Distilling Causal Models

Model Averaging, Federated Learning and More

Marco Scutari
scutari@bnlearn.com

Dalle Molle Institute for
Artificial Intelligence (IDSIA)

September 4, 2024

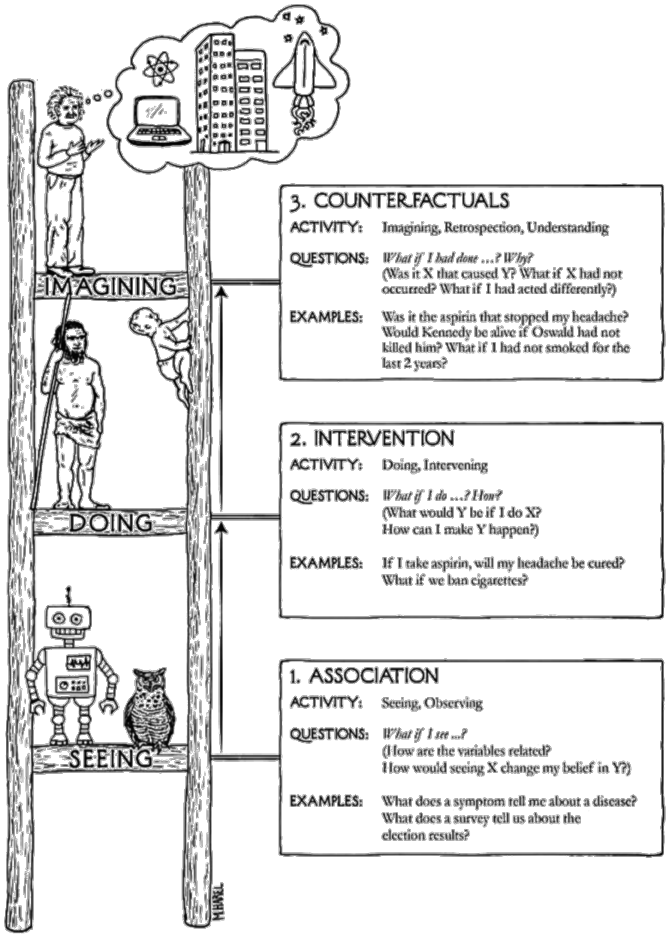
Causality is a Network

Machine learning creates black boxes that use probabilistic associations for prediction. Scientific questions are inherently causal.

We have [13, 21] a rigorous theory of causality that uses directed (acyclic) graphs (DAGs) to represent causes and effects. With it, we can reason about

- what we see,
- affecting change,
- hypothetical situations.

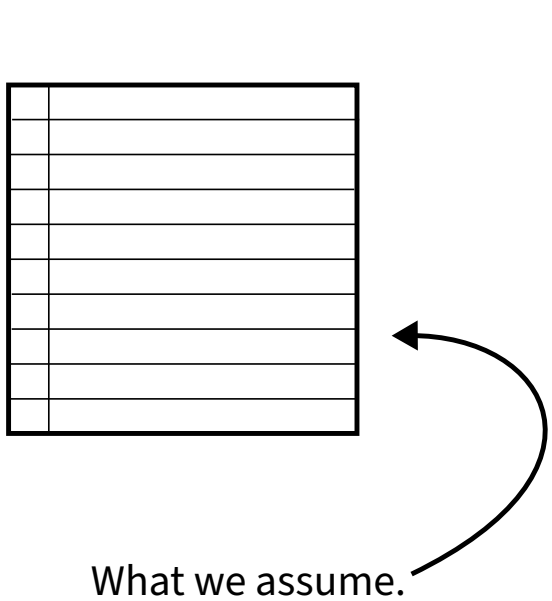
How can we learn causal DAGs from clinical data?

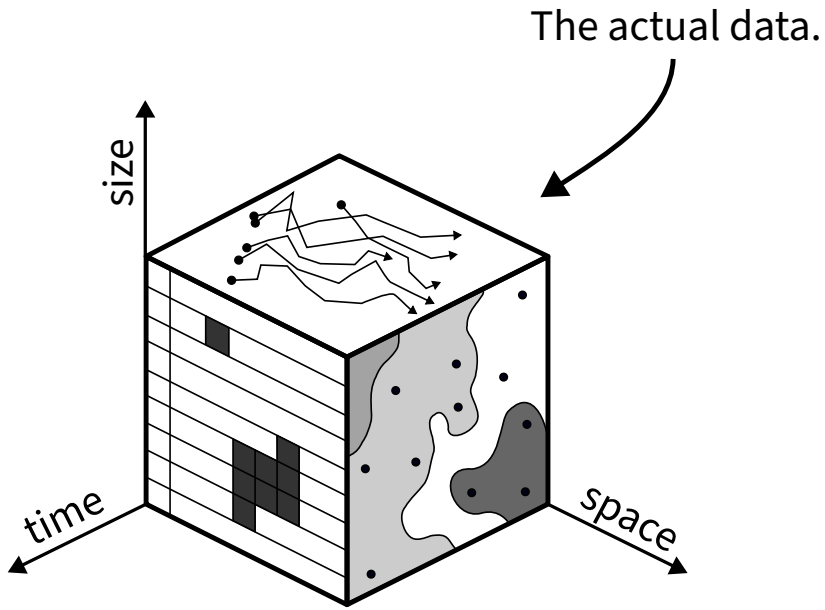


Learning a causal network means learning its structure $\mathcal{G}$ and parameters Θ, much like Bayesian networks:

$$\underbrace{\mathrm{P}(\mathcal{G}, \Theta \mid \mathcal{D})}_{\text{learning}} = \underbrace{\mathrm{P}(\mathcal{G} \mid \mathcal{D})}_{\text{structure learning}} \cdot \underbrace{\mathrm{P}(\Theta \mid \mathcal{G}, \mathcal{D})}_{\text{parameter learning}}.$$

We used to ask domain experts for information [6, 7]; now we rely more and more on learning algorithms and the data $\mathcal{D}$ [24].





Bayesian networks [23] are defined by:

- a network structure, a directed acyclic graph (DAG) $\mathcal{G}$, in which each node corresponds to a random variable X_i;
- a global probability distribution $\mathbf{X}$ with parameters Θ, which factorises into smaller local probability distributions according to the arcs in $\mathcal{G}$:

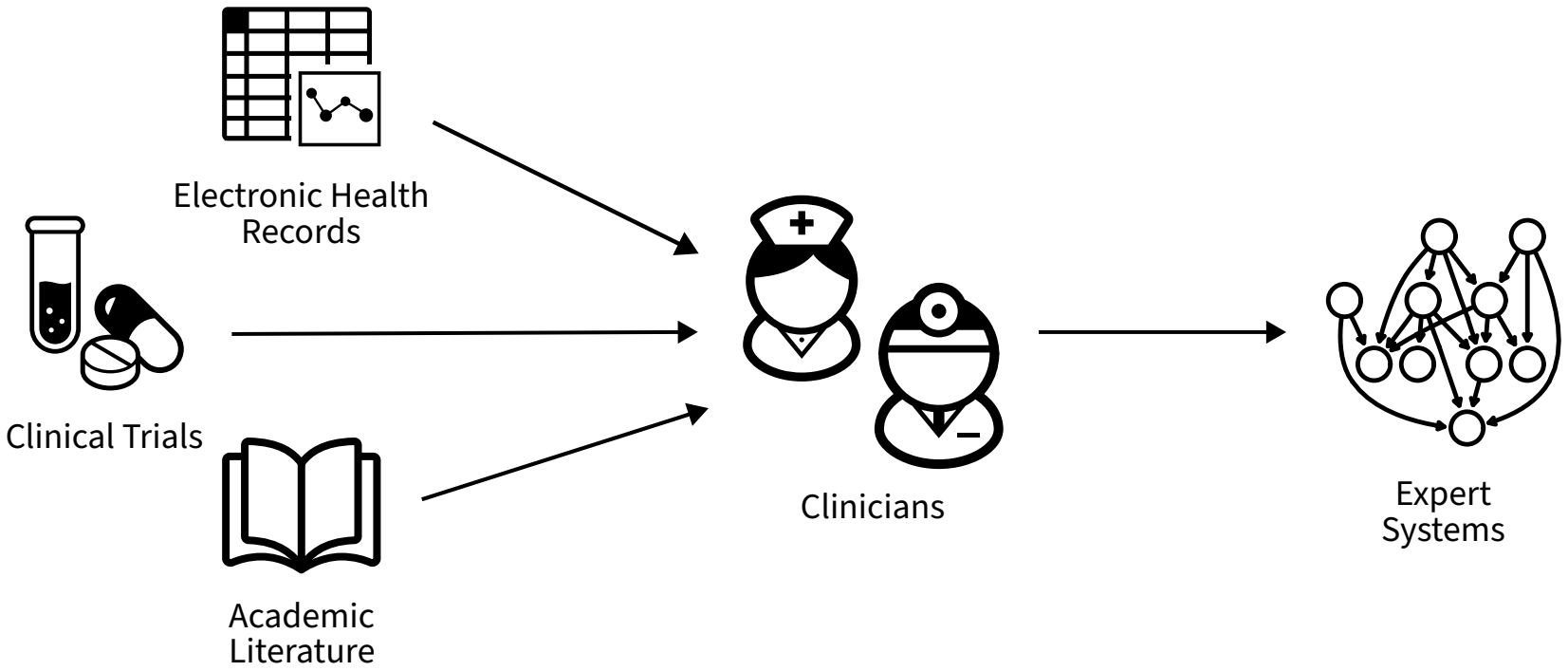
$$\mathrm{P}(\mathbf{X}, \Theta) = \prod_{i=1}^{N} \mathrm{P}(X_i \mid \Pi_{X_i}; \Theta_{X_i}) \quad \text{where} \quad \Pi_{X_i} = \{\text{parents of } X_i\}.$$

Causal networks are Bayesian networks with extra assumptions, such as causal sufficiency, that allow for interpreting arcs as causal effects.

- ✓ Causal Discovery
- → **Here Comes Clinical Data**
- Reality-Centric Causal Networks
- Conclusions
- Acknowledgements

The Role of Clinical Experts

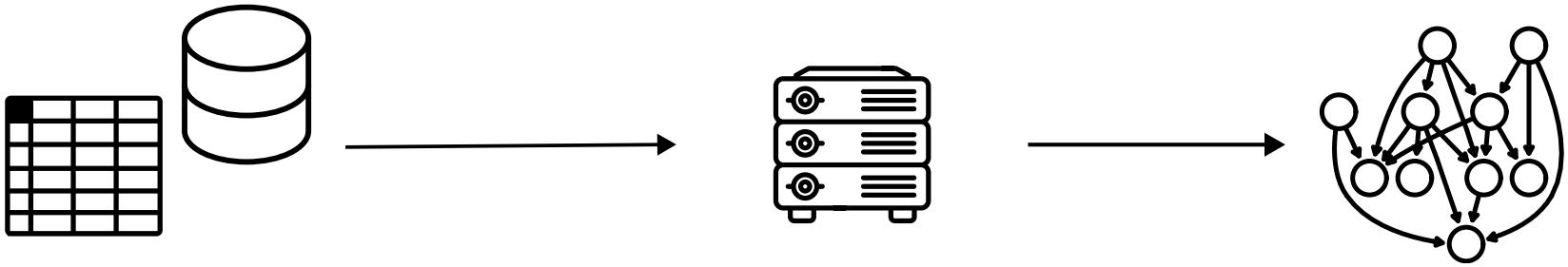
First, there were expert systems [3, 4] compiled with clinicians.



The clinicians distilled all the available knowledge into a set of causal effects ($\mathcal{G}$) and their effect sizes (Θ).

Enter Machine Learning

However, clinicians cannot scale to complex phenomena and often do not agree with each other. Machine learning then took a completely data-driven approach.

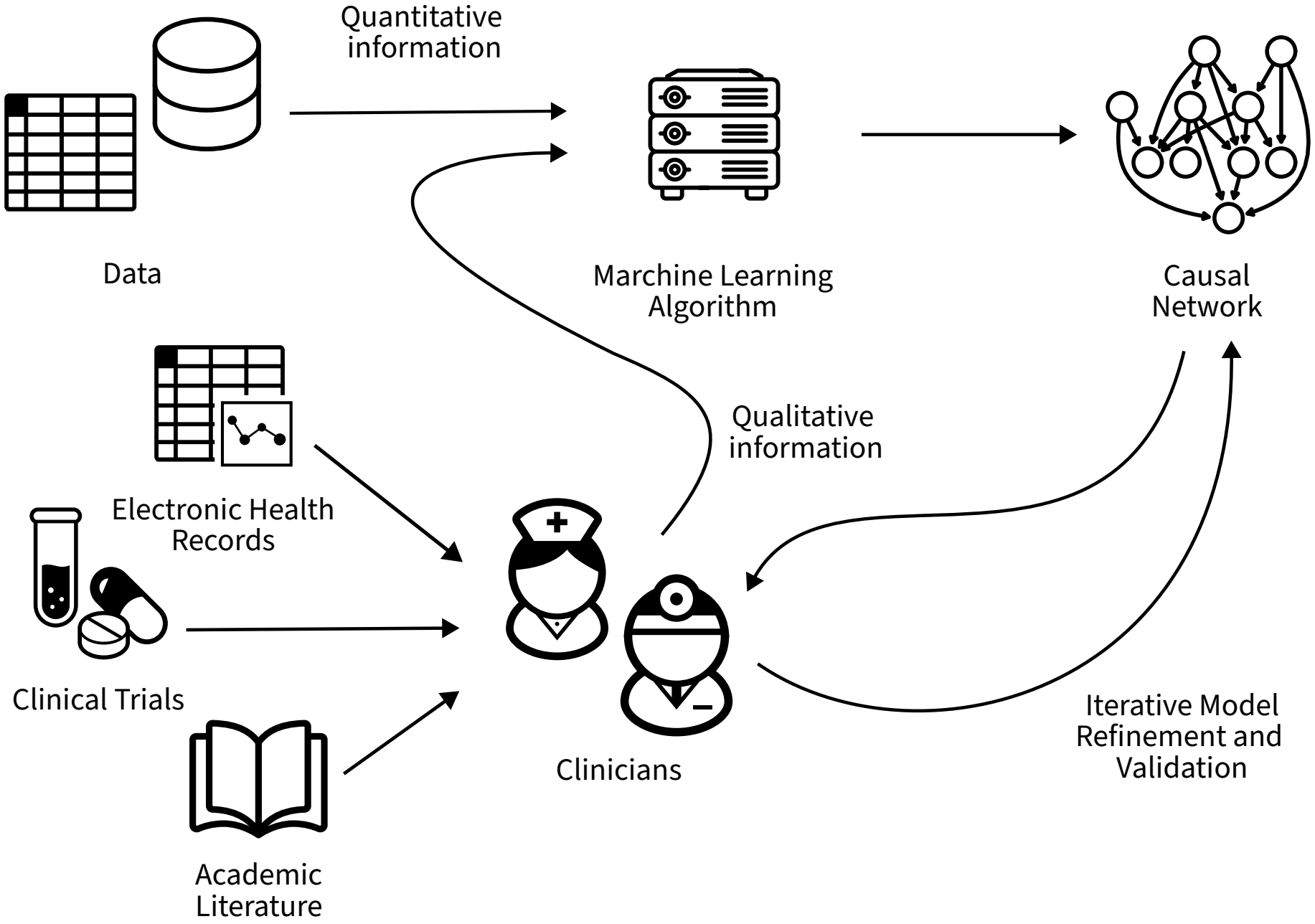


This is the current paradigm of structure learning (for Bayesian networks) and causal discovery (for causal networks) [14, 28]. Many, many such algorithms in the literature [24].

The Limits of a Purely Data-Driven Approach

- We do not have enough data to capture the complexity of modern clinical problems.
- Causal discovery algorithms have unrealistic assumptions and are (mostly) unusable on clinical data.
- Most of the field has forgotten the accumulated statistical wisdom on the design of experiments, hierarchical data, model validation, etc.
- There is an excess focus on artificial benchmarks that have nothing to do with the challenges and the risk evaluations of real clinical practice.
- The expertise of clinicians is under-appreciated and under-utilised in driving causal discovery.

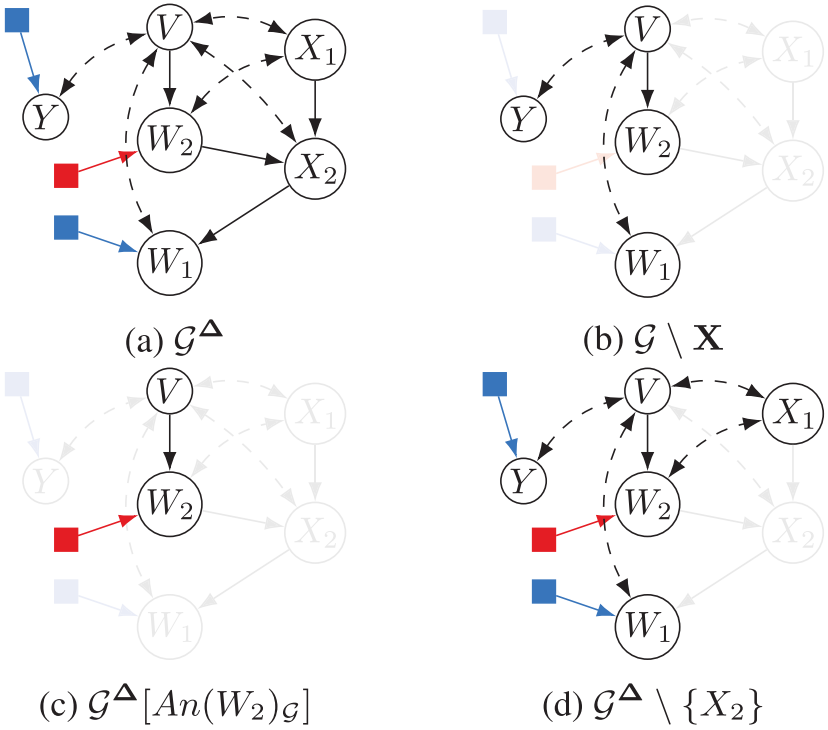
Producing Realistic, Useful Models



Technical Challenges: Missing Data and Latent Confounders

Causal networks require causal sufficiency: no latent confounders. Both sampling bias and incomplete data are such if left unmodelled.

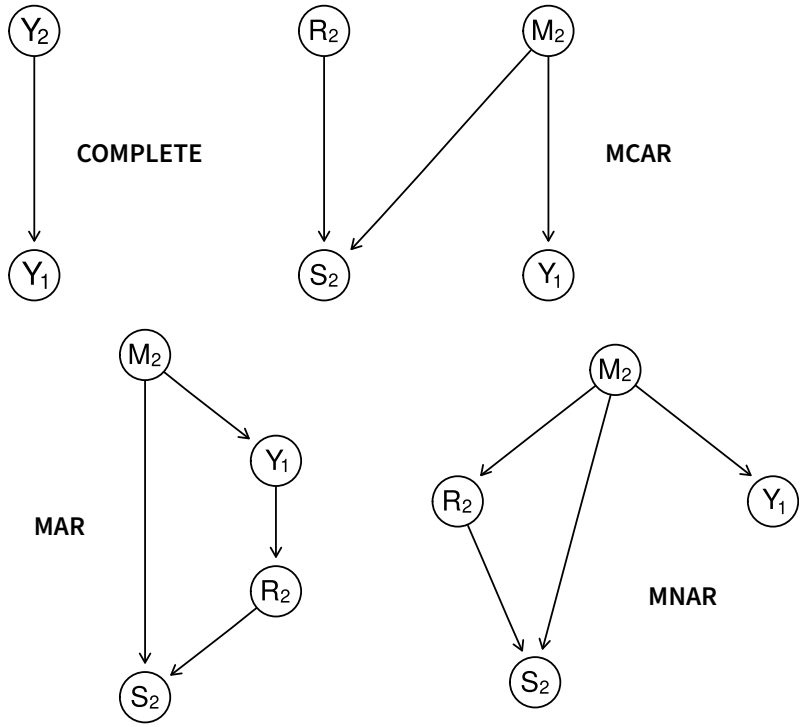
Selection diagrams [15]



(a) $\mathcal{G}^{\Delta}$ (b) $\mathcal{G} \setminus \mathbf{X}$

(c) $\mathcal{G}^{\Delta}[An(W_2)_{\mathcal{G}}]$ (d) $\mathcal{G}^{\Delta} \setminus \{X_2\}$

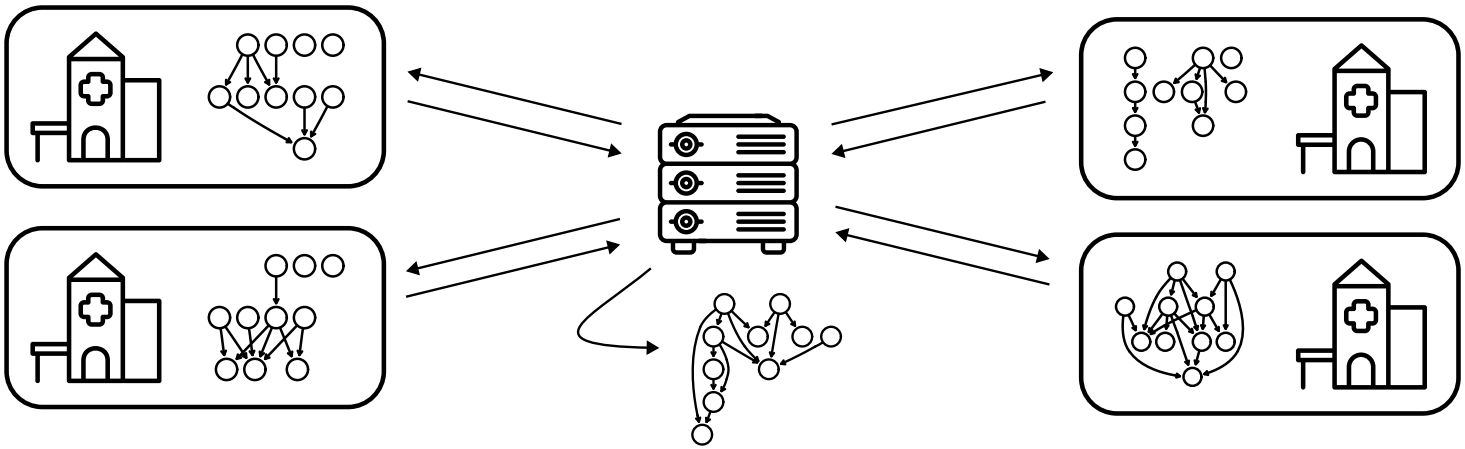
Missingness graphs [20]



Technical Challenges: Federated Learning

How do we aggregate data across centres when we cannot share it?

- Score-based: computing local scores for parent sets in the centres and aggregating them to minimise regret [18, 19].
- Constraint-based: aggregating partial networks learned using conditional independence tests [12].



The “true network” may be different in each centre, but causal discovery assumes that there is a single one that is valid for all centres...

Technical Challenges: Structured Populations

Individuals are not independent and identically distributed!

- Panel data are correlated over time.
- Epidemiological data are correlated over space.
- Subgroups of patients with differential treatment effects.
- Centre heterogeneity is unavoidable in multi-centre clinical trials.

And yet mostly $X_i = f(\Pi_{X_i}; \Theta_{X_i}) + \varepsilon_i$ with $\mathrm{COV}(\varepsilon_{i,j}, \varepsilon_{i,k}) = 0$.

- We need hierarchical models sharing information between centres, as in [1, 27], or to borrow ideas from federated learning.
- We definitely need to borrow models for state-space data from the ecological modelling literature.

Technical Challenges: Evaluation and Validation

MLOps [26] makes an important distinction:

- Model evaluation: is the statistical performance good?
- Model validation: what about the domain metrics we care about?

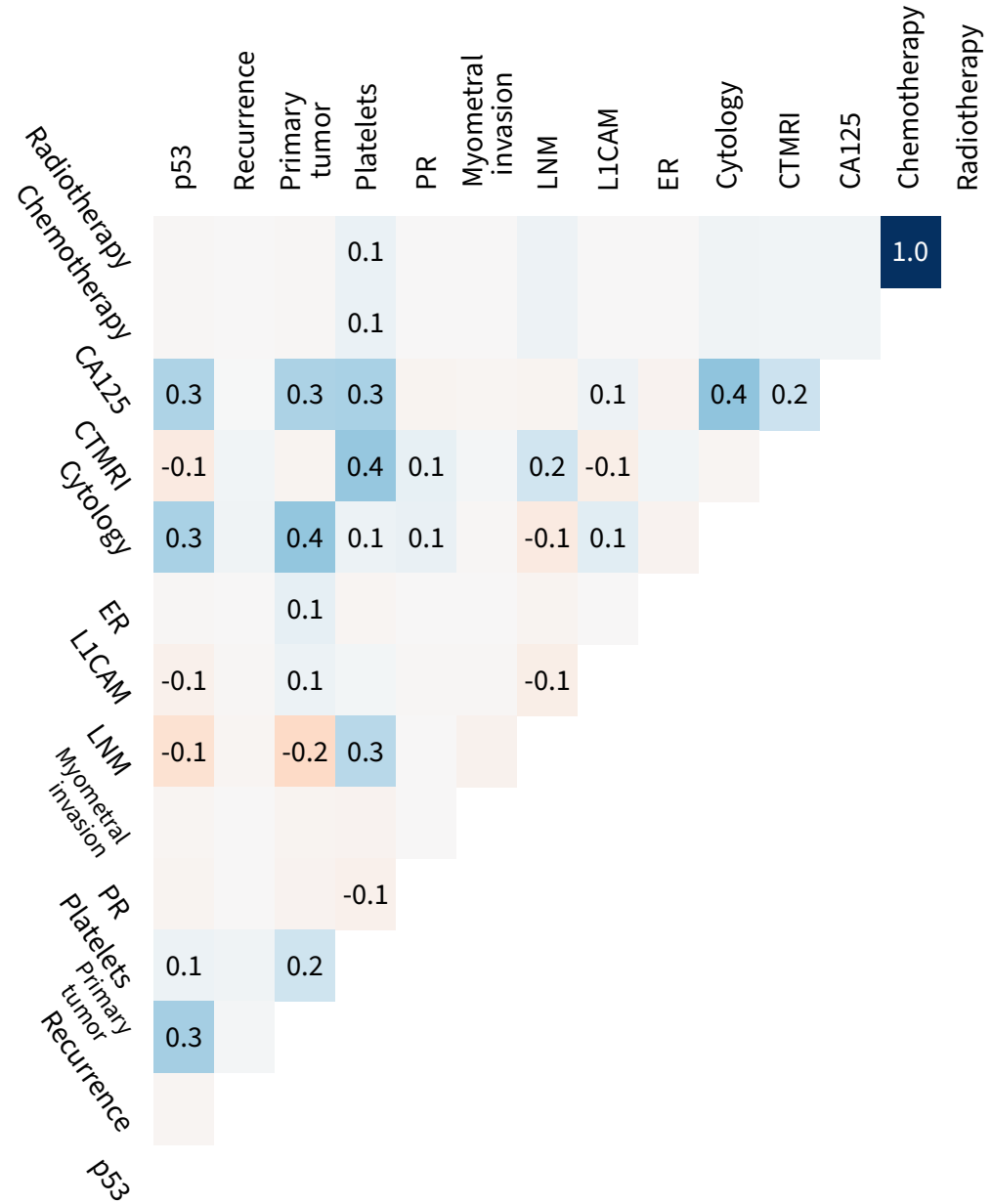
Machine learning only cares about the former, and even then there is no consensus on how to do that.

- Many existing metrics fail to exhibit a strong correlation with the quality of the approximation to the true model posterior [17].
- Different causal structural metrics with contrasting interpretations [11]. SID [22] is inconsistent and very costly to compute.
- Even quantitative metrics such as [5] are not framed in clinical terms (say, ATE). An exception may be [8].

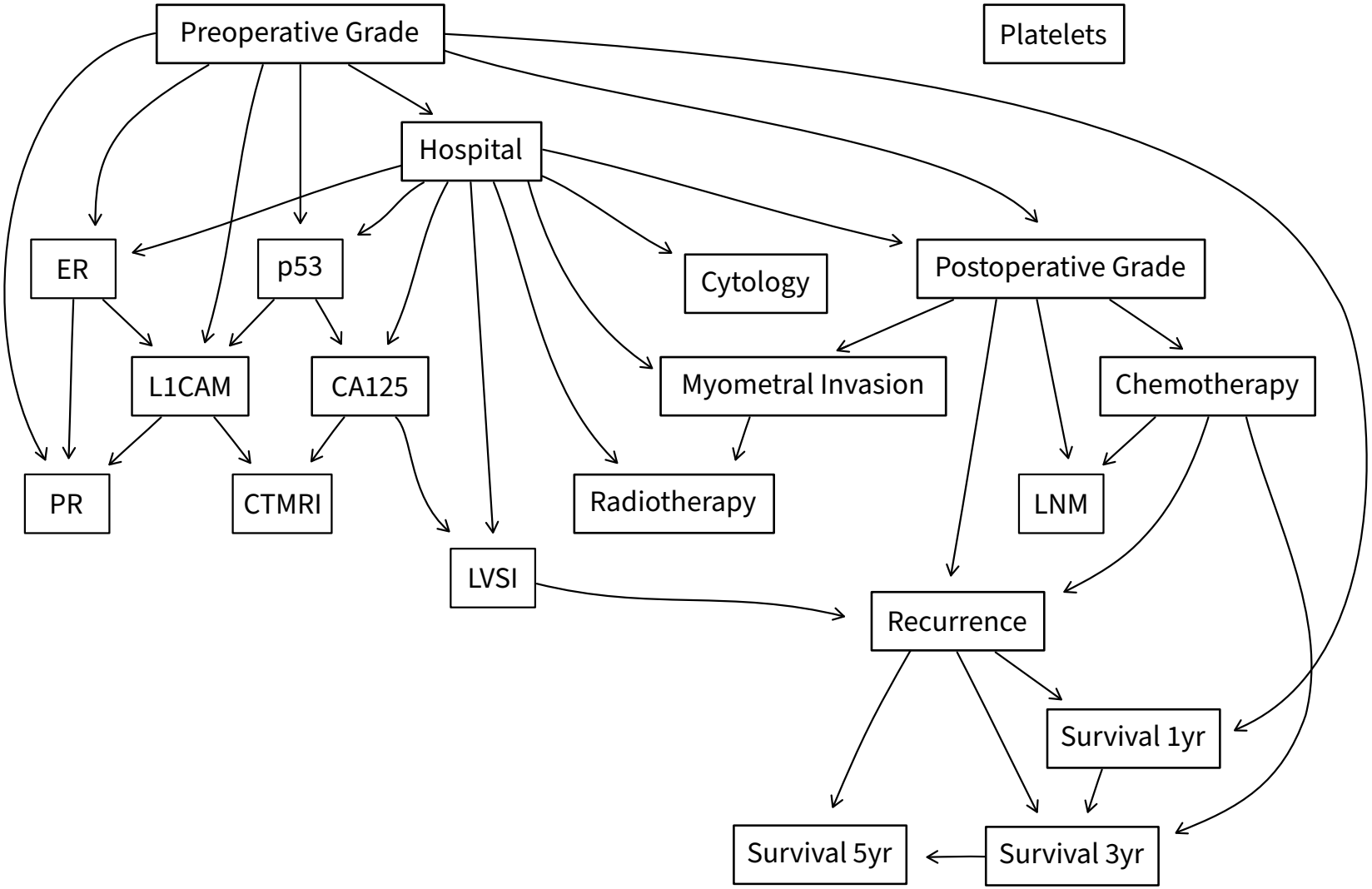
- ✓ Causal Discovery
- ✓ Here Comes Clinical Data
- → **Reality-Centric Causal Networks**
- Conclusions
- Acknowledgements

Lymph Node Metastases in Endometrial Cancer Patients

- Ten clinics in the European Network for Individualised Treatment of Endometrial Cancer (ENITEC).
- Key Issues:
  - Small sample size.
  - High missingness ratio.
  - Centre heterogeneity.
- Modelling:
  - Prior knowledge from clinicians.
  - Model averaging.
  - MNAR missingness.

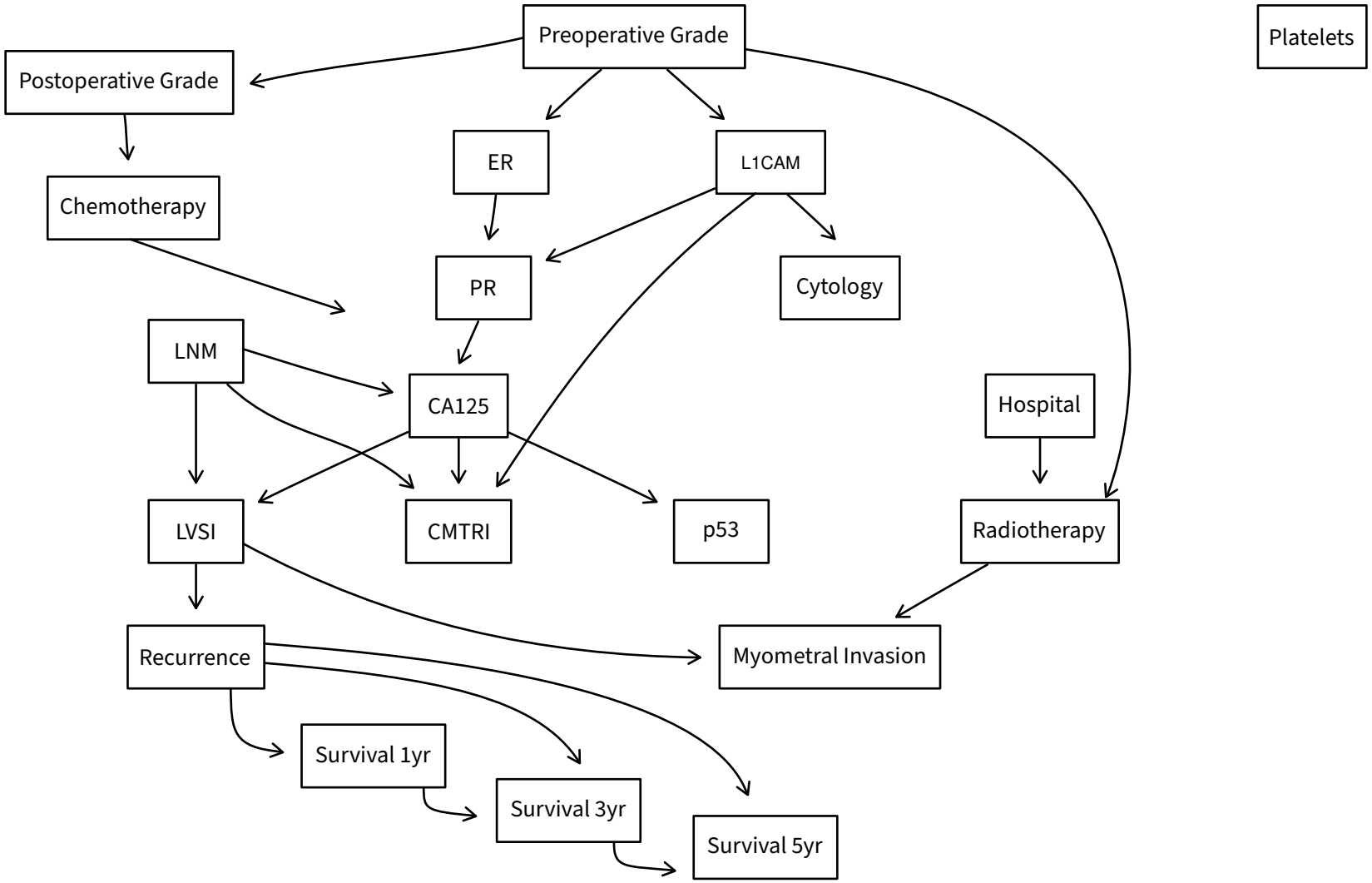


Disregarding the Missing-Not-At-Random Mechanism



Baseline model learned with SEM [9, 10] from our earlier AIxIA paper [29].

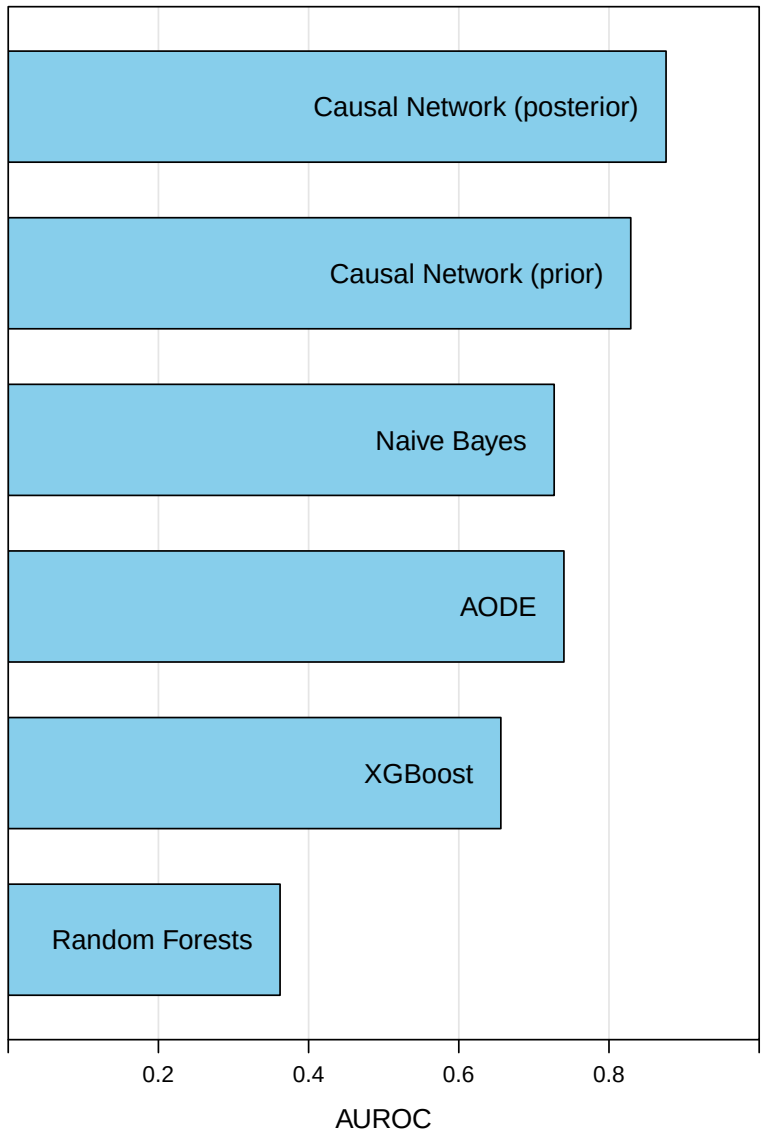
Modelling the Missing-Not-At-Random Mechanism



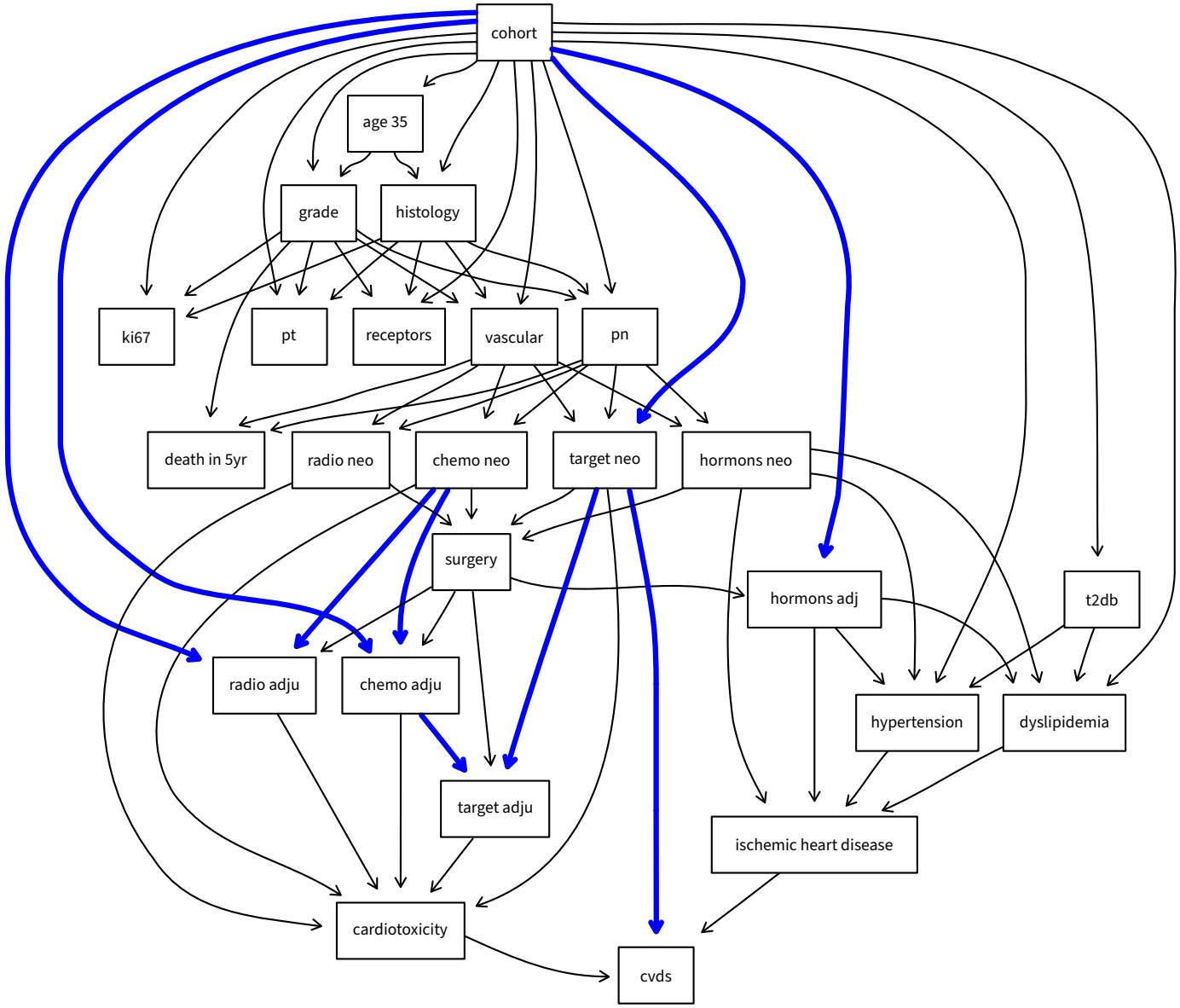
Unbiased model learned with HC-aIPW [16] in our AIME paper [30].

CVDs in Adolescents and Young Adults with Breast Cancer

- A population cohort and a clinical cohort from Istituto Nazionale dei Tumori di Milano.
- Key Issues:
  - Different sampling criteria.
  - Different sets of variables.
  - Low prevalence / incidence.
- Modelling:
  - Prior knowledge from clinicians.
  - Model averaging.
  - MNAR missingness.
  - Sampling diagrams.



Modelling Multiple Cohorts Together



Model learned with SEM in our AIxIA paper [2].

Infodemiology: Dermatological and Mental Conditions

Google COVID-19 Open Data: 400 health conditions, 4 countries (county-level in the US), weekly search frequencies for 2020-2023 normalised by NLP.

Weather stations in 1652 counties with and satellite images.

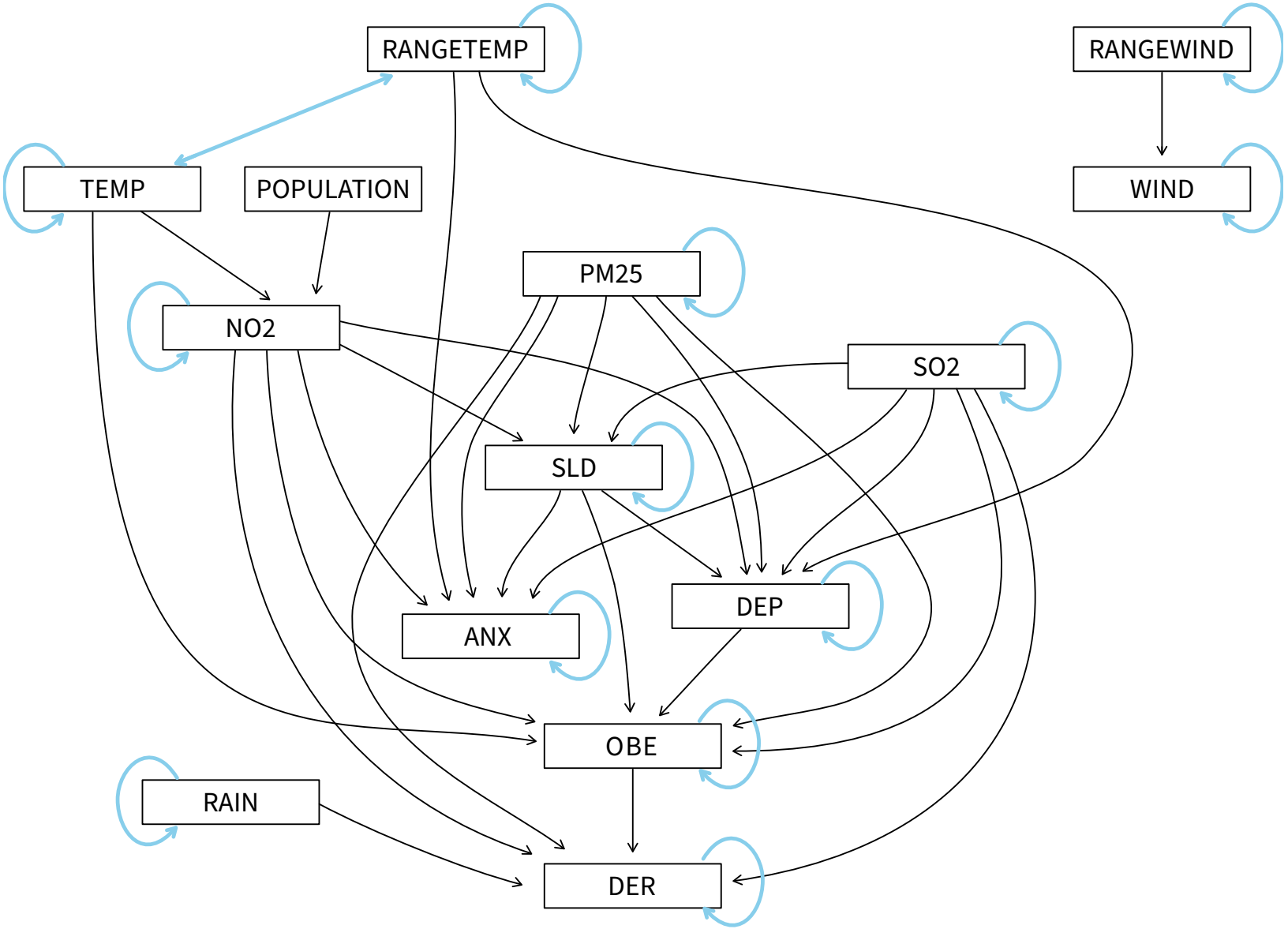
Monitoring stations in 1470 counties with hourly measurements of NOx, SOx, O3, PMx.

Socio-economic data at the population level to avoid confounding.

Key Issues: incomplete and heterogeneous state space-data (≈53k observations over ≈500 US counties and 134 weeks.)

Modelling: GLS, IRLS, node-averaged likelihood, model averaging.

Disregarding the State-Space Nature of the Data



Biased model learned with a dynamic network as in my previous work [25].

Failing on Most Counts

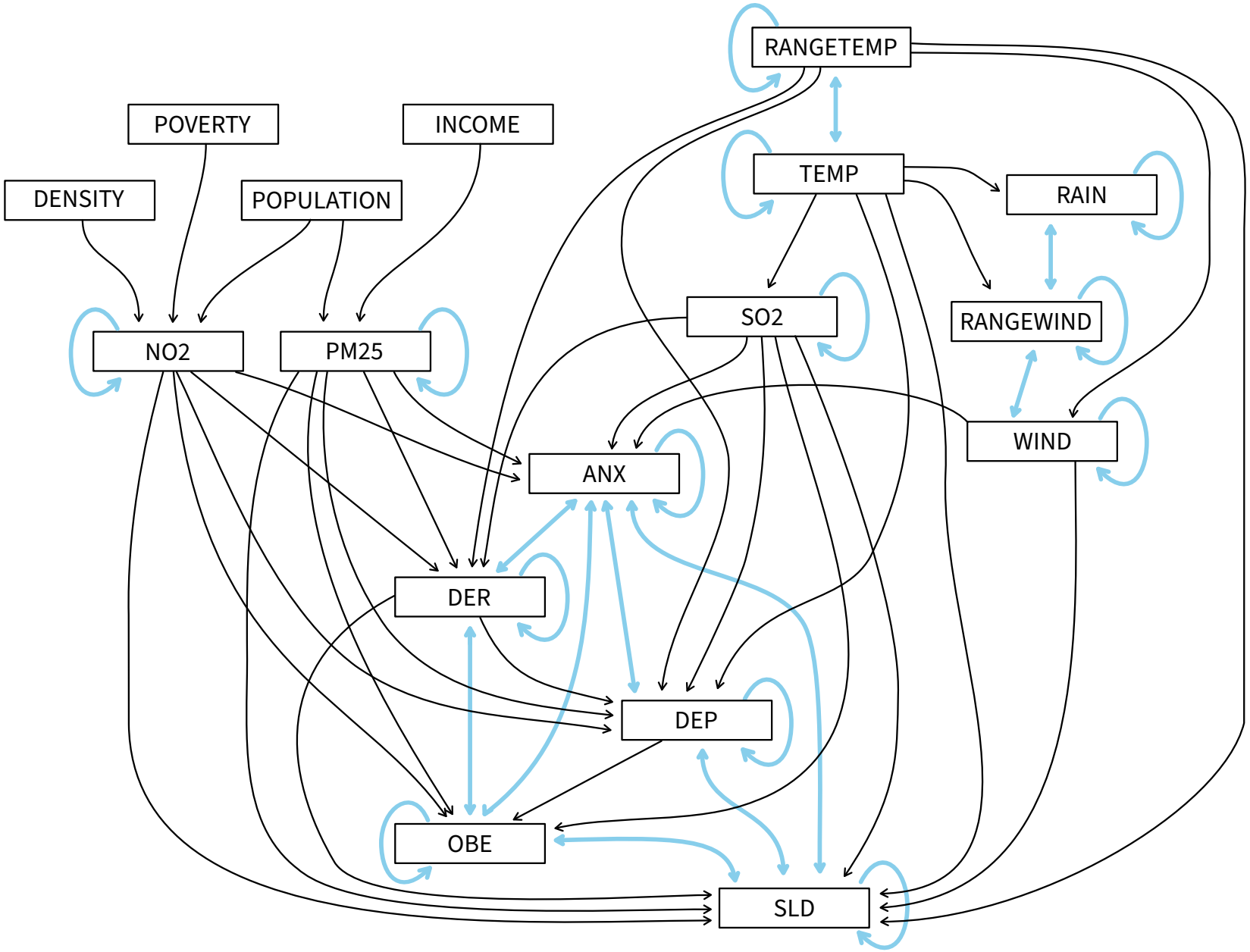
Residuals are reasonably free from autocorrelation, but they are full of spatial correlation and heteroscedastic!

	lag 1	lag 2	lag 3	lag 4	space	heteroscedastic
ANX	0.024	0.000	0.000	0.048	0.460	4×10^{-169}
DEP	0.016	0.000	0.000	0.000	0.325	1×10^{-212}
DER	0.032	0.000	0.000	0.000	0.754	0
OBE	0.000	0.000	0.000	0.000	0.563	6×10^{-100}
SLD	0.092	0.007	0.007	0.000	0.381	1×10^{-154}

Reality chooses a better model:

- Adding a spatial correlation matrix: BF = 81.59.
- Adding different standard errors for different states: BF = 25.31.
- Reducing sparsity to let known arcs in: BF = 72.15.

Modelling the State-Space Nature of the Data



Unbiased model that I am currently working to improve.

- ✓ Causal Discovery
- ✓ Here Comes Clinical Data
- ✓ Reality-Centric Causal Networks
- → Conclusions

Acknowledgements

Conclusions

- Causal networks are a principled and versatile tool to model clinical data in their complexity.
- Statistical and clinical experts are fundamental: using causal networks as a purely machine learning model is doomed to fail.
- State-space data, mixed variable types, missing values, population structure, non-stationarity: we can deal with them!
- There are technical challenges in translating clinical requirements and knowledge into causal discovery approaches, but we are actively working on them.

- ✓ Causal Discovery
- ✓ Here Comes Clinical Data
- ✓ Reality-Centric Causal Networks
- ✓ Conclusions
- → Acknowledgements

Acknowledgements

Peter Lucas

University of Twente

Alice Bernasconi

Alessio Zanga (funded by Roche)

Fabio Stella

Università degli Studi di Milano-Bicocca

Samir Salah

Delphine Kerob

L'Oréal, La Roche-Posay

My former student: Tjebbe Bodewes (University of Oxford).

That's all!

Happy to discuss in more detail.

References I

- L. Azzimonti, G. Corani, and M. Scutari.
  A Bayesian Hierarchical Score for Structure Learning from Related Data Sets.
  International Journal of Approximate Reasoning, 142:248–265, 2021.
- A. Bernasconi, A. Zanga, P. J. F. Lucas, M. Scutari, and F. Stella.
  Towards a Transportable Causal Network Model Based on Observational Healthcare Data.
  In *Proceedings of the 2nd Workshop on Artificial Intelligence for Healthcare, 22nd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2023)*, pages 67–82, 2023.
- E. Castillo, J. M. Gutiérrez, and A. S. Hadi.
  Expert Systems and Probabilistic Network Models.
  Springer, 1997.
- R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter.
  Probabilistic Networks and Expert Systems.
  Springer, 1999.
- M. Dhanakshirur, F. Laumann, J. Park, and M. Barahona.
  A Continuous Structural Intervention Distance to Compare Causal Graphs.
  https://arxiv.org/abs/2307.16452, 2023.

References II

- M. J. Druzdel and L. C. van der Gaag.
  Elicitation of Probabilities for Belief Networks: Combining Qualitative and Quantitative Information.
  In *UAI*, pages 141–148, 1995.
- M. J. Druzdel and L. C. van der Gaag.
  Building Probabilistic Networks: "Where Do the Numbers Come From?".
  IEEE Transactions on Knowledge and Data Engineering, 12(4):481–486, 200.
- M. Eigenmann, S. Mukherjee, and M. Maathuis.
  Evaluation of Causal Structure Learning Algorithms via Risk Estimation.
  Proceedings of Machine Learning Research, 124 (UAI):151–160, 2020.
- N. Friedman.
  Learning Belief Networks in the Presence of Missing Values and Hidden Variables.
  In *ICML*, pages 125–133, 1997.
- N. Friedman.
  The Bayesian Structural EM Algorithm.
  In *UAI*, pages 129–138, 1998.

References III

- L. Henckel, T. Würtzen, and S. Weichwald.
  Adjustment Identification Distance: A gadjid for Causal Structure Learning.
  In *UAI*, 2024.
- J. Huang, X. Guo, K. Yu, F. Cao, and J. Liang.
  Towards Privacy-Aware Causal Structure Learning in Federated Setting.
  IEEE Transactions on Big Data, 9(6):1525–1535, 2023.
- G. W. Imbens.
  Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics.
  Journal of Economic Literature, 58(4):1129–1159, 2020.
- D. Koller and N. Friedman.
  Probabilistic Graphical Models: Principles and Techniques.
  MIT Press, 2009.
- S. Lee, J. Correa, and E. Bareinboim.
  General Transportability – Synthesizing Observations and Experiments from Heterogeneous Domains.
  AAAI Conference on Artificial Intelligence, 34(6):10210–10217, 2020.

References IV

- Y. Liu and A. C. Constantinou.
  Greedy Structure Learning from Data that Contain Systematic Missing Values.
  Machine Learning, 111:3867–3896, 2022.
- A. M. K. Mamaghan, P. Tigas, K. H. Johansson, Y. Gal, Y. Annadani, and S. Bauer.
  Challenges and Considerations in the Evaluation of Bayesian Causal Discovery.
  https://arxiv.org/abs/2406.03209, 2024.
- O. Mian, D. Kaltenpoth, and M. Kamp.
  Regret-based Federated Causal Discovery.
  Proceedings of Machine Learning Research, 185 (KDD):61–69, 2022.
- O. Mian, D. Kaltenpoth, M. Kamp, and J. Vreeken.
  Nothing but Regrets — Privacy-Preserving Federated Causal Discovery.
  Proceedings of Machine Learning Research, 206 (AISTATS):8263–8278, 2023.
- K. Mohan and J. Pearl.
  Graphical Models for Processing Missing Data.
  Journal of the American Statistical Association, 116(534):1023–1037, 2018.

References V

- J. Pearl and D. Mackenzie.
  The Book of Why: the New Science of Cause and Effect.
  Basic Books, 2018.
- J. Peters and P. Bühlmann.
  Structural Intervention Distance (SID) for Evaluating Causal Graphs.
  Neural Computation, 27:771–799, 2015.
- M. Scutari and J.-B. Denis.
  Bayesian Networks with Examples in R.
  Chapman & Hall, 2nd edition, 2021.
- M. Scutari, C. E. Graafland, and J. M. Gutiérrez.
  Who Learns Better Bayesian Network Structures: Accuracy and Speed of Structure Learning Algorithms.
  International Journal of Approximate Reasoning, 115:235–253, 2019.
- M. Scutari, D. Kerob, and S. Salah.
  Inferring Skin-Brain-Skin Connections from Infodemiology Data Using Dynamic Bayesian Networks.
  Scientific Reports, 14:10266, 2024.

References VI

- M. Scutari and M. Malvestio.
  The Pragmatic Programmer for Machine Learning: Engineering Analytics and Data Science Solutions.
  Chapman & Hall, 2023.
- M. Scutari, C. Marquis, and L. Azzimonti.
  Using Mixed-Effect Models to Learn Bayesian Networks from Related Data Sets.
  Proceedings of Machine Learning Research (PGM 2022), 186:73–84, 2022.
- P. Spirtes, C. Glymour, and R. Scheines.
  Causation, Prediction, and Search.
  MIT Press, 2000.
- A. Zanga, A. Bernasconi, P. Lucas, H. Pijnenborg, C. Reijnen, M. Scutari, and F. Stella.
  Risk Assessment of Lymph Node Metastases in Endometrial Cancer Patients: A Causal Approach.
  In *Workshop on Artificial Intelligence for Healthcare, AIxIA*, pages 1–15, 2022.
- A. Zanga, A. Bernasconi, P. J. F. Lucas, H. Pijnenborg, C. Rejinen, M. Scutari, and F. Stella.
  Causal Discovery with Missing Data in a Multicentric Clinical Study.
  In *AIME*, pages 40–44, 2023.