-up visits.

Confirmation bias
Information is only collected when a condition is expected.

*Traditional missingness dichotomisation does not capture the **group-specific nature** of medical missingness*

Current Imputation Practices

Current imputation practices

1. Aim to **minimise reconstruction error**

$$L^I = \mathbb{E}_x[||\tilde{x}^I - x||_2^2]$$

Current imputation practices

1. Aim to **minimise reconstruction error**
2. Rely on a **single imputation** based upon unrealistic missingness assumptions

Current imputation practices

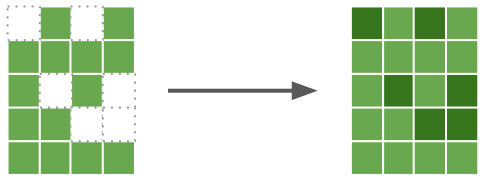
1. Aim to **minimise reconstruction error**
2. Rely on a **single imputation** based upon unrealistic missingness assumptions
3. When algorithmic fairness, encourage **group-specific imputation**

Fairness without imputation: A decision tree approach for fair prediction with missing values by Jeong et al., 2022

Current imputation practices

1. Aim to **minimise reconstruction error**
2. Rely on a **single imputation** based upon unrealistic missingness assumptions
3. When algorithmic fairness, encourage **group-specific imputation**

Population Mean



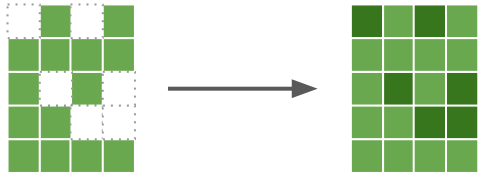
$\tilde{x} = \mathbb{E}\left[x\right]$

Fairness without imputation: A decision tree approach for fair prediction with missing values by Jeong et al., 2022

Current imputation practices

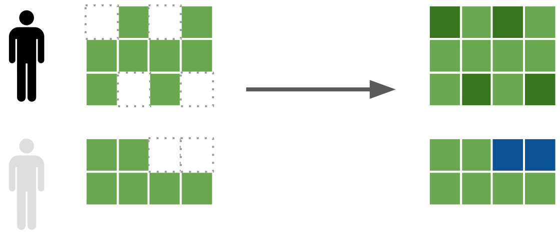
1. Aim to **minimise reconstruction error**
2. Rely on a **single imputation** based upon unrealistic missingness assumptions
3. When algorithmic fairness, encourage **group-specific imputation**

Population Mean



$$\tilde{x} = \mathbb{E}\left[x\right]$$

Group Mean



$$\tilde{x_g} = \mathbb{E}_{x|g}\left[x\right]$$

Fairness without imputation: A decision tree approach for fair prediction with missing values by Jeong et al., 2022

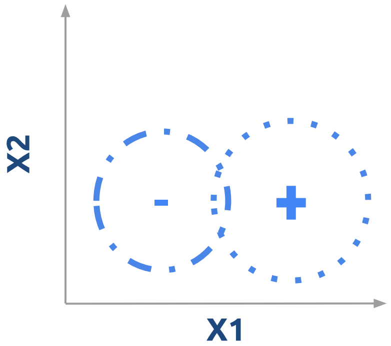
Empirical Comparison of Imputation Strategies

Simulations

Outcome **Group**

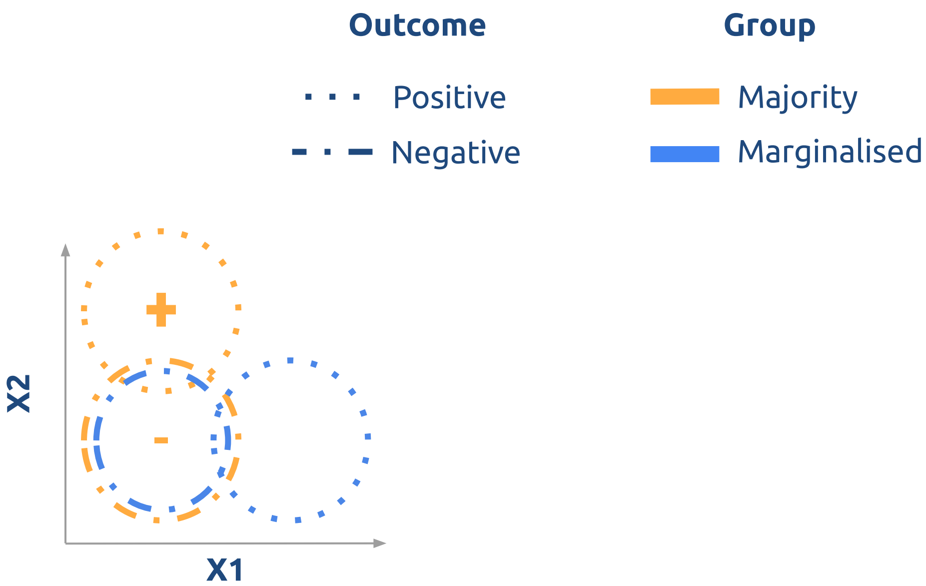
Positive

Negative Marginalised



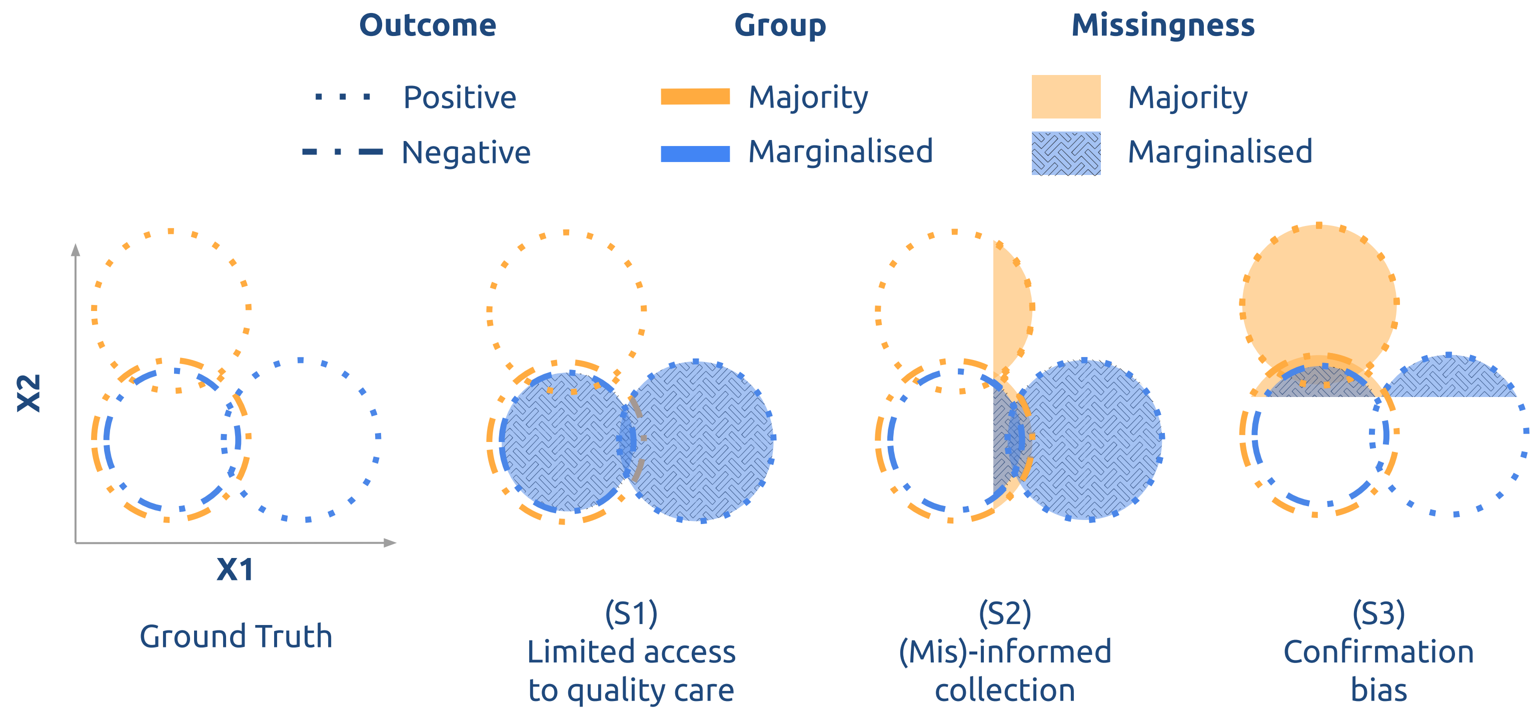
Ground Truth

Simulations



Ground Truth

Simulations



Pipeline

- **Single mean imputation** (Mean) - Missing data are replaced by the population mean.
- **Hot Deck** - Missing data replaced with closest patients' covariates.
- **Multiple Imputation using Chained Equation** (MICE) - Missing covariates are iteratively drawn from a regression model built over all other available covariates with median initialisation.
- **MICE Missing** - Missingness indicators are concatenated to the input data to leverage informative missingness.

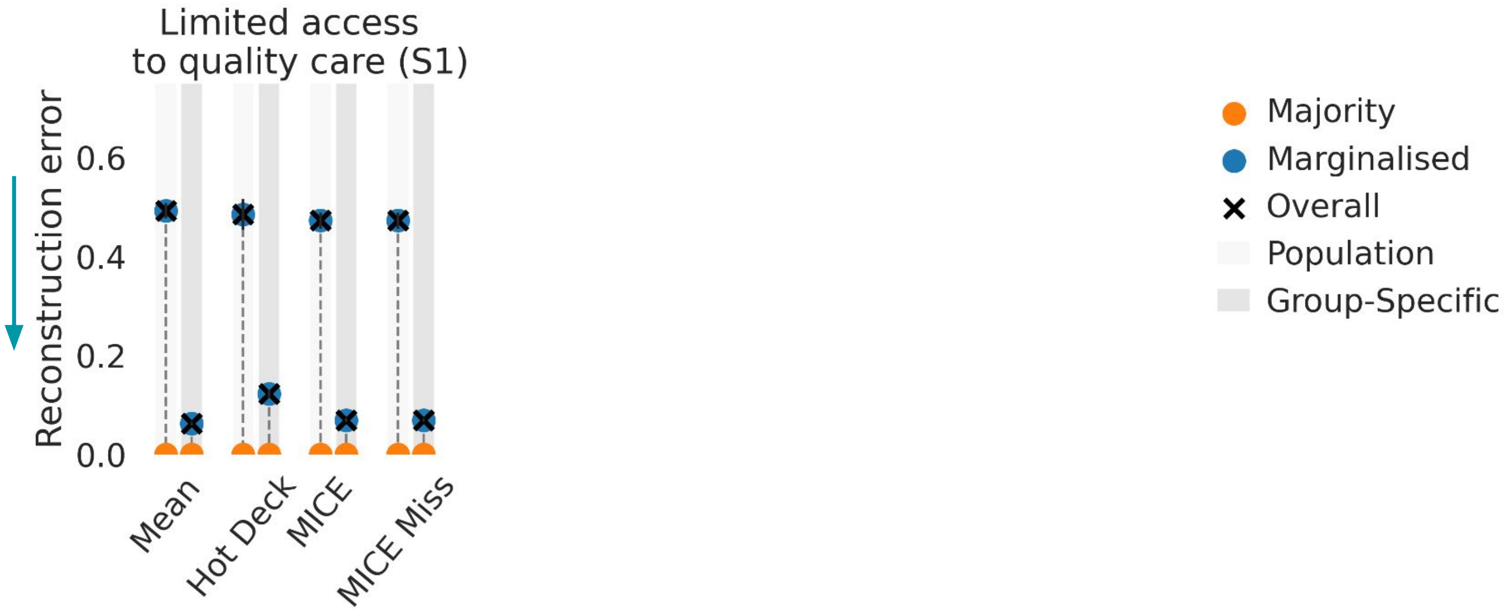
Pipeline

- **Single mean imputation** (Mean) - Missing data are replaced by the population mean.
- **Hot Deck** - Missing data replaced with closest patients' covariates.
- **Multiple Imputation using Chained Equation** (MICE) - Missing covariates are iteratively drawn from a regression model built over all other available covariates with median initialisation.
- **MICE Missing** - Missingness indicators are concatenated to the input data to leverage informative missingness.
- **Group Alternatives** - Group membership is added to render the MAR assumption more plausible.

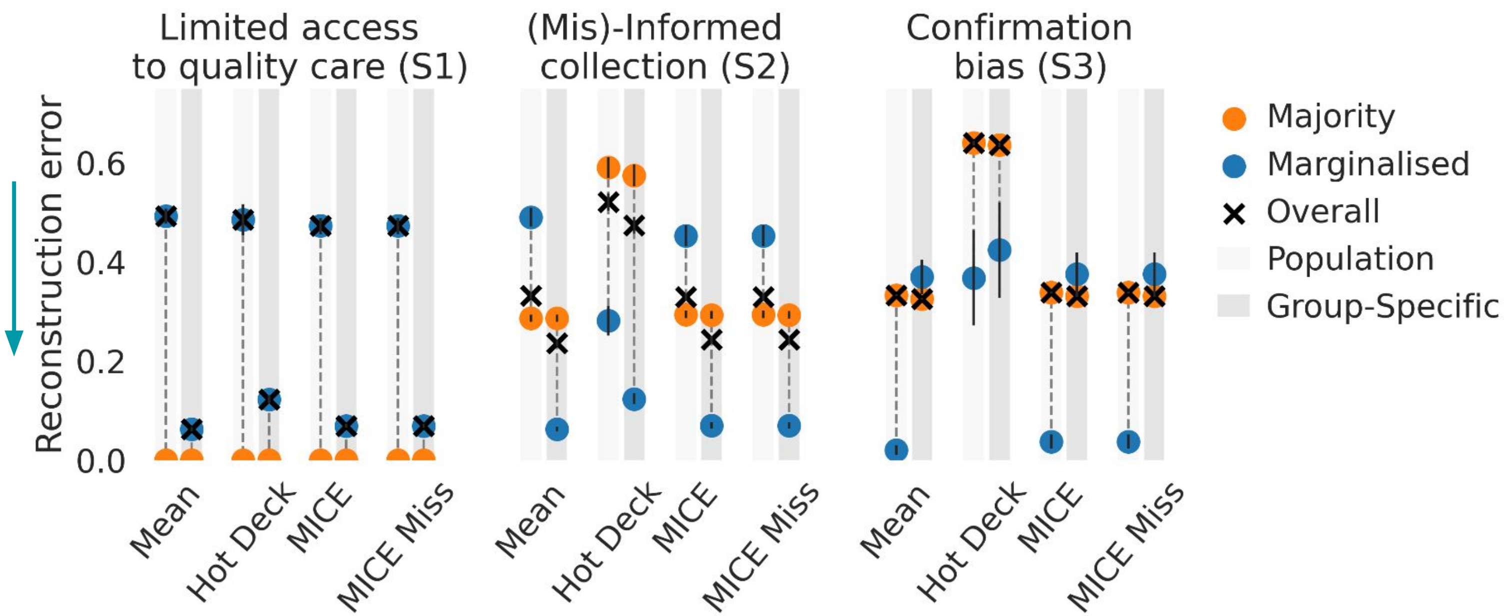
Reconstruction error

Reconstruction error

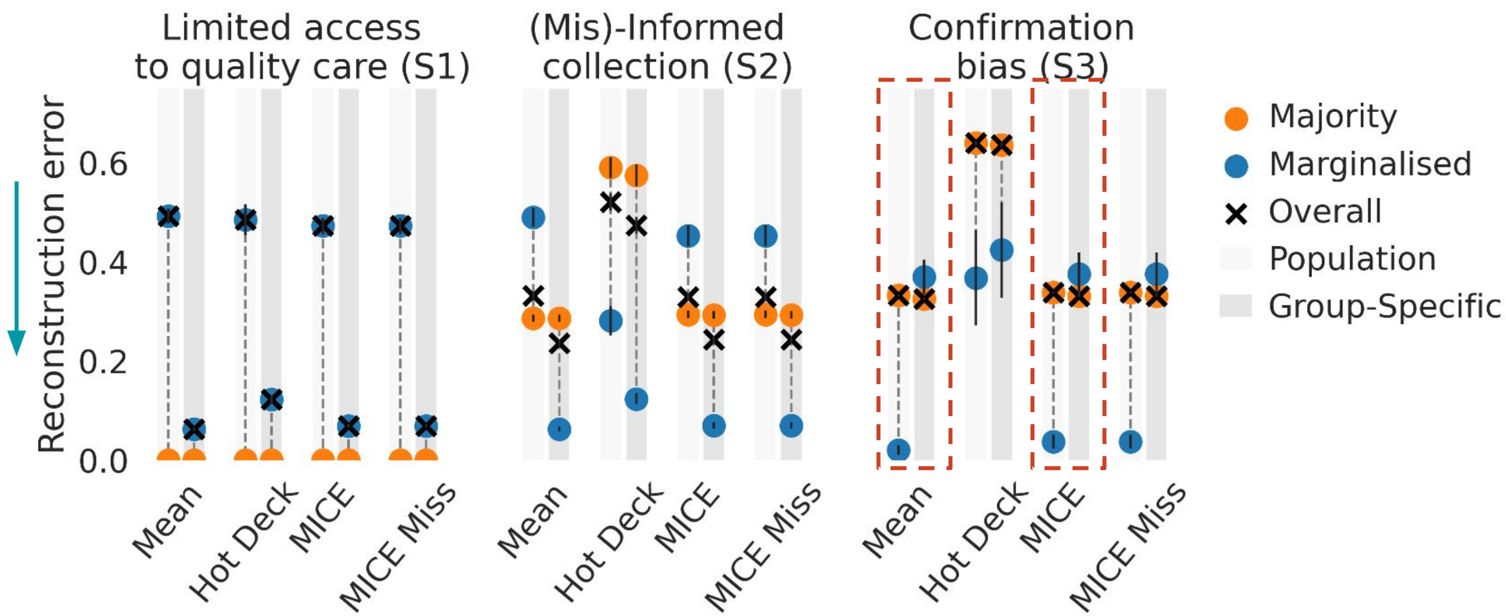
Reconstruction error



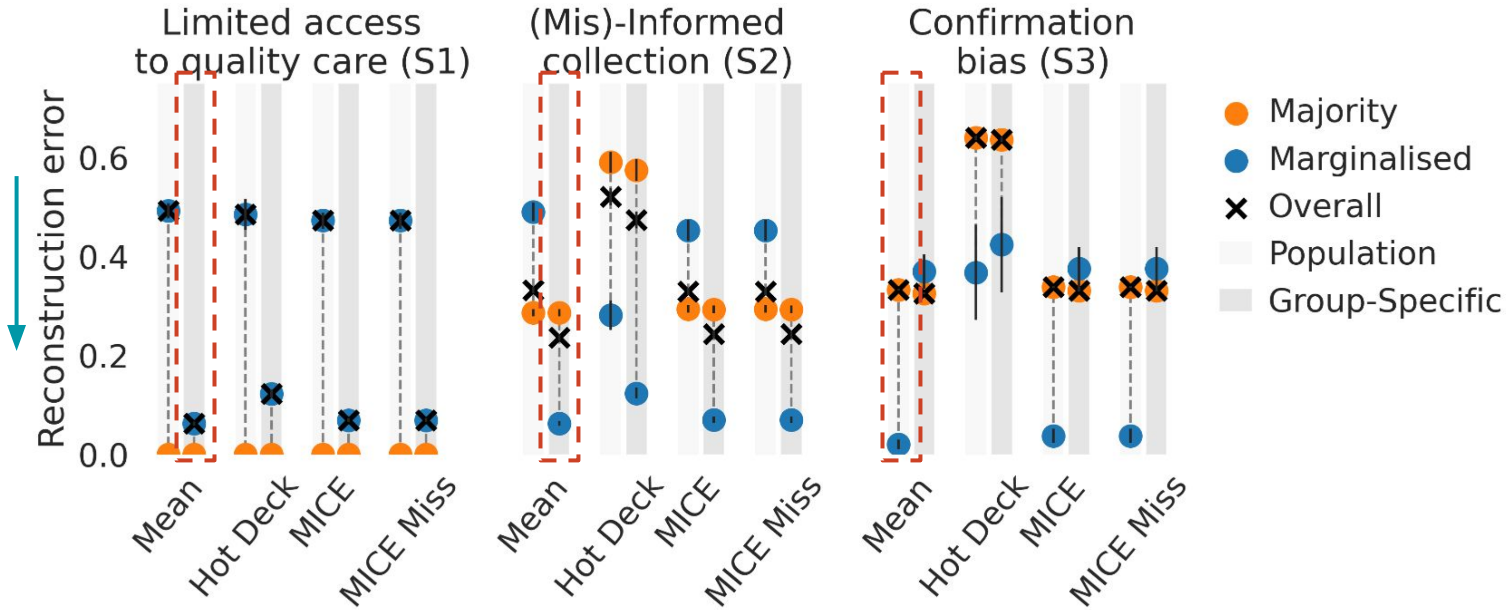
Reconstruction error



Group-imputation can lead to worse performance

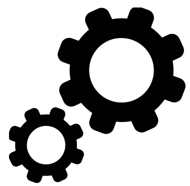


No imputation is best over all settings



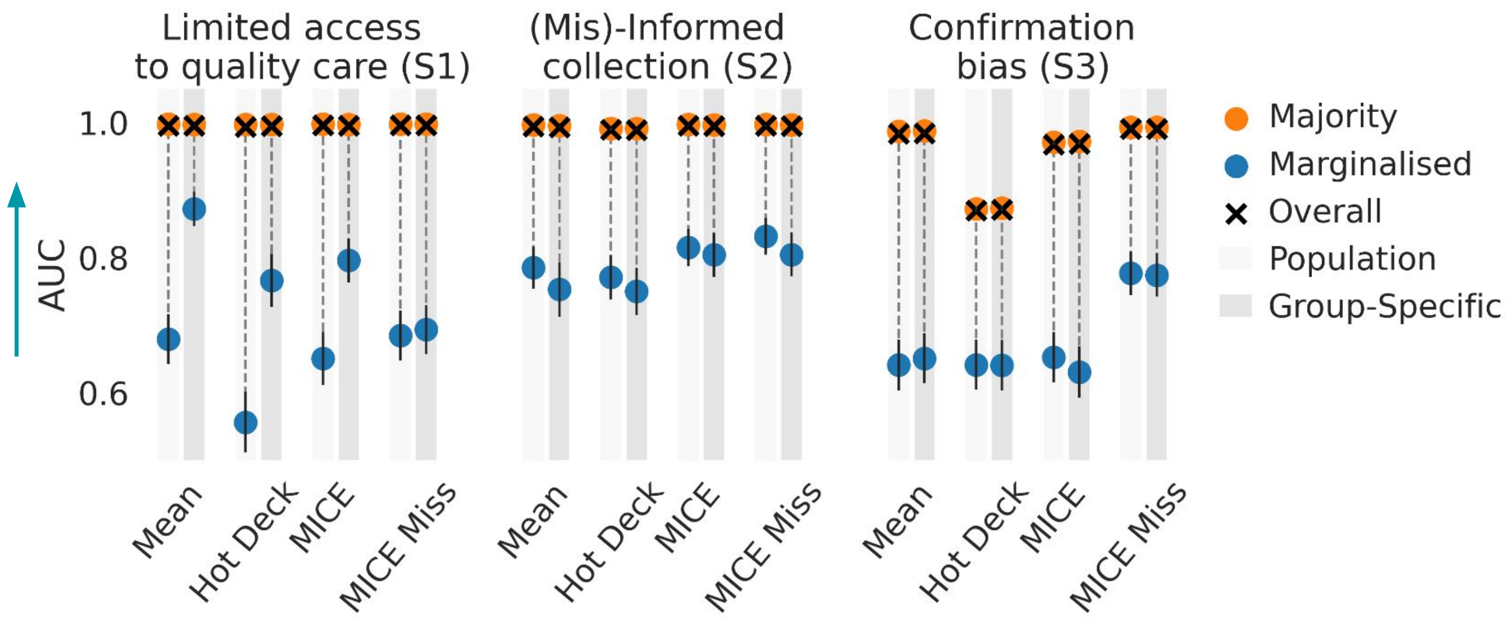
Pipeline

- ➔ **Single mean imputation** (Mean) - Missing data are replaced by the population mean.
- ➔ **Hot Deck** - Missing data replaced with closest patients' covariates.
- ➔ **Multiple Imputation using Chained Equation** (MICE) - Missing covariates are iteratively drawn from a regression model built over all other available covariates with median initialisation.
- ➔ **MICE Missing** - Missingness indicators are concatenated to the input data to leverage informative missingness.
- ➔ **Group Alternatives** - Group membership is added to render the MAR assumption more plausible.

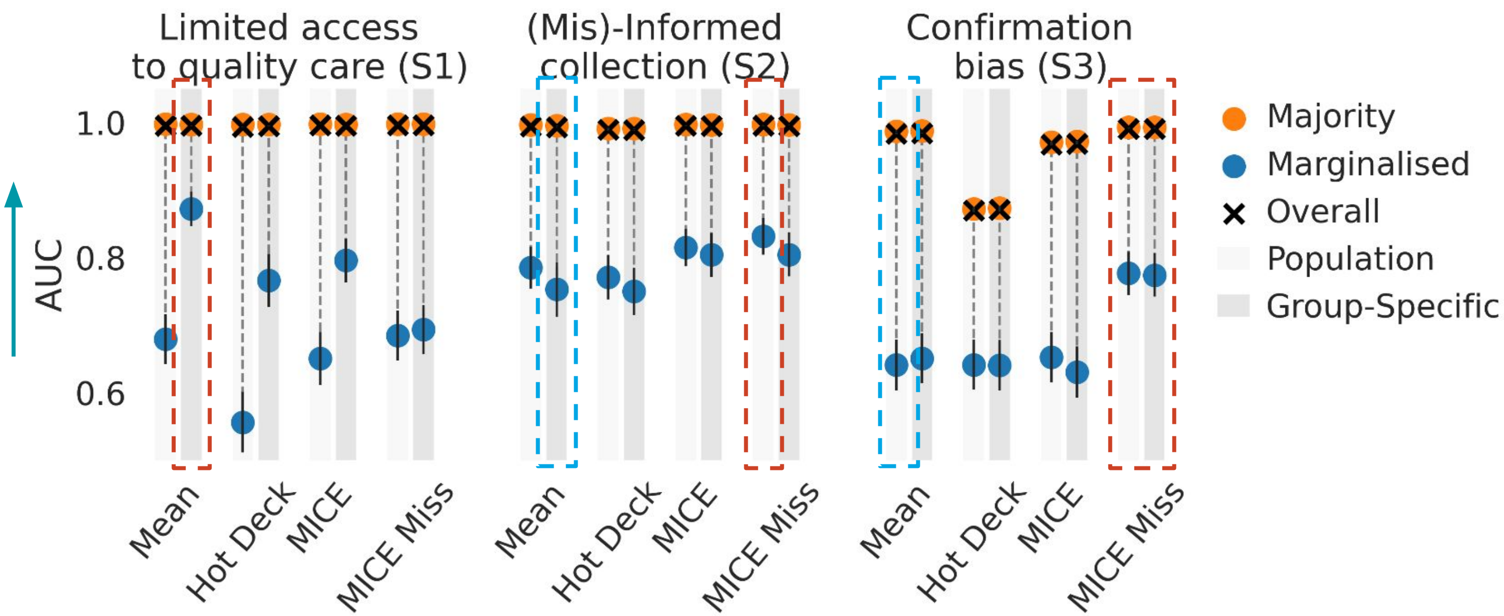


Logistic
Regression

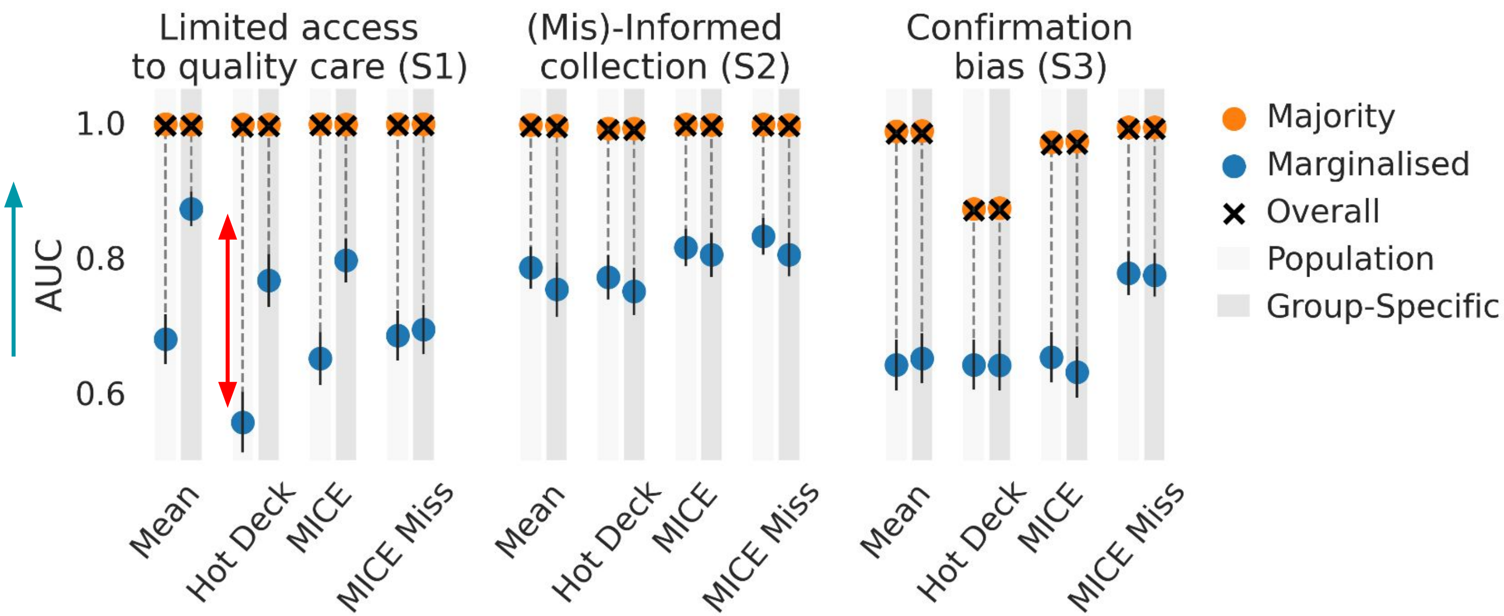
Downstream performance



No imputation is best over all settings



No imputation is best over all settings



Current practices are flawed

Current practices

1. Aim to **minimise reconstruction error**
2. Rely on a **single imputation** based upon unrealistic missingness assumptions
3. When algorithmic fairness, encourage **group-specific imputation**

Counter arguments

1. **Impossible** to measure reconstruction error and **disconnected from downstream** algorithmic fairness

Current practices are flawed

Current practices

1. Aim to **minimise reconstruction error**
2. Rely on a **single imputation** based upon unrealistic missingness assumptions
3. When algorithmic fairness, encourage **group-specific imputation**

Counter arguments

1. **Impossible** to measure reconstruction error and **disconnected from downstream** algorithmic fairness
2. **No best** imputation strategies over all missingness patterns

Current practices are flawed

Current practices

1. Aim to **minimise reconstruction error**
2. Rely on a **single imputation** based upon unrealistic missingness assumptions
3. When algorithmic fairness, encourage **group-specific imputation**

Counter arguments

1. **Impossible** to measure reconstruction error and **disconnected from downstream** algorithmic fairness
2. **No best** imputation strategies over all missingness patterns
3. Group-specific imputation can **increase** reconstruction error and downstream gaps

Current practices are flawed

Current practices

1. Aim to **minimise reconstruction error**
2. Rely on a **single imputation** based upon unrealistic missingness assumptions
3. When algorithmic fairness, encourage **group-specific imputation**

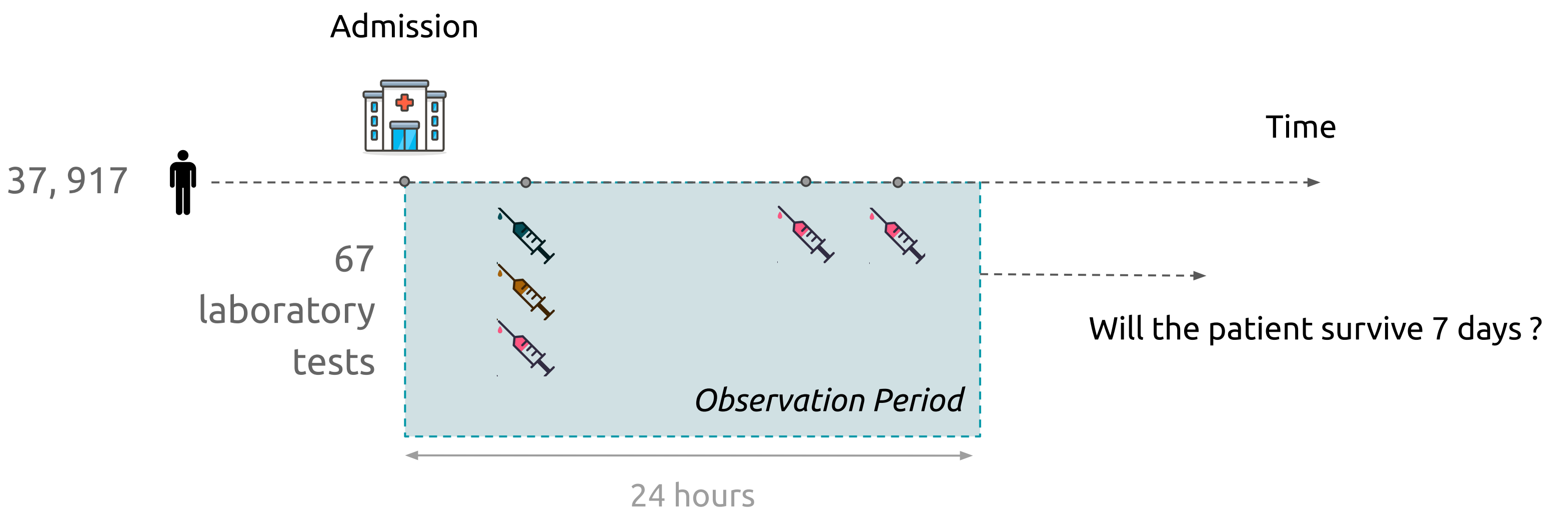
Counter arguments

1. **Impossible** to measure reconstruction error and **disconnected from downstream** algorithmic fairness
2. **No best** imputation strategies over all missingness patterns
3. Group-specific imputation can **increase** reconstruction error and downstream gaps

Practitioners in healthcare must change their imputation practices

Informing Imputation Choice in a Case-Study

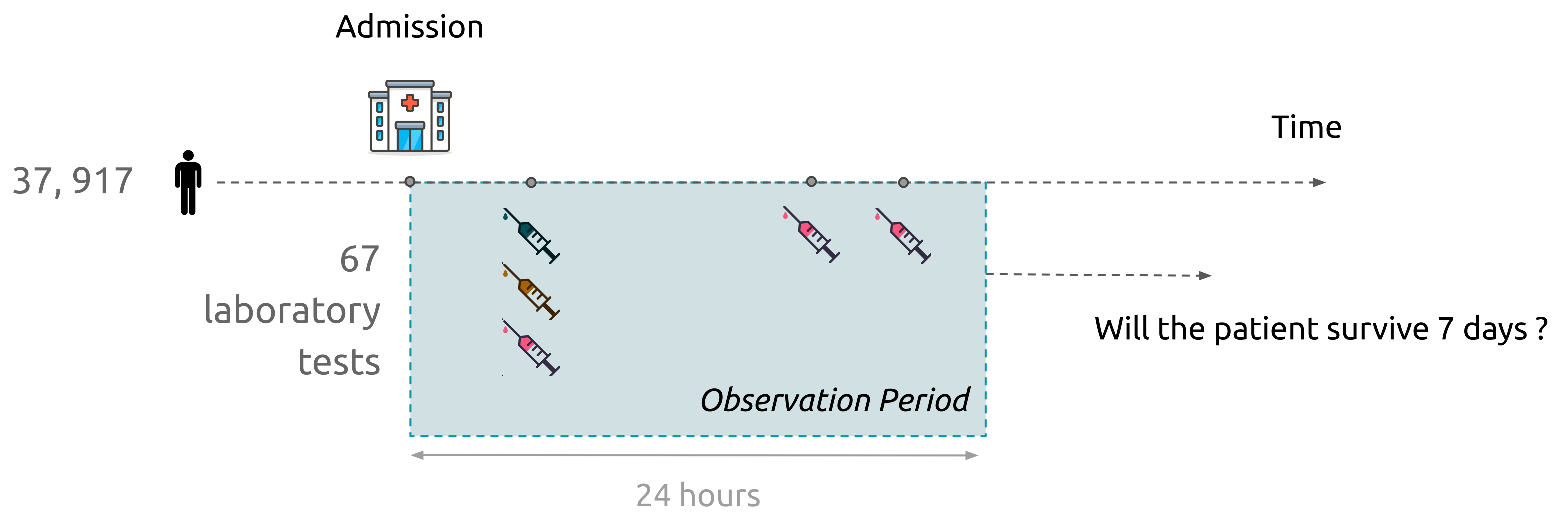
Building a predictive model on MIMIC III



MIMIC-III, a freely accessible critical care database by A Johnson & al.

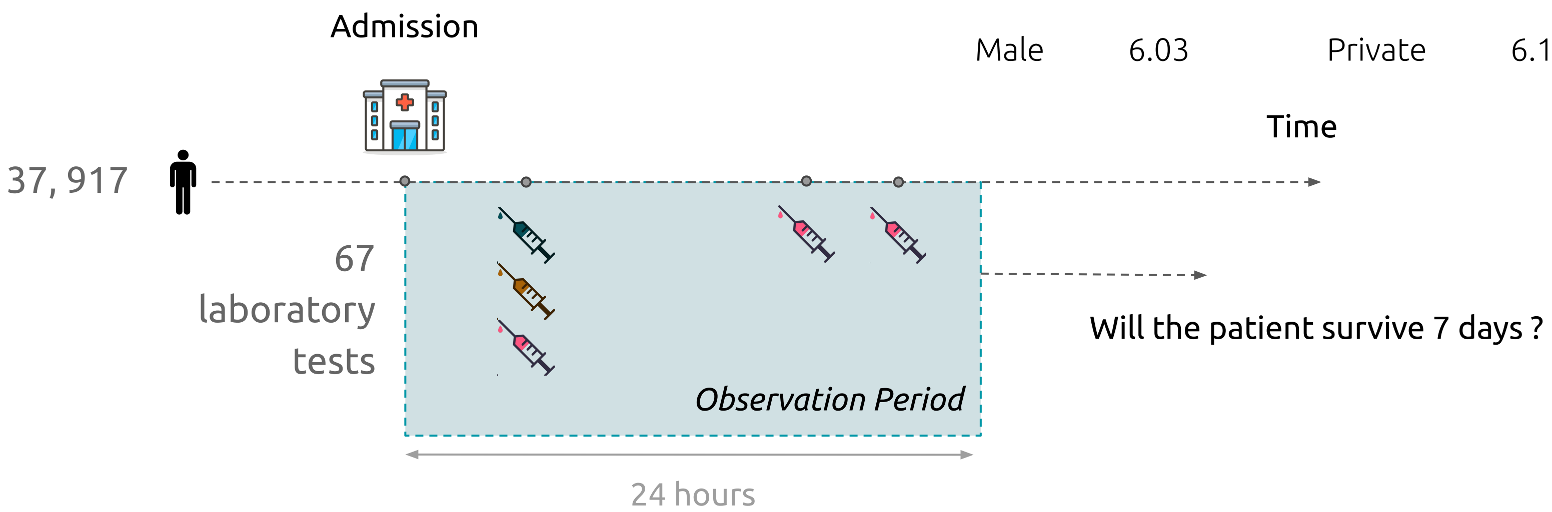
Patterns of observation

	Orders
Alive	**5.68**
Dead	7.57



Patterns of observation

	Orders		Orders
Alive	**5.68**	Black	**5.24**
Dead	7.57	Other	5.86
Female	**5.54**	Public	**5.67**
Male	6.03	Private	6.11



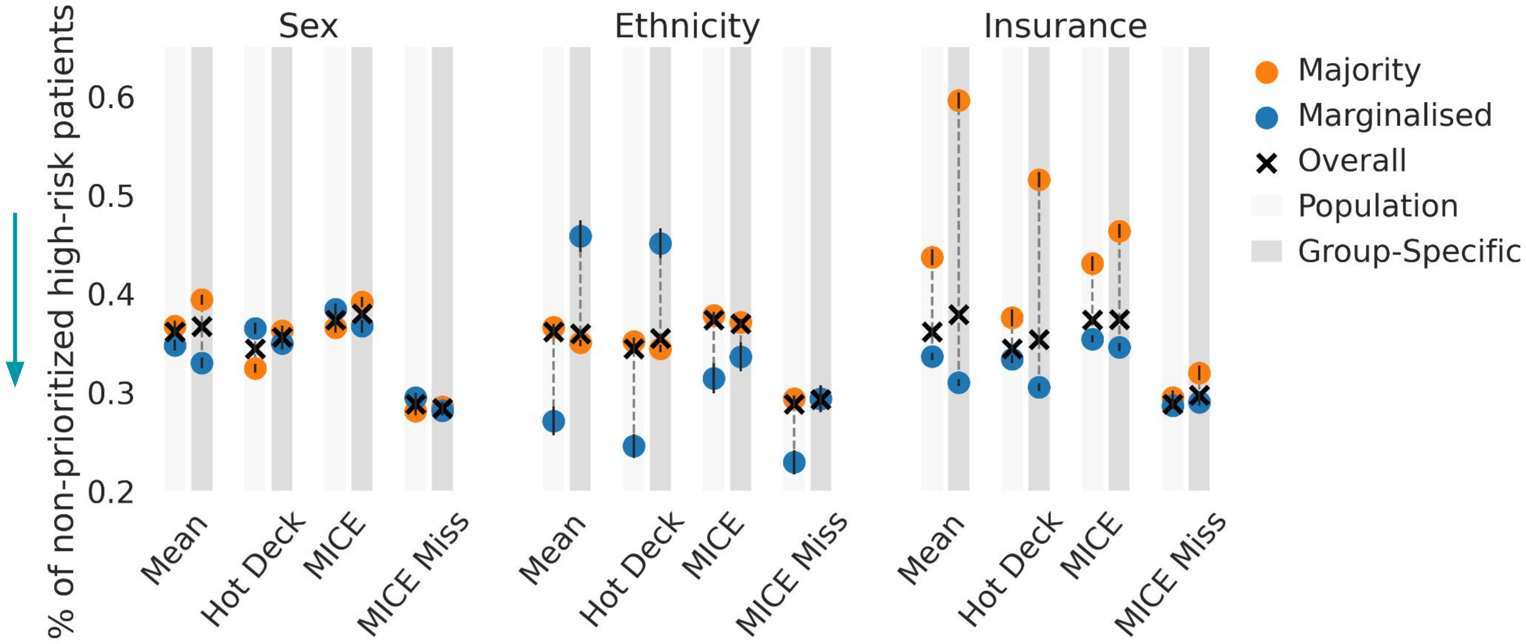
MIMIC-III, a freely accessible critical care database by A Johnson & al.

Framework

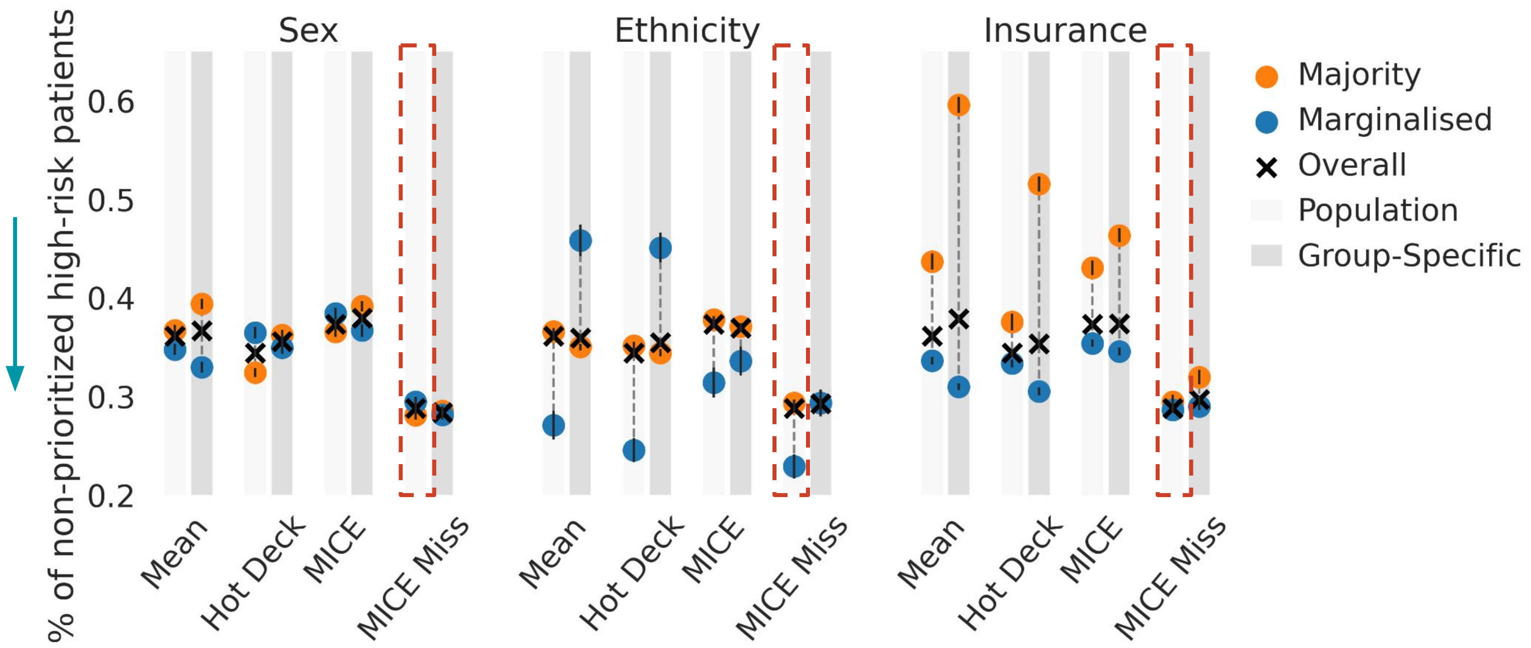
1. **Identify** imputation strategies
2. **Measure** impact on downstream performances and algorithmic fairness
3. **Select** imputation considering trade-off

Informing imputation choice

Informing imputation choice

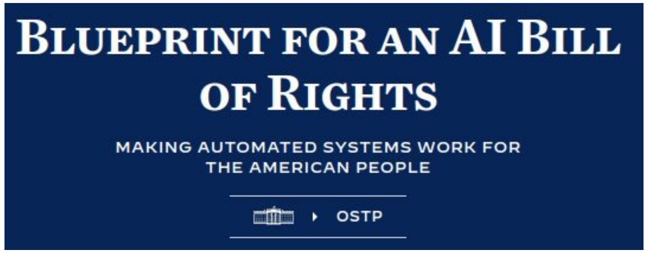


Informing imputation choice



Framework

1. **Identify** imputation strategies
2. **Measure** impact on downstream performances and algorithmic fairness
3. **Select** imputation considering trade-off



Framework

1. **Identify** imputation strategies
2. **Measure** impact on downstream performances and algorithmic fairness
3. **Select** imputation considering trade-off
4. **Report**

- Factors:
  - ☐ Marginalised groups
  - ☐ Environment
- Missingness process:
  - ☐ Known mechanisms
  - ☐ Potential influences
- Descriptive statistics
- Considered pipelines:
  - ☐ Imputation strategies
  - ☐ Models
- Metrics
- Quantitative results
- Caveats and recommendations

Imputation Cards

Short-term survival prediction

Factors:
- Marginalised groups: Sex (43.3% female), ethnicity (Black (7.7%) vs non-Black) and insurance (public (66.6%) vs rest).
- Environment: All data from Intensive Care Units (ICU) in a teaching hospital in Boston, Massachusetts, USA.

Missingness process:
- Known mechanisms: Standardised procedures following ICU guidelines.
- Potential influences: Experts' intuition and potential biases from limited access to care, historical biases and confirmation biases.

Descriptive statistics:
- Range of missingness rate at the end of 24 hours of observation across the different covariates [1.90 - 99.98] with average 60.53%. The following table presents the results stratified for the different groups.
- Percentage of patients with more than 50% of tests observed: 85.56%

Groups	Marginalised	Majority
Sex	60.82 [2.15 - 99.98]	60.31 [1.70 - 99.99]
Ethnicity	61.10 [2.62 - 100]	60.48 [1.83 - 99.98]
Insurance	60.93 [1.88 - 99.98]	60.81 [1.93 - 99.99]

Missingness percentage (mean [min - max]) stratified per groups.

Considered pipelines:
- Imputation: Mean imputation, Hot Deck, MICE, MICE Missing (using a missingness indicator as input to the model) and their group-specific alternatives.
- Modelling: Logistic regression with l2 penalty on the imputed data.

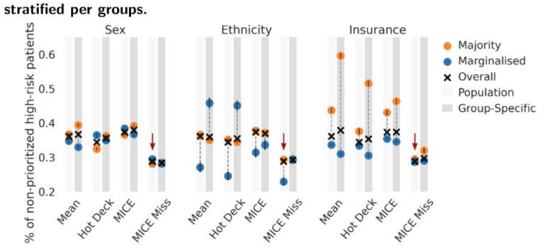
Metrics: Use of False Negative Rate at 30% (current threshold of prioritisation) to reflect the percentage of patients that would not be prioritised despite being at risk, both at the population level and stratified by groups.

Quantitative results: The following figure describes the performance stratified by groups. Overall performance ranges from 28.4% to 38.0% FNR highlighting a large impact of imputation on performance.

Groups	Gap Range	Best
Sex	[-6.42 - 1.87]	-3.73
Ethnicity	[-9.52 - 10.83]	0.01
Insurance	[-28.6 - -0.81]	-0.81

Range FNR performance gaps (in percent) stratified per group.

Caveats and recommendations: Assuming a stable missingness process and population distribution at deployment, MICE with missingness indicator minimises the number of patients missed across and within each group.

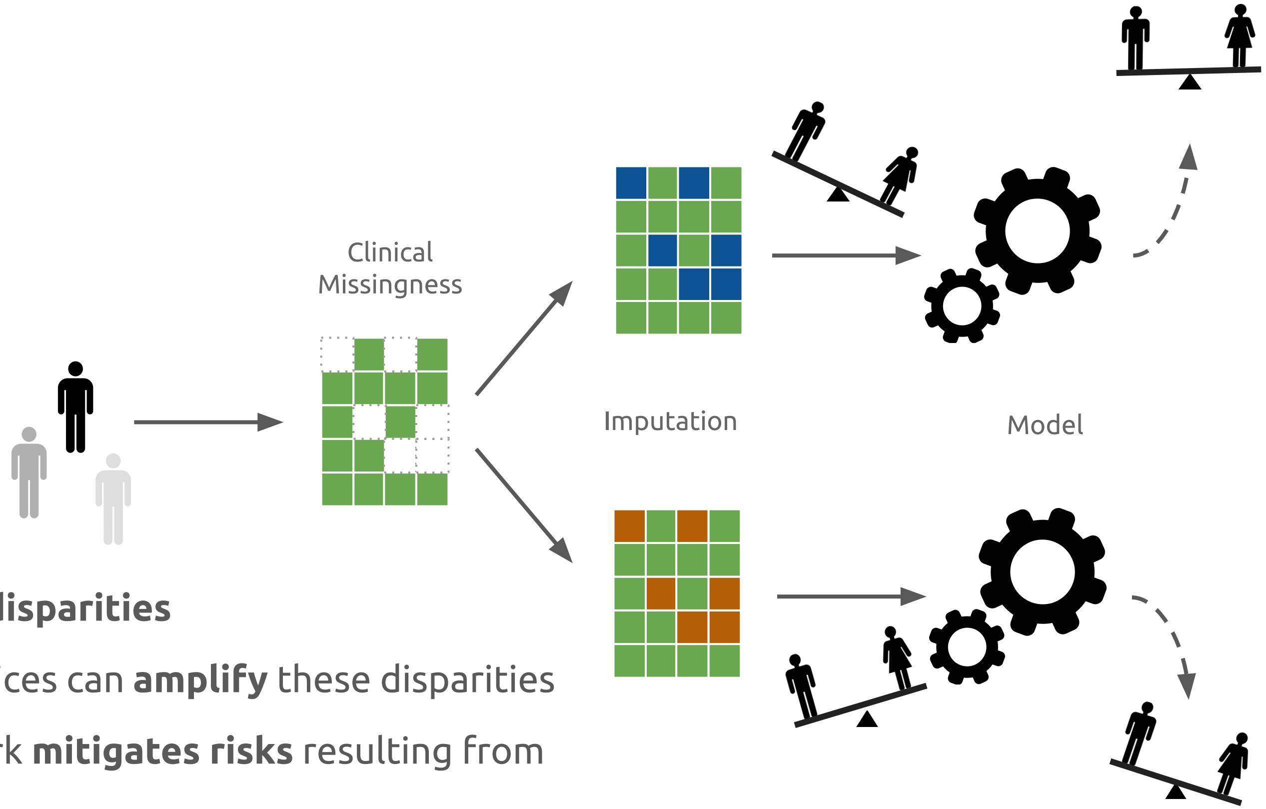


Model performance stratified per group and imputation strategies.

Figure 9 Imputation card for short-term prediction in the MIMIC dataset.

UNIVERSITY OF CAMBRIDGE

Conclusion



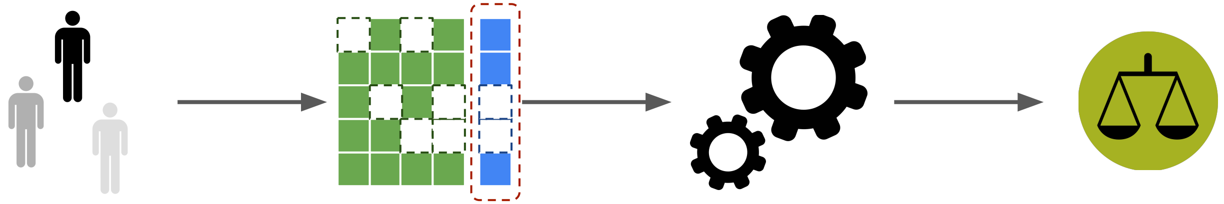
1. **Missingness** can reflect **disparities**
2. Current imputation practices can **amplify** these disparities
3. The introduced framework **mitigates risks** resulting from imputation

Jeanselme, V., De-Arteaga, M., Zhang, Z., Barrett, J., & Tom, B. (2022). *Imputation Strategies Under Clinical Presence: Impact on Algorithmic Fairness.* In Machine Learning for Health (pp. 12-34). PMLR. - Reject and resubmit at Management Science (2nd round)

Ignoring Competing Risks: Impact on Algorithmic Fairness

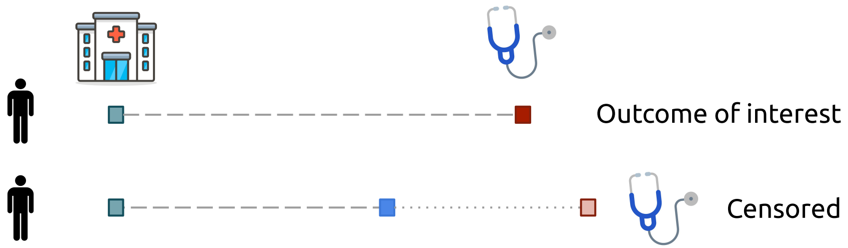
V. Jeanselme, C. Yoon, J. Barrett and B. Tom

Clinical presence concerns more than covariates

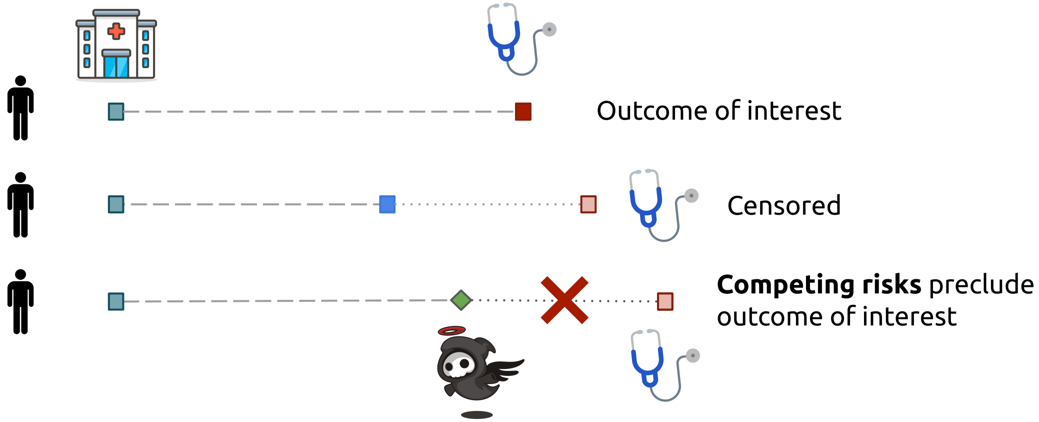


Outcomes are not always observed

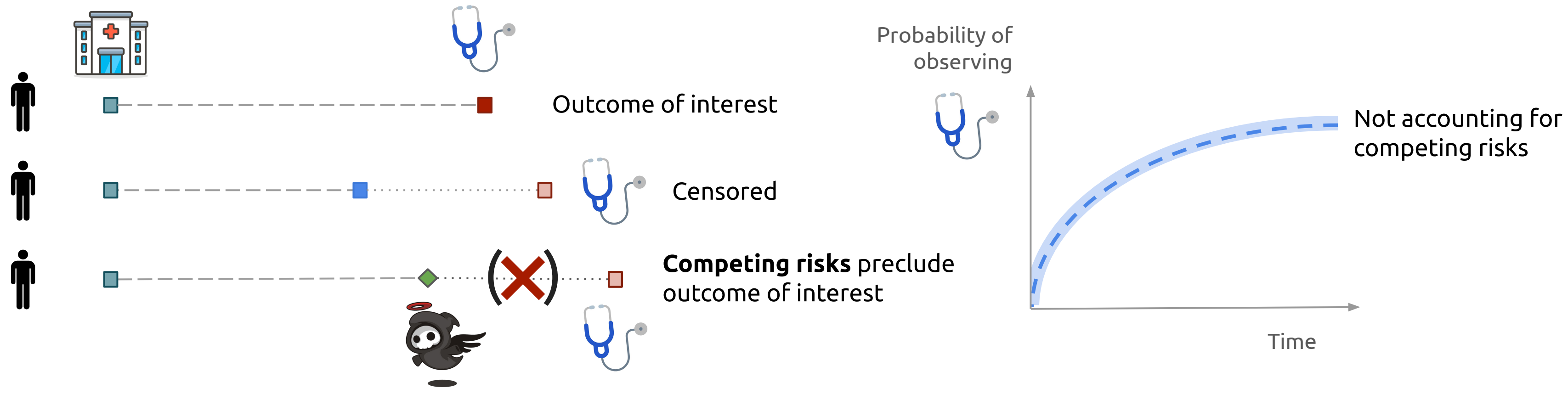
Outcomes are not always observed



Competing risks preclude the outcome of interest



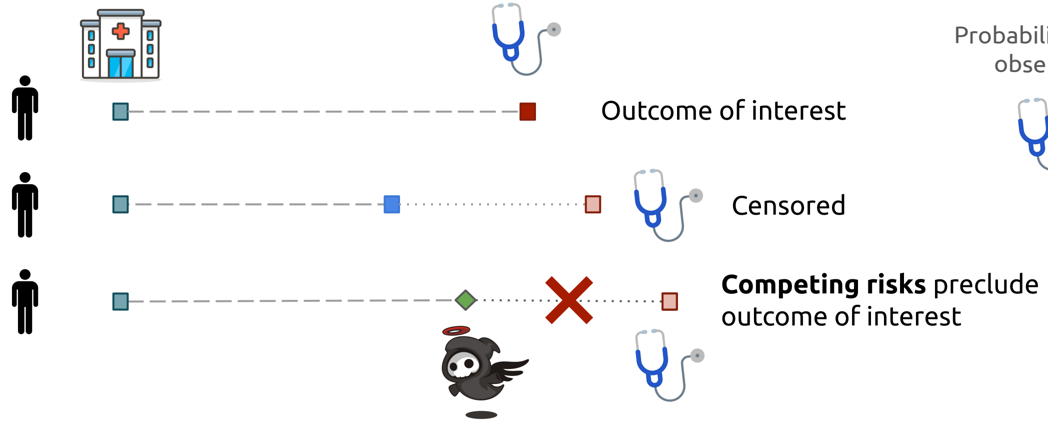
Considering competing risks as censoring

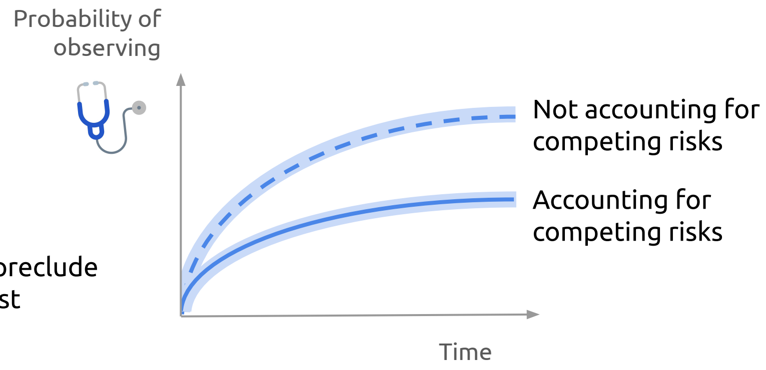


50% of studies do not account for competing risks

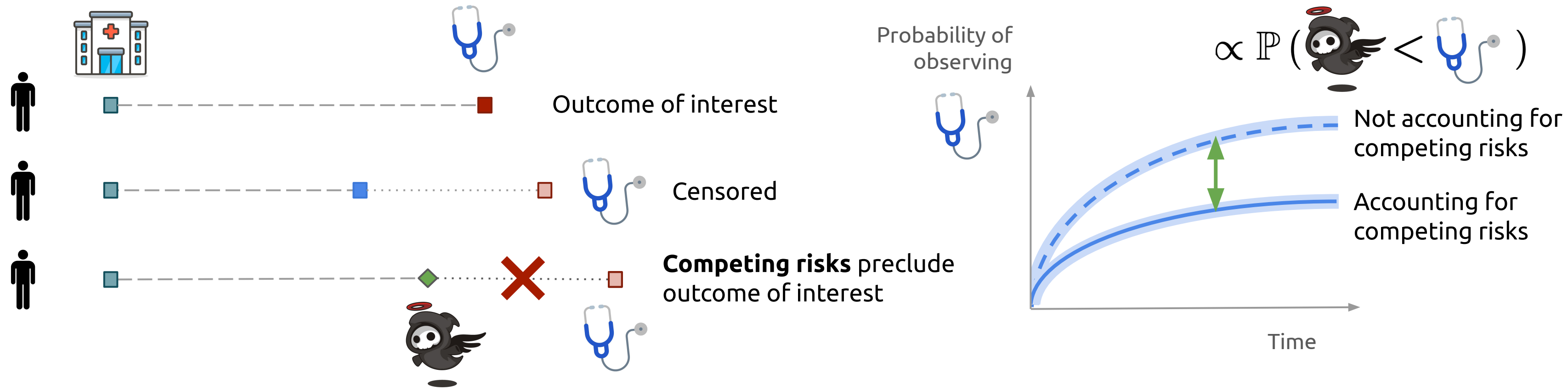
Competing risks and the clinical community: irrelevance or ignorance? by Koller et al., 2012

This practice biases estimates

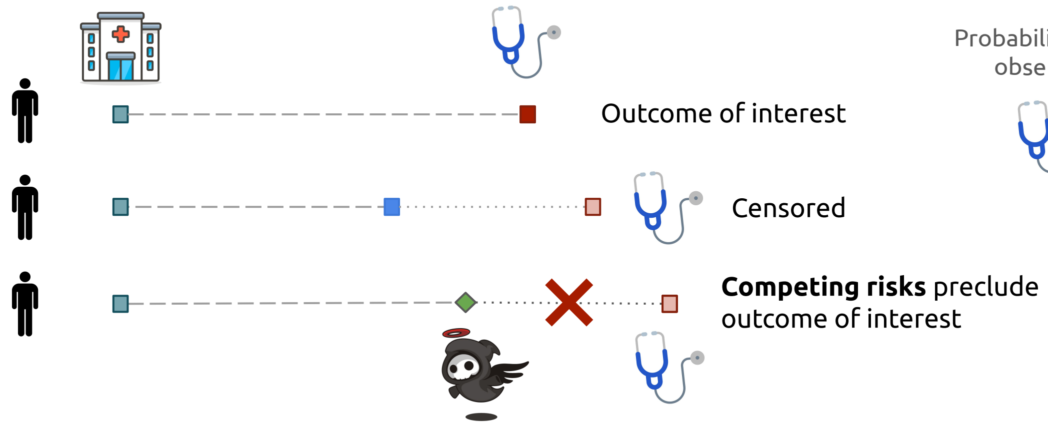


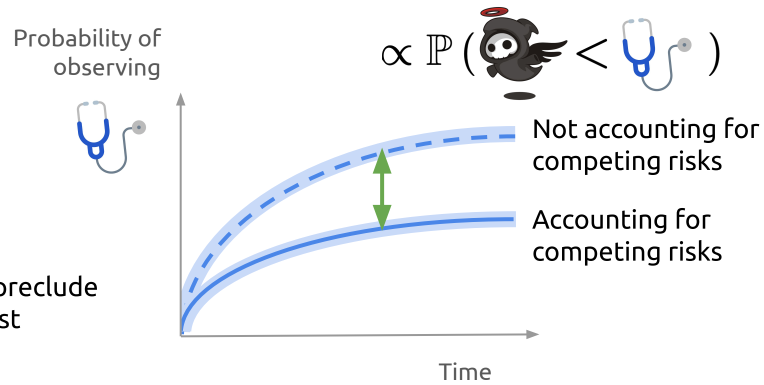


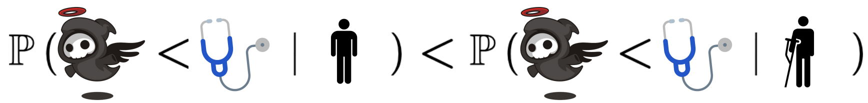
This practice biases estimates



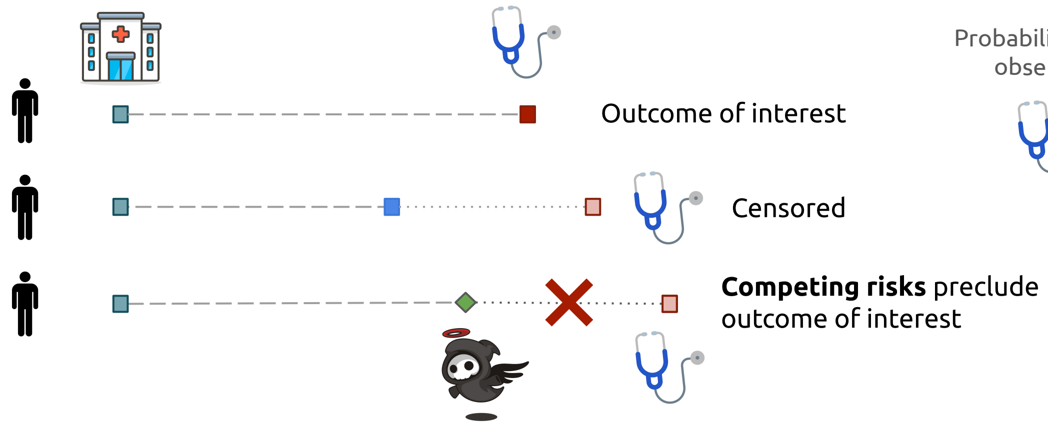
Different groups may not present the same risk

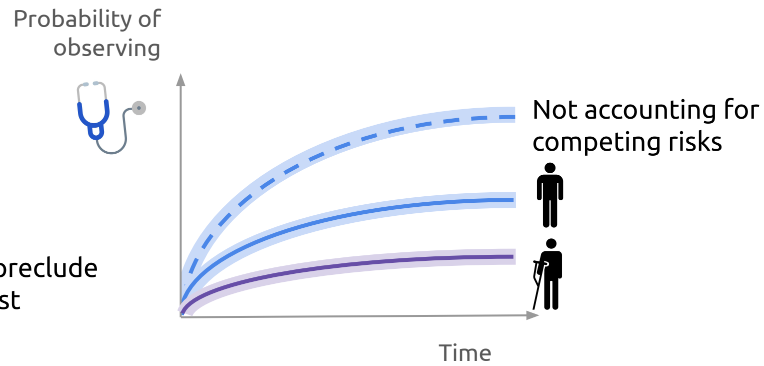


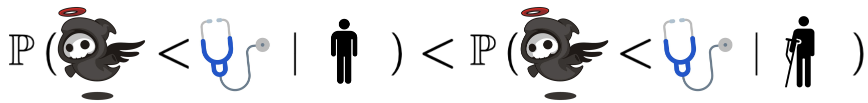




Different groups are impacted differently

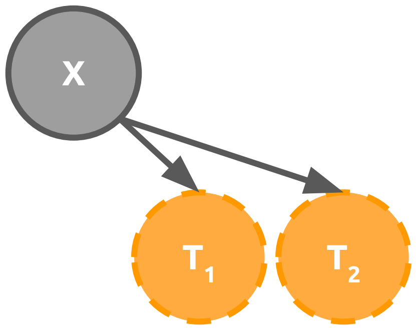




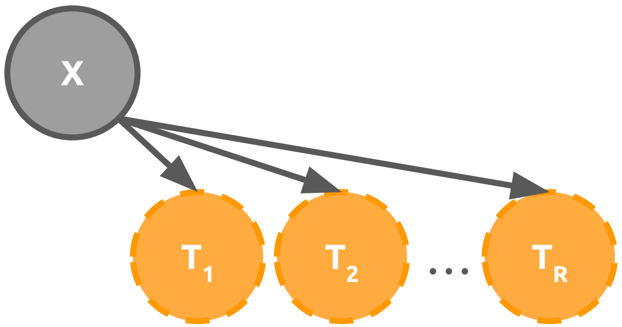


Quantifying the error associated with current practice

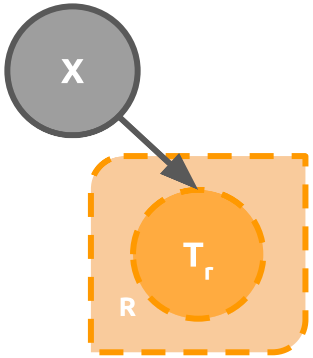
Modelling competing risks



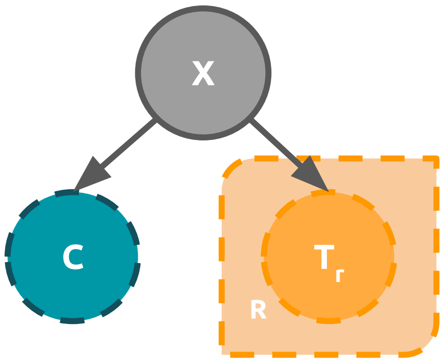
Modelling competing risks



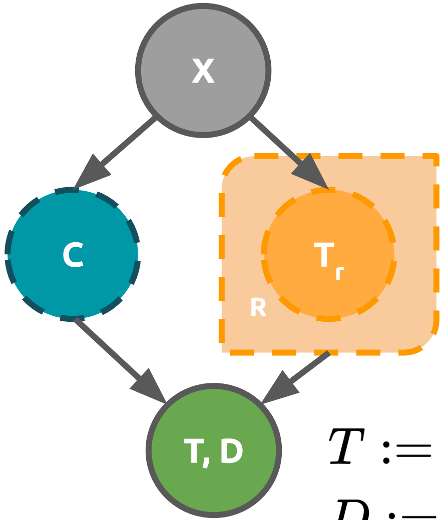
Modelling competing risks



Modelling competing risks

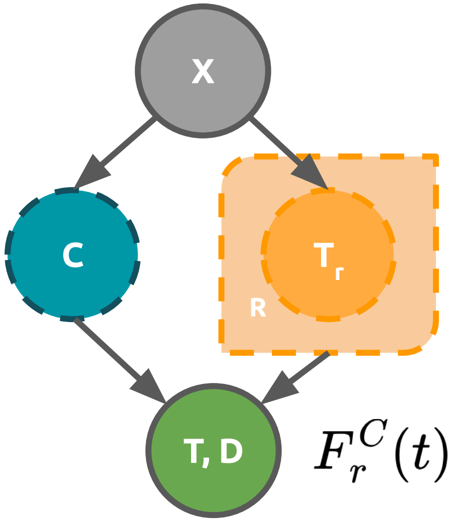


Modelling competing risks



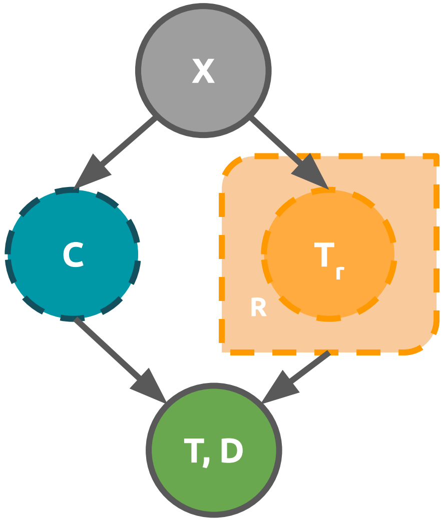
$$T := \min\left(C,\, T_1, \ldots, T_R\right)$$
$$D := \arg\min\left(C,\, T_1, \ldots, T_R\right)$$

Modelling competing risks

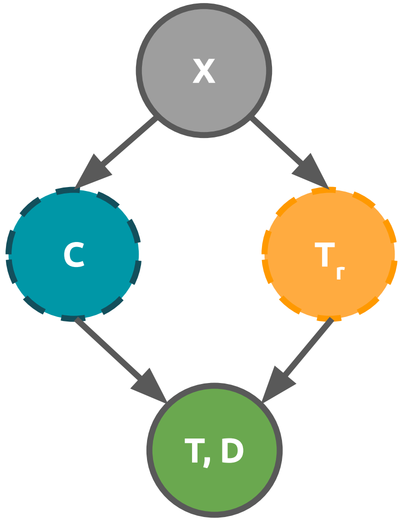


$$F_r^C(t) := \mathbb{P}(\underline{T_r < t}, \underline{\forall s\, T_r < T_s})$$

Quantifying the error between the two



$$F_r^C(t) := \mathbb{P}(\underline{T_r < t}, \forall s\, T_r < T_s)$$



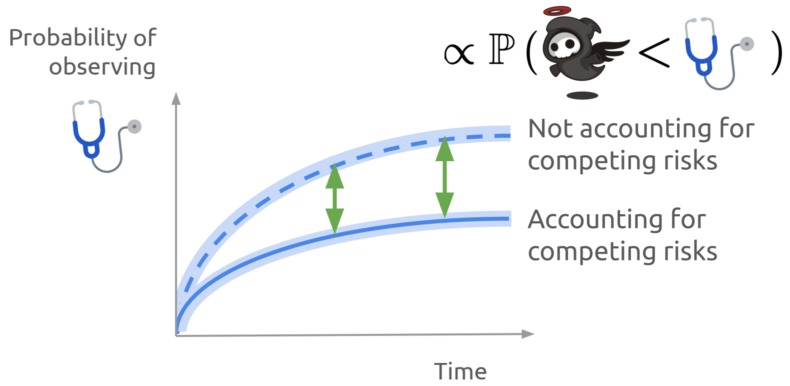
$$F_r^{NC}(t) := \mathbb{P}(\underline{T_r < t})$$

Relative cumulative incidence discrepancy

$$L^r(t,x) := \frac{F_r^{NC}(t \mid x) - F_r^{C}(t \mid x)}{\max(F_r^{NC}(t \mid x), F_r^{C}(t \mid x))}$$

Relative cumulative incidence discrepancy

$$L^r(t, x) := \frac{F_r^{NC}(t \mid x) - F_r^{C}(t \mid x)}{\max(F_r^{NC}(t \mid x), F_r^{C}(t \mid x))}$$
$$= \mathbb{P}(\exists s, T_s < T_r \mid x)$$



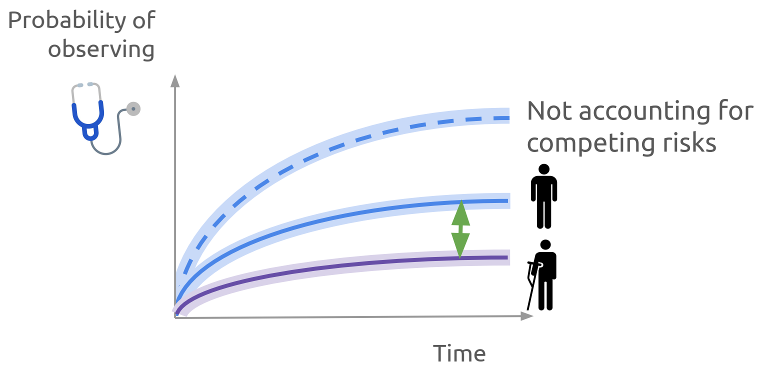
Inter-group discrepancy

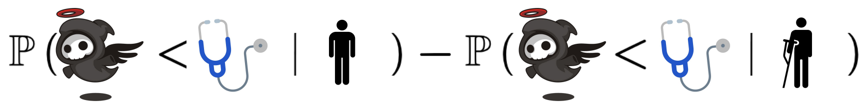
$$\Delta_g^r := \mathbb{E}_{x|g}\left[L^r(x)\right] - \mathbb{E}_{x|\neq g}\left[L^r(x)\right]$$

Does modelling competing risks as censoring have **algorithmic fairness consequences** ?

Different groups are impacted differently

$$\Delta_g^r := \mathbb{E}_{x|g}\left[L^r(x)\right] - \mathbb{E}_{x|\neq g}\left[L^r(x)\right]$$
$$= \mathbb{P}(\exists s,\, T_s < T_r \mid g) - \mathbb{P}(\exists s,\, T_s < T_r \mid \neg g)$$





Modelling competing risks

Modelling competing risks

One is interested in estimating the **cumulative incidence function**:

$$F_r^C(t \mid x) := \mathbb{P}(T_r < t, \forall s\, T_r < T_s \mid x)$$

Challenges in modelling competing risks

One is interested in estimating the **cumulative incidence function**:

$$F_r^C(t \mid x) := \mathbb{P}(T_r < t, \forall s\, T_r < T_s \mid x)$$

Often by maximising the associated likelihood of observed outcomes:

$$l := \sum_r \sum_{i,\, d_i = r} \log \underline{\frac{\partial F_r^C(t \mid x_i)}{\partial t}\Big|_{t=t_i}} + \sum_{i,\, d_i = 0} \log \Big[1 - \sum_r \underline{F_r^C(t_i \mid x_i)}\Big]$$

Observed Events Censored

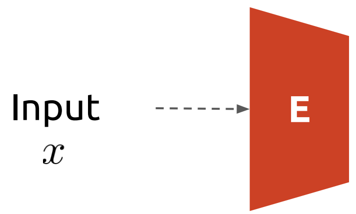
Traditional approximations

$$l := \sum_r \sum_{i,\, d_i = r} \log \underline{\frac{\partial F_r^C(t \mid x_i)}{\partial t}}\Big|_{t=t_i} + \sum_{i,\, d_i = 0} \log \left[1 - \sum_r \underline{F_r^C(t_i \mid x_i)}\right]$$

Proposed approach

$$l := \sum_r \sum_{i,\, d_i = r} \log \frac{\partial F_r^C(t \mid x_i)}{\partial t}\bigg|_{t=t_i} + \sum_{i,\, d_i = 0} \log \left[1 - \sum_r F_r^C(t_i \mid x_i)\right]$$

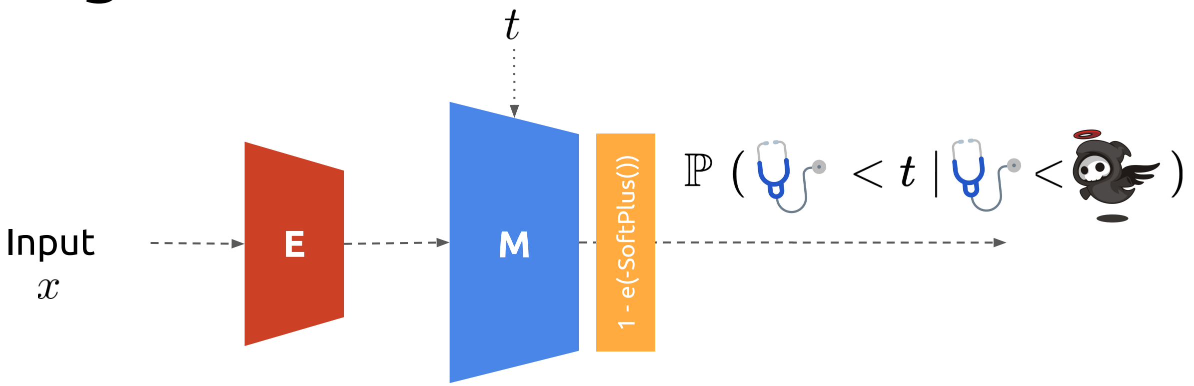
Embedding covariates



Multi Layer Perceptron

Monotonic Neural Network

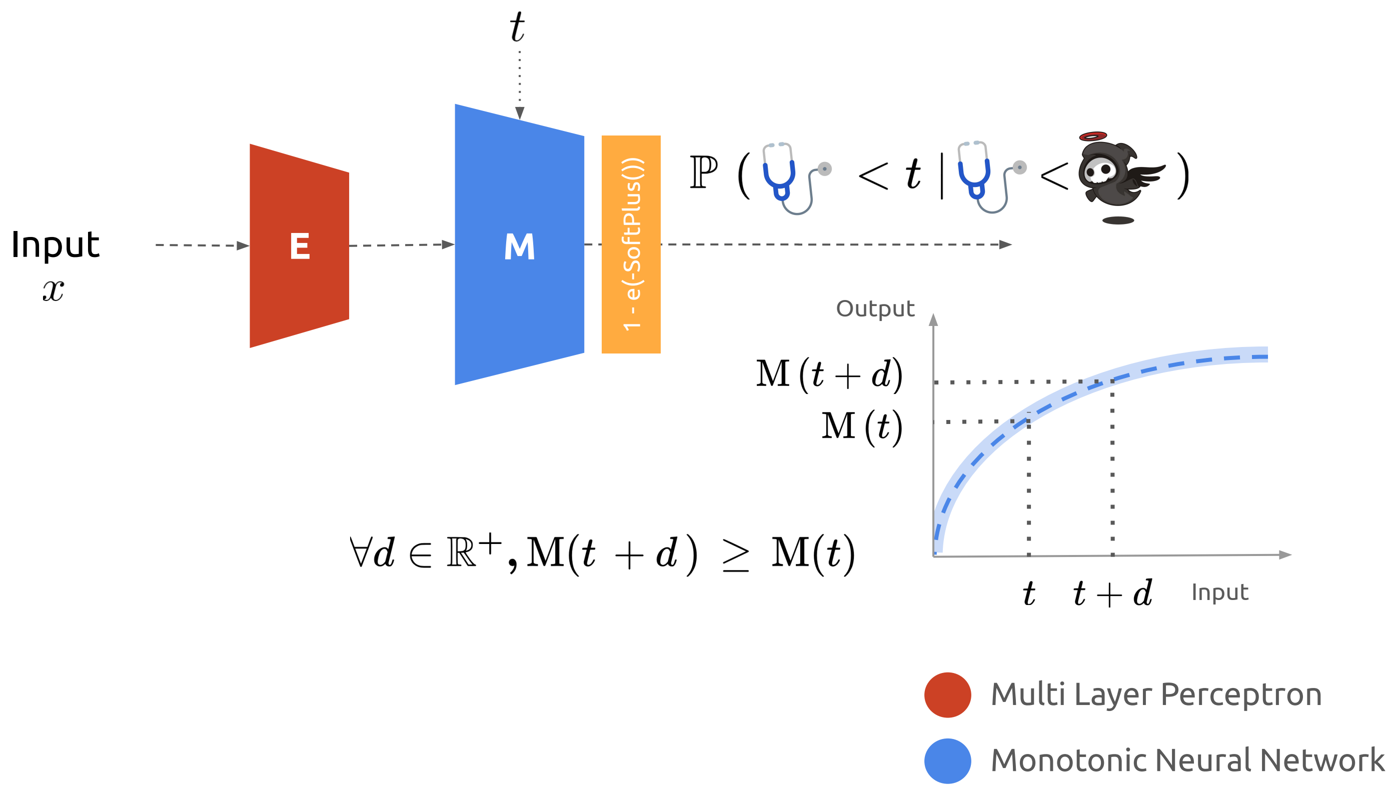
Modelling conditional outcome



Multi Layer Perceptron

Monotonic Neural Network

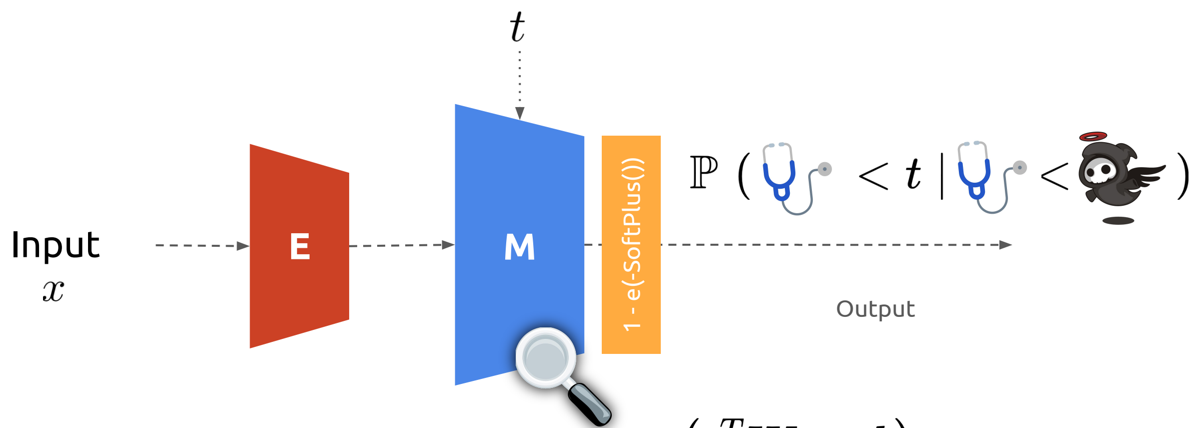
What is a monotonic neural network ?



$$\forall d \in \mathbb{R}^{+}, \mathrm{M}(t+d) \geq \mathrm{M}(t)$$

Monotonic networks by J. Sill, 1997

What is a monotonic neural network ?



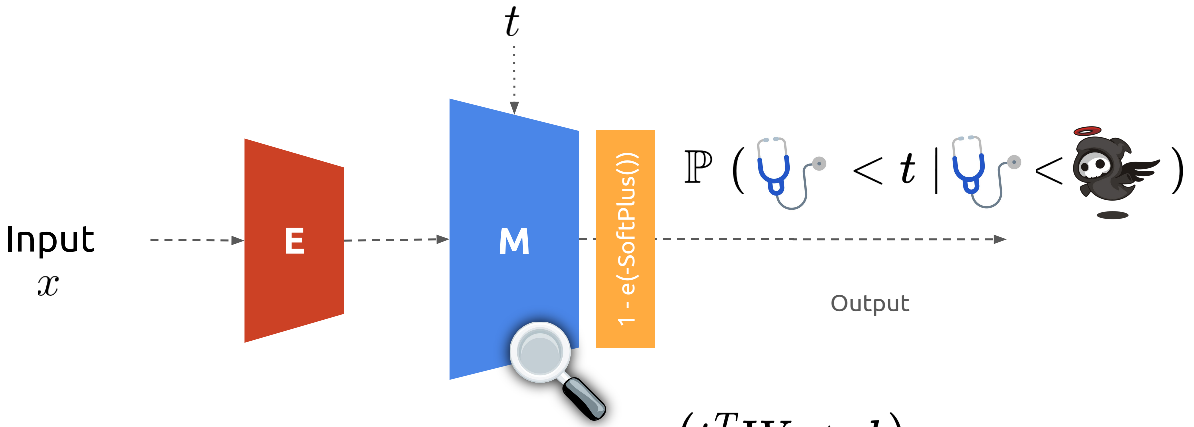
$$\sigma\left(t^T \underline{W} + \underline{b}\right)$$

Multi Layer Perceptron

Monotonic Neural Network

Monotonic networks by J. Sill, 1997

What is a monotonic neural network ?



$$\sigma\left(t^T \underline{W} + \underline{b}\right)$$

$$W = w^2 > 0$$

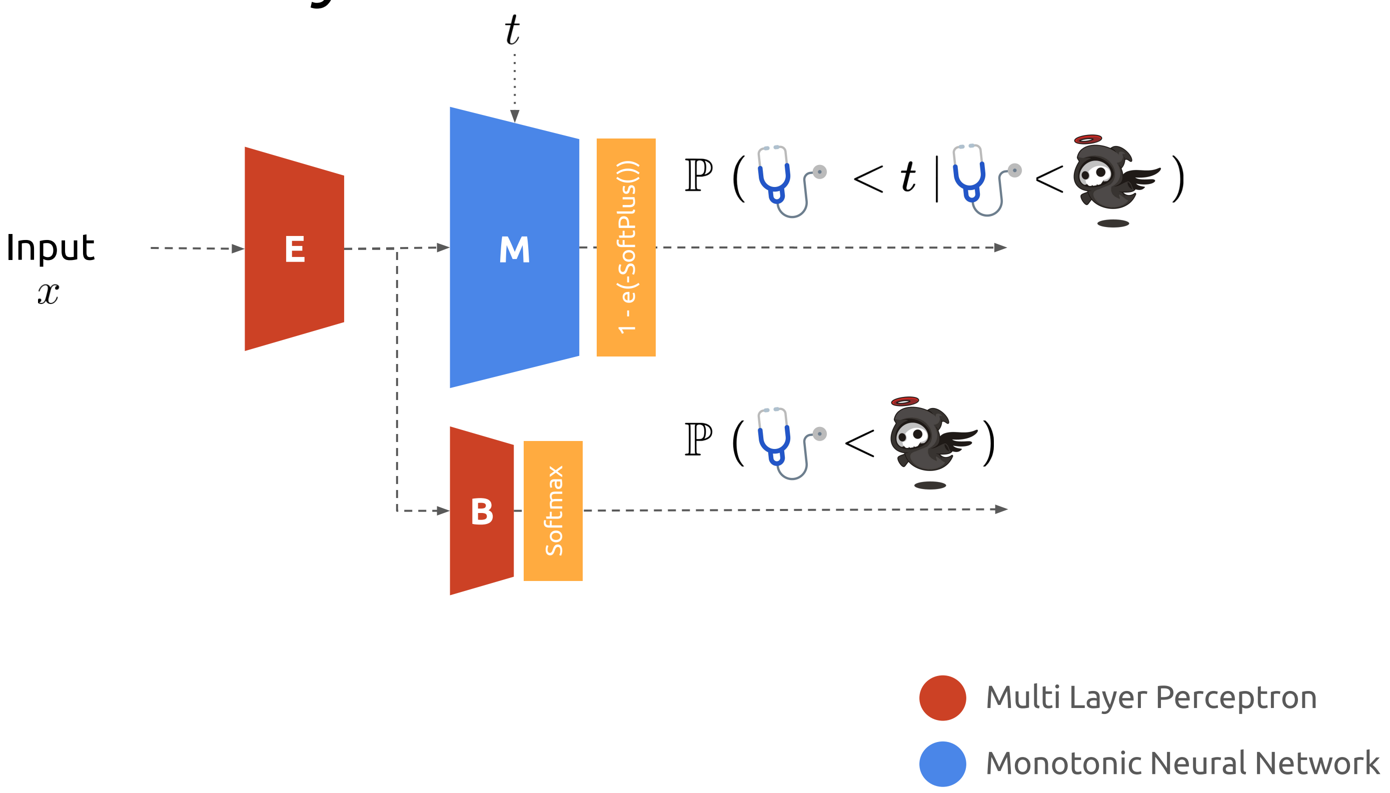
Positively weighted neural networks are **universal monotonic approximators**.

Multi Layer Perceptron

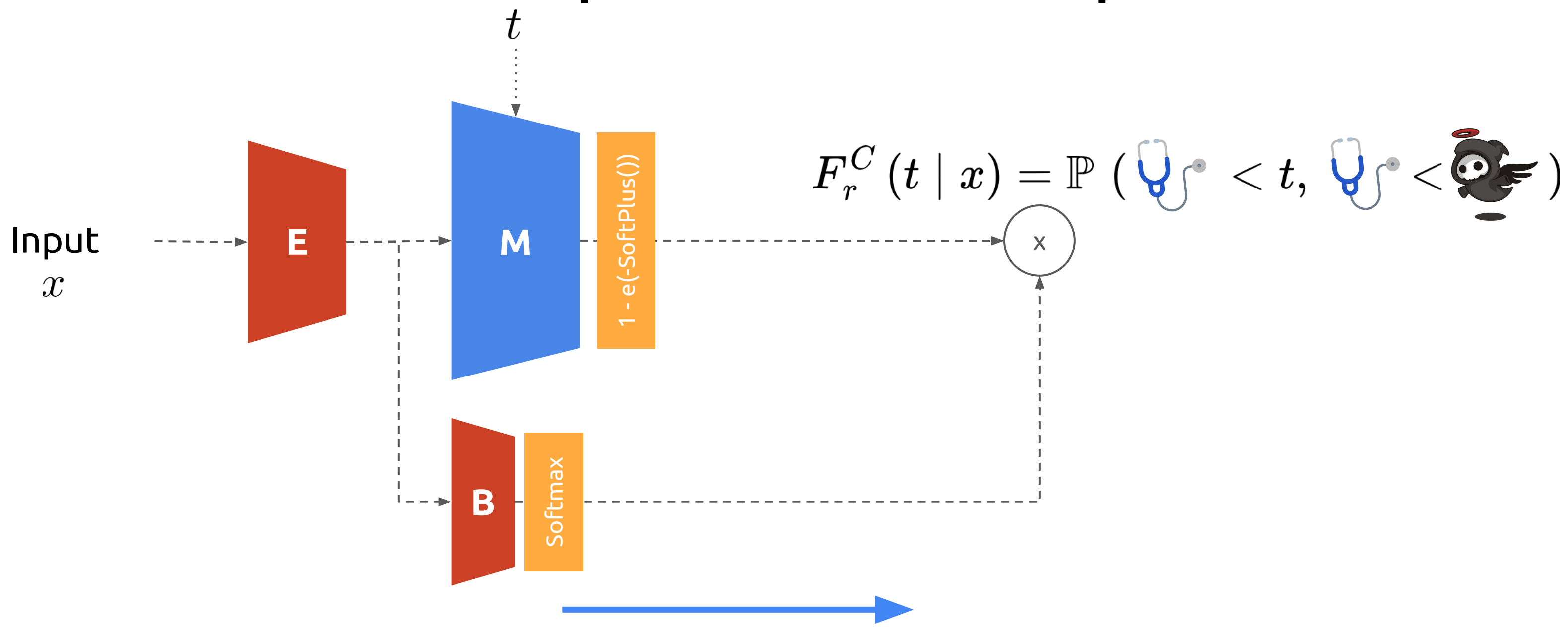
Monotonic Neural Network

Monotonic Multi-layer Perceptron Networks as Universal Approximators by Lang, 2005

Neural Fine-Gray

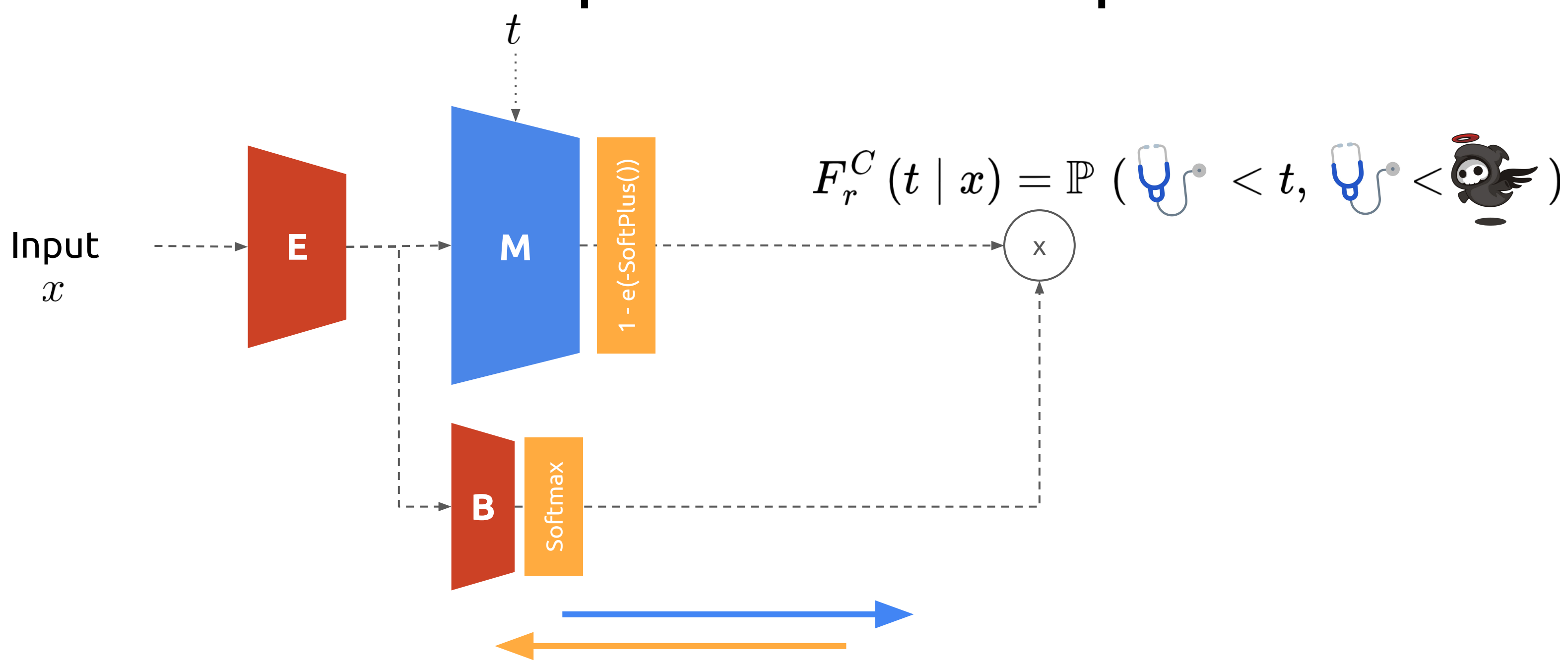


Efficient and exact computation of all quantities



$$l := \sum_r \sum_{i,\, d_i = r} \log \left. \frac{\partial F_r^C(t \mid x_i)}{\partial t} \right|_{t=t_i} + \sum_{i,\, d_i = 0} \log \left[1 - \sum_r F_r^C(t_i \mid x_i) \right]$$

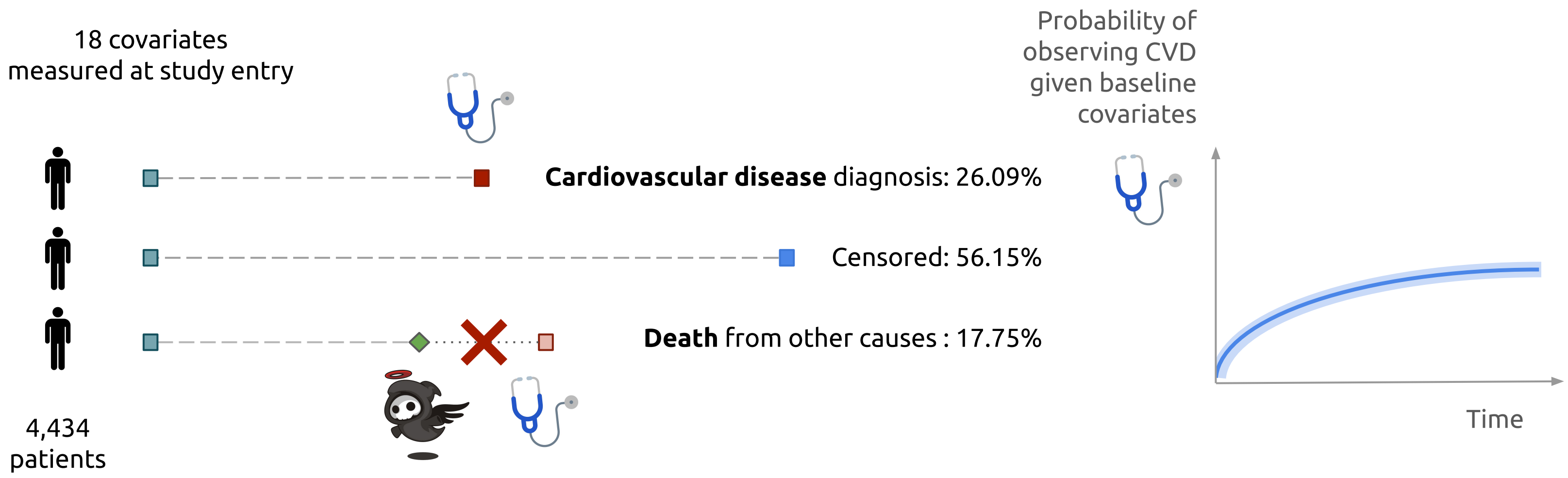
Efficient and exact computation of all quantities



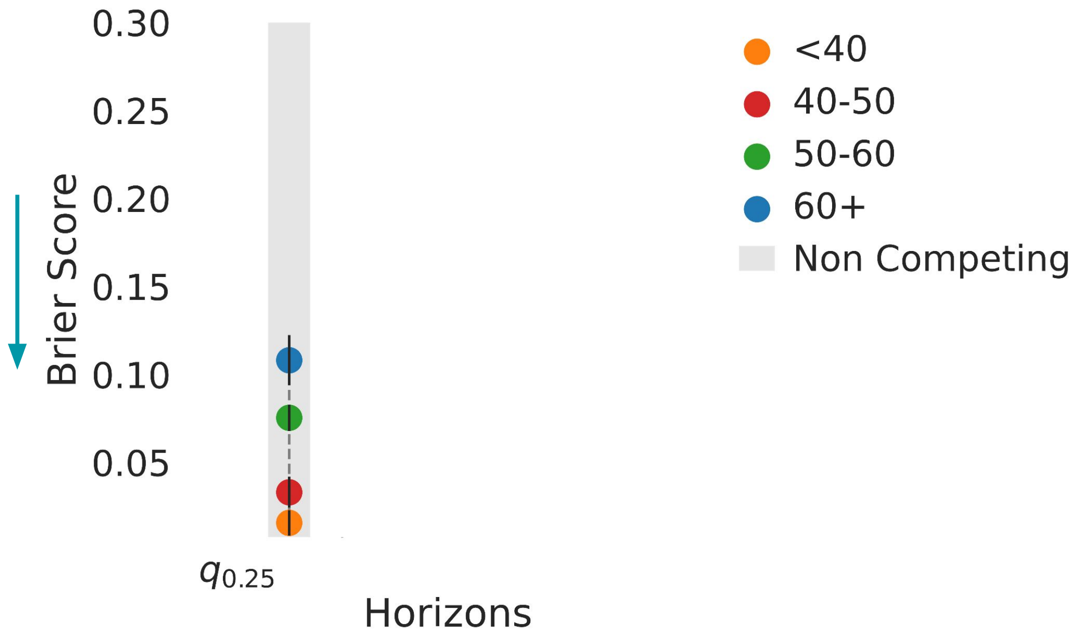
$$l := \sum_r \sum_{i,\, d_i = r} \log \frac{\partial F_r^C(t \mid x_i)}{\partial t}\Big|_{t=t_i} + \sum_{i,\, d_i=0} \log \left[1 - \sum_r F_r^C(t_i \mid x_i) \right]$$

Impact on Cardiovascular Care Management

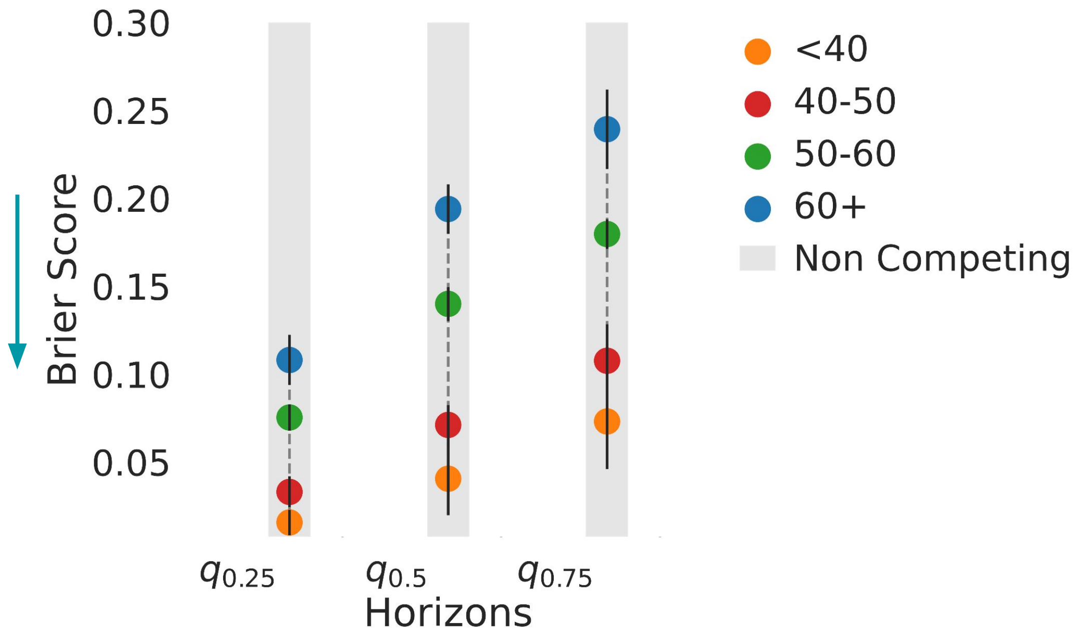
Experimental settings



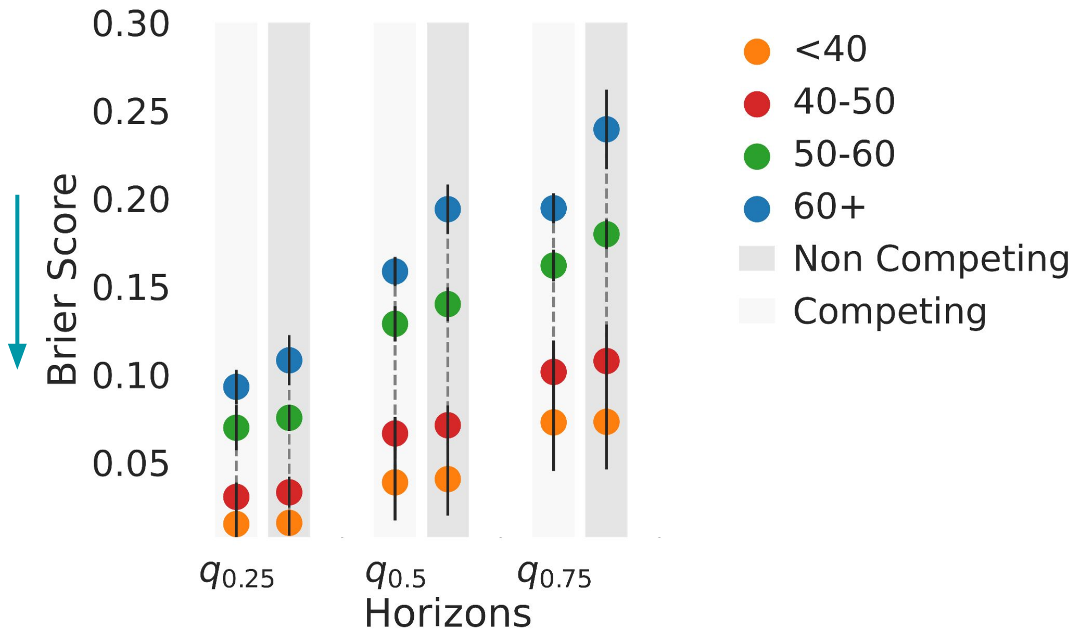
Ignoring competing risks



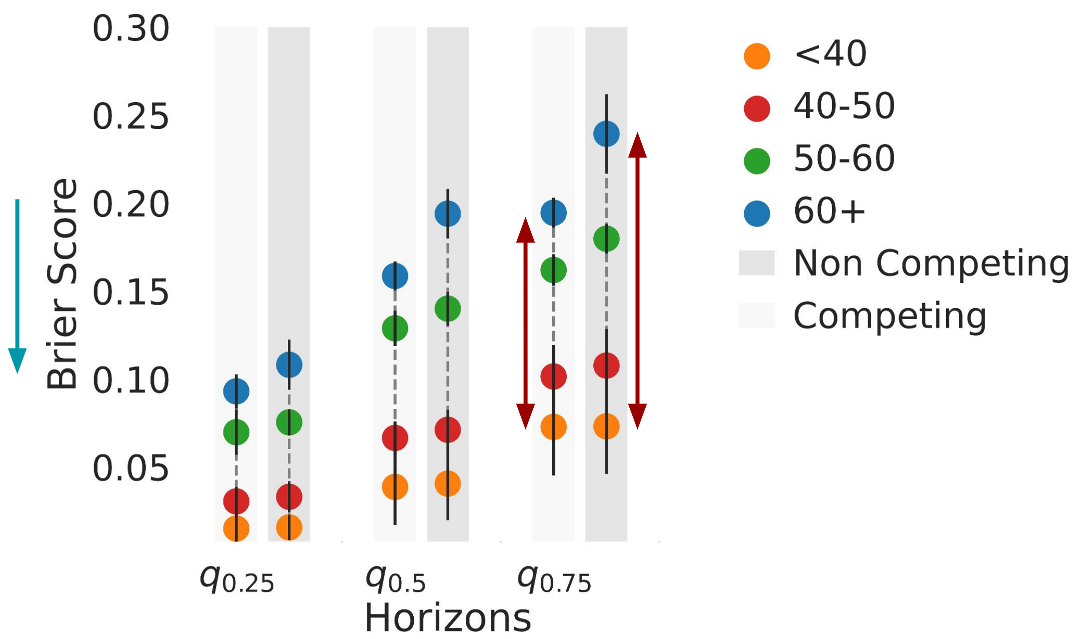
Performances decrease with longer horizons



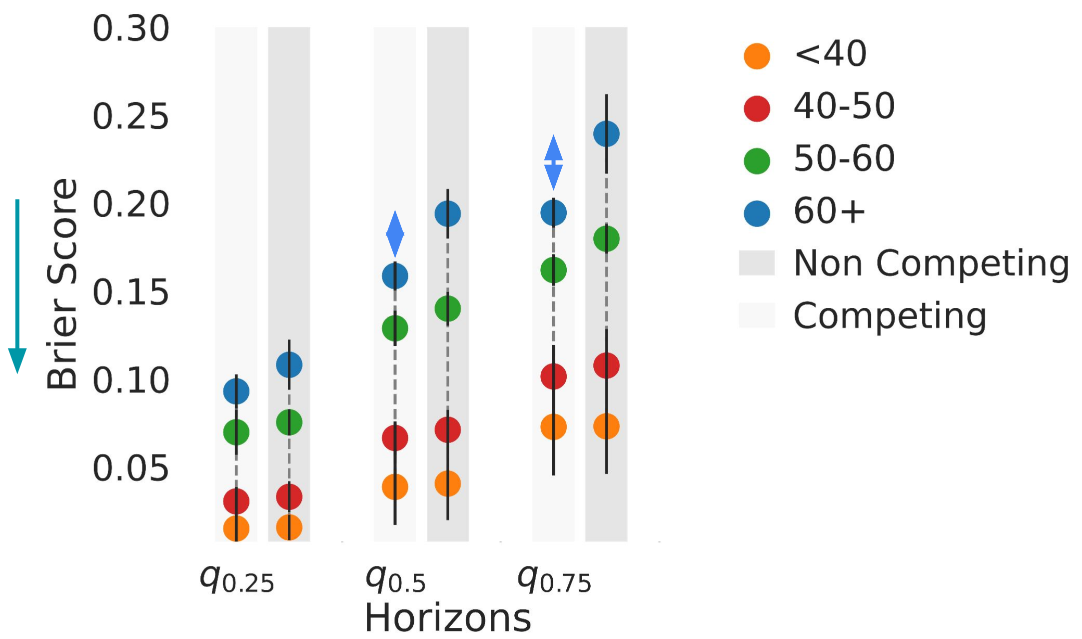
Modelling competing risks improves performance



Modelling competing risks reduces gap



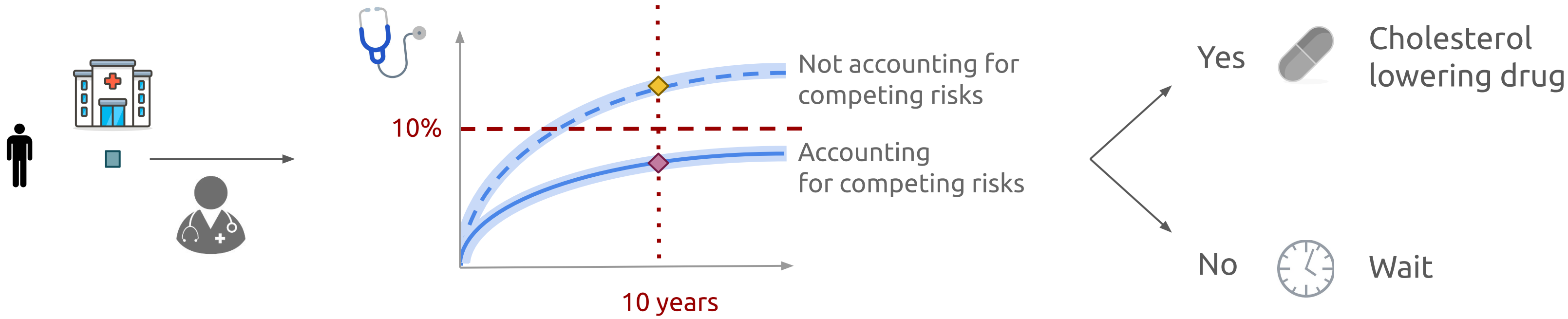
Groups benefit differently



Patients the **most at risk** for the competing risks benefit the most.

Impact on medical practice

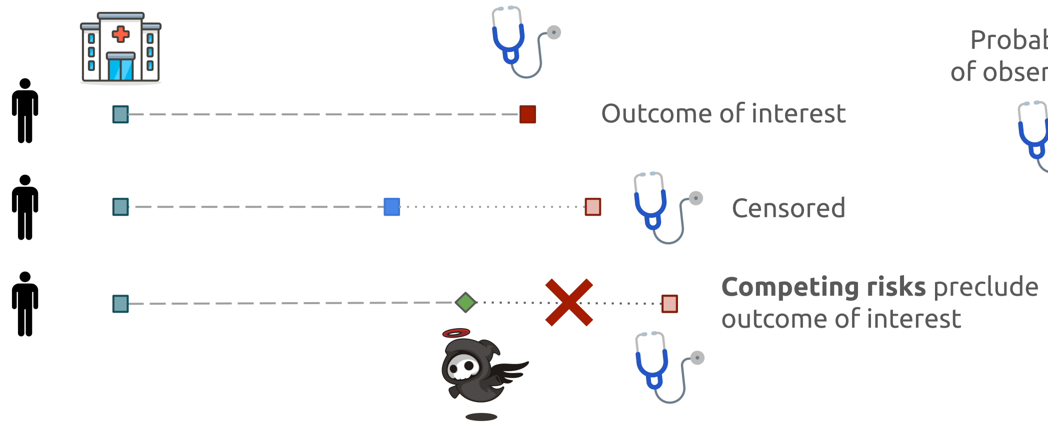
Is the patient's risk more than 10% in the 10 next years ?

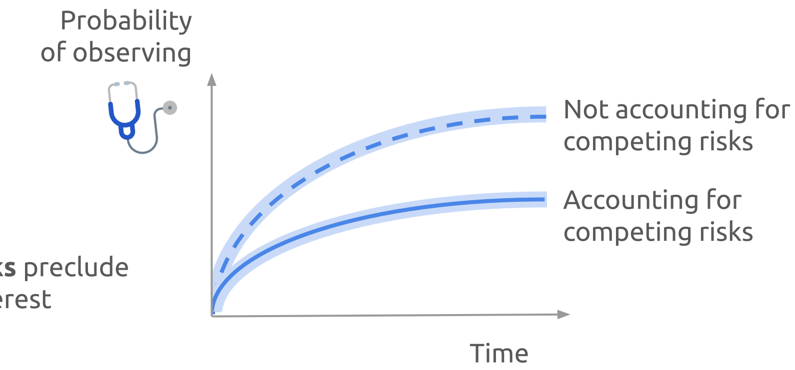


16% of 60+ patients would avoid treatment

Conclusions

1. Modelling competing risks as censoring results in **overestimating risks** and impacts **algorithmic fairness**
2. The proposed **Neural Fine Gray** models competing risks exactly and efficiently





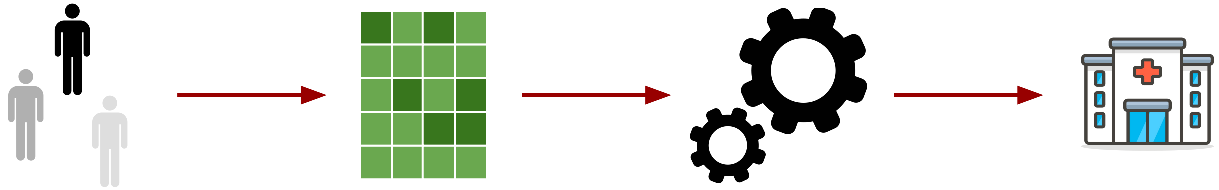
Jeanselme, V., Yoon, C. H., Tom, B., & Barrett, J. (2023). *Neural Fine-Gray: Monotonic neural networks for competing risks*. In Conference on Health, Inference, and Learning (pp. 379-392). PMLR.

Jeanselme, V., Yoon, C. H., Tom, B., & Barrett, J. *Improper Modelling of Competing Risks: Impact on Risk Estimation and Algorithmic Fairness*

Future Directions

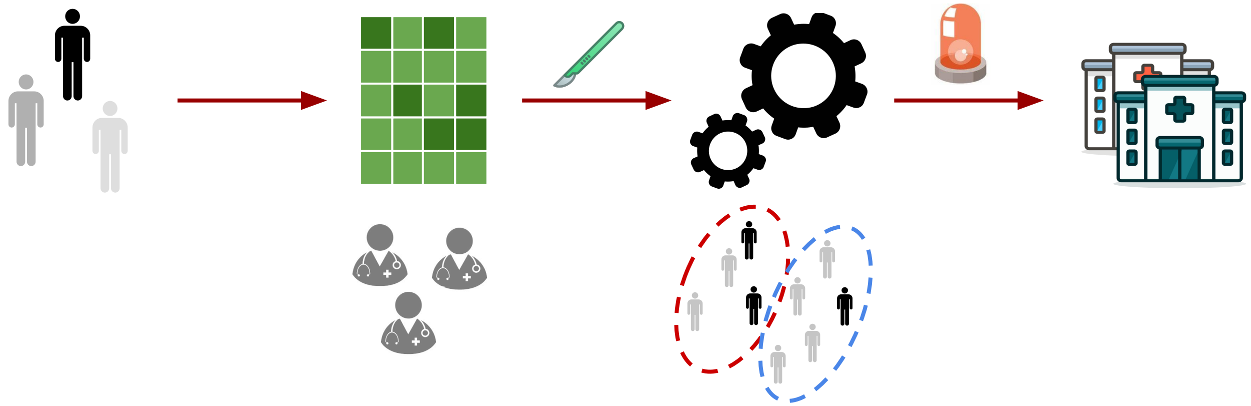
Future directions

Develop predictive models for medical decision-making and addressing socio-medical disparities present in medical data.
How can we improve prediction from data and labels resulting from imperfect decisions?



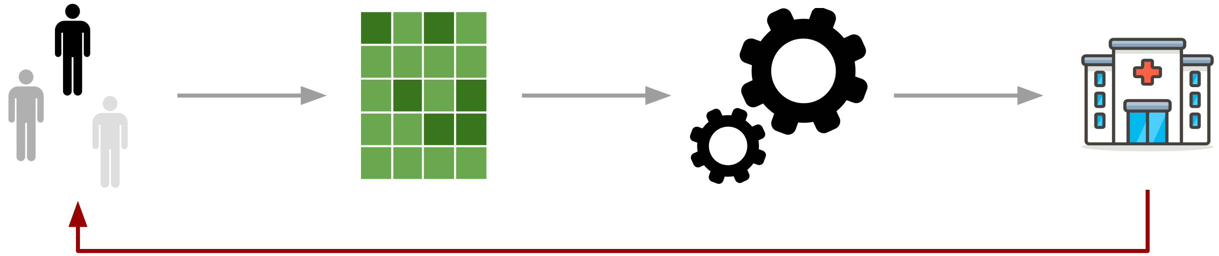
Future directions

Develop predictive models for medical decision-making and addressing socio-medical disparities present in medical data.
How can we improve prediction from data and labels resulting from imperfect decisions?



Future directions

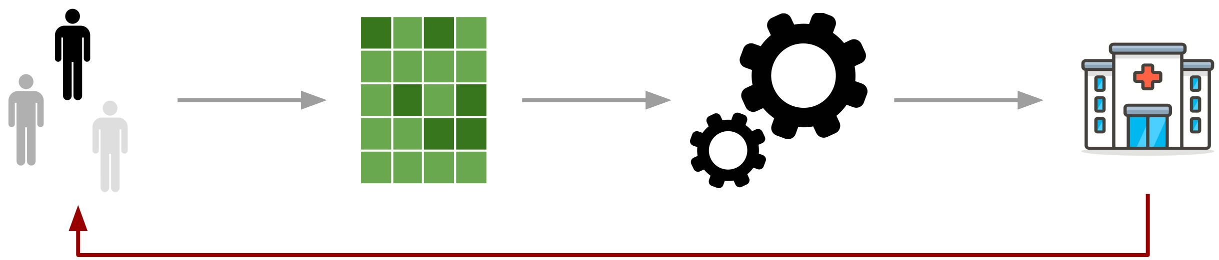
Develop predictive models for medical decision-making and addressing socio-medical disparities present in medical data.
How can we improve prediction from data and labels resulting from imperfect decisions?



Deploy and measure impact on care and practice
How can we improve medical decisions?

Future directions

Develop predictive models for medical decision-making and addressing socio-medical disparities present in medical data.
How can we improve prediction from data and labels resulting from imperfect decisions?



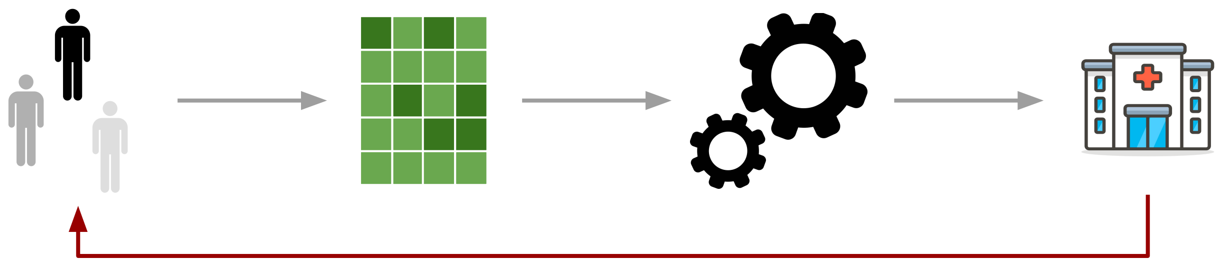
Deploy and measure impact on care and practice
How can we improve medical decisions?

riSCC
A personalized risk calculator for cutaneous squamous cell carcinoma

Jambusaria-Pahlajani, A.*, Jeanselme, V.*, et al. (2024)
riSCC: A personalized risk model for the development of poor outcomes in cutaneous squamous cell carcinoma
Journal version under review at JAMA Network Open.

Future directions

Develop predictive models for medical decision-making and addressing socio-medical disparities present in medical data.
How can we improve prediction from data and labels resulting from imperfect decisions?

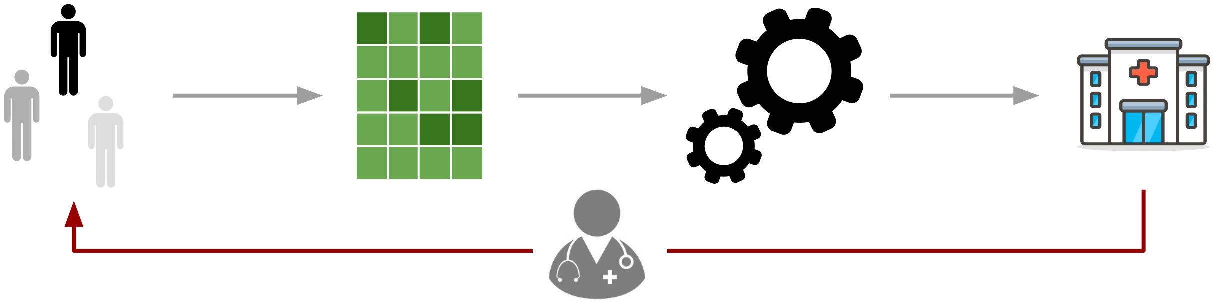


Deploy and measure impact on care and practice
How can we improve medical decisions?

1. *Develop trials to quantify the benefit of ML*
2. *Consider all dimensions of medical decisions*

Future directions

Develop predictive models for medical decision-making and addressing socio-medical disparities present in medical data.
How can we improve prediction from data and labels resulting from imperfect decisions?



Deploy and measure impact on care and practice
How can we improve medical decisions?

1. *Develop trials to quantify the benefit of ML*
2. *Consider all dimensions of medical decisions*
3. *Human-Centered AI: Consider decisions as part of the pipeline*

Jeanselme, V. Agarwal, N., Wang C. (2024) *Review of Language Models for Survival Analysis.* In AAAI 2024 Spring Symposium Series Clinical FMs

Jambusaria-Pahlajani, A.*, Jeanselme, V.*, Wang, D., Ran, N., Granger, E., Cañuet, J., Brodland, D., Carr, D., Carter, J., Carucci, J., Hirotsu, K., Koyfman, S., Mangold, A., Girardi, F., Shahwan, K., Srivastava, D., Vidimos, A., Willenbrink, T., Wysong, A., Lotter, W., Ruiz, E. (2024) *riSCC: A personalized risk model for the development of poor outcomes in cutaneous squamous cell carcinoma* - Journal version under review at Journal of Clinical Oncology.

De-Arteaga, M., Jeanselme, V., Dubrawski, A., Chouldechova, A. (2024). *Leveraging expert consistency to improve algorithmic decision support.* - Accepted in Management Science.

Jeanselme, V., Yoon, C., Falck, F., Tom, B., Barrett, J. Identifying treatment response subgroups in observational time-to-event data, under review at ICLR

Jeanselme, V. *Improper Modelling of Competing Risks: Impact on Risk Estimation and Algorithmic Fairness,* work in progress

Jeanselme, V., Yoon, C. H., Tom, B., Barrett, J. (2023). *Neural Fine-Gray: Monotonic neural networks for competing risks.* In Conference on Health, Inference, and Learning (pp. 379-392). PMLR.

Jeanselme, V., Tom, B., Barrett, J. (2022). *Neural Survival Clustering: Non-parametric mixture of neural networks for survival clustering.* In Conference on Health, Inference, and Learning (pp. 92-102). PMLR.

Jeanselme, V., De-Arteaga, M., Zhang, Z., Barrett, J., Tom, B. (2022). *Imputation Strategies Under Clinical Presence: Impact on Algorithmic Fairness.* In Machine Learning for Health (pp. 12-34). PMLR. and Journal version under review in Management Science.

Jeanselme, V., Martin, G., Peek, N., Sperrin, M., Tom, B., Barrett, J. (2022). *Deepjoint: Robust survival modelling under clinical presence shift.* NeurIPS 2022 Workshop on Learning from Time Series for Health.

Nagpal, C.*, Jeanselme, V.*, Dubrawski, A. (2021). *Deep parametric time-to-event regression with time-varying covariates.* In AAAI Spring Symposium Survival Prediction-Algorithms, Challenges and Applications (pp. 184-193). PMLR.

Jeanselme, V., De-Arteaga, M., Elmer, J., Perman, S. M., Dubrawski, A. (2021). *Sex differences in post-cardiac arrest discharge locations.* Resuscitation Plus, 8, 100185.

Yoon, J. H.*, Jeanselme, V.*, Dubrawski, A., Hravnak, M., Pinsky, M. R., Clermont, G. (2020).
Prediction of hypotension events with physiologic vital sign signatures in the intensive care unit. Critical Care, 24(1), 1-9.