Bayesian and Causal Networks for Clinical and Epidemiological Data

Concepts, Implementation and Interpretation

Marco Scutari
scutari@bnlearn.com

Dalle Molle Institute for Artificial Intelligence

October 22, 2025

Bayesian Networks

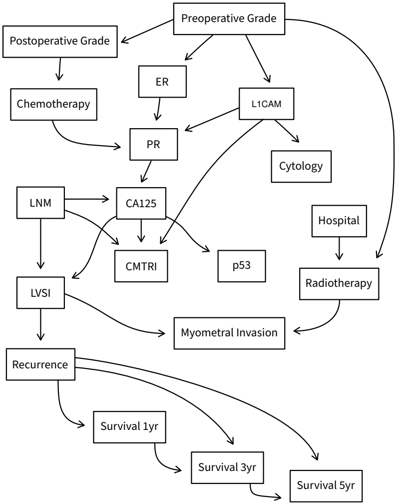
A Bayesian network [17, BN] is defined by:

- a directed acyclic graph (DAG) $\mathcal{G}$ where each node maps to a random variable $X_i \in \mathbf{X}$;
- a probability distribution $\mathrm{P}(\mathbf{X}; \Theta)$, factorising into smaller local distributions following the arcs in $\mathcal{G}$.

The DAG $\mathcal{G}$ expresses the conditional independencies among the X_i through graphical separation, leading to:

$$\mathrm{P}(\mathbf{X}) = \prod_{i=1}^{N} \mathrm{P}\left(X_i \mid \Pi_{X_i}; \Theta_{X_i}\right),$$

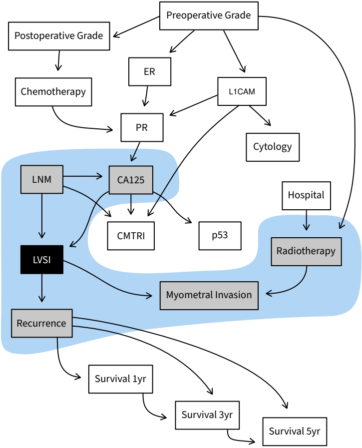
$$\Pi_{X_i} = \{\text{parents of } X_i\}.$$



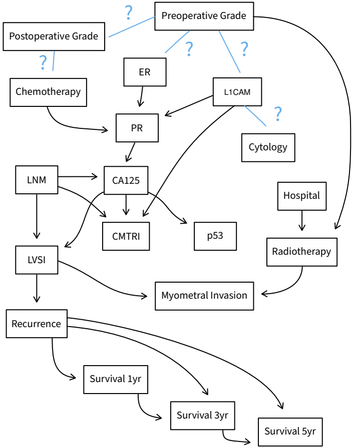
Endometrial cancer, AIME 2023 [21].

Key Properties on Bayesian Networks

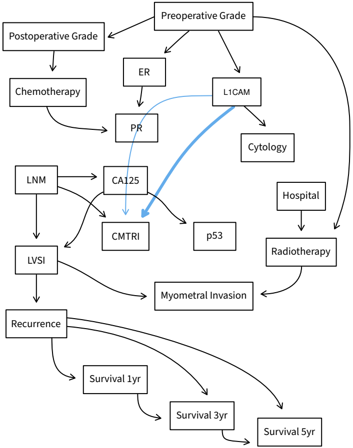
Markov Blankets: Feature Selection



Equivalence Classes: Identifiability



Path Analysis: Mediation



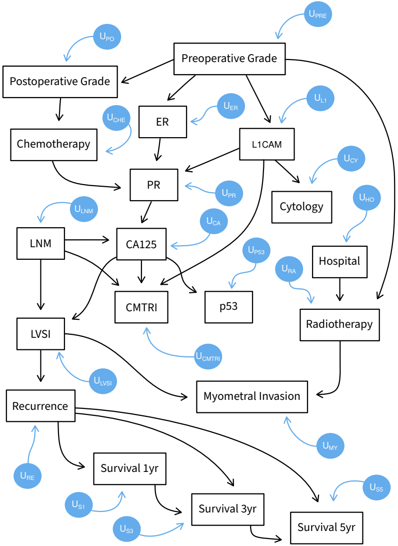
A structural causal model [1, SCM] is defined by:

- some endogenous (deterministic) variables $\mathbf{X}$;
- some exogenous (stochastic) variables $\mathbf{U}$;
- a DAG $\mathcal{G}$ with nodes mapping to $X_i, U_i \in \mathbf{X} \cup \mathbf{U}$;
- one structural equation $f_i(\cdot)$ for each X_i.

Then the X_i are defined by the structural equations:

$$X_i := f_i(\Pi_{X_i}, U_i), \Pi_{X_i} \subset \mathbf{X}.$$

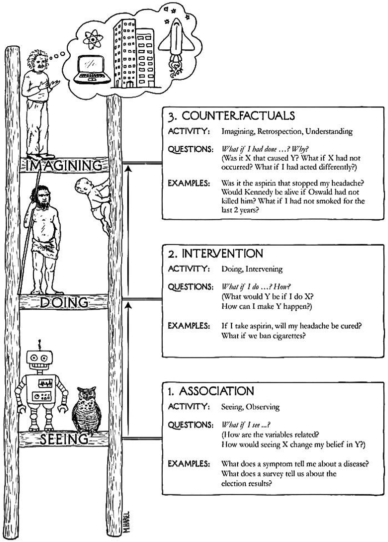
Also structural equations models (SEMs) or causal BNs.



They Look Similar, But They Have Completely Different Semantics

Bayesian Networks	Structural Causal Models
Probabilistic independence relationships are symmetric: if $X_i \perp\!\!\!\perp X_j$, then $X_j \perp\!\!\!\perp X_i$.	Causes and effects are not symmetric: treating symptoms does not cure the underlying disease.
$\mathbf{X}_i$ as generalised linear models [9, 17]: $g(X_i) = \eta(\Pi_{X_i}) + \varepsilon_i$ main effects only.	X_i as additive noise models [15, ANMs]: $f_i(\Pi_{X_i}, U_i) \approx f_i(\Pi_{X_i}) + U_i$, main effects only, fitted with regularised regressions.
Heavily rely on sufficiency (assume no latent confounders).	Heavily rely on causal sufficiency (assume no latent confounders).

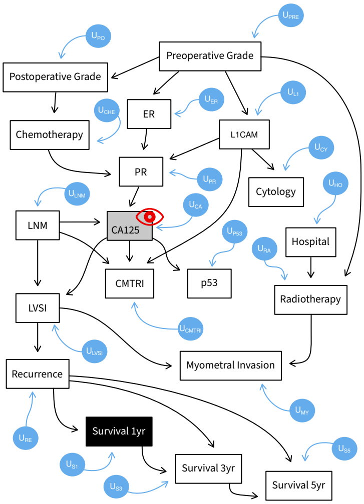
The Ladder of Causation



These network models can be used at various levels of sophistication, the step in the Ladder of Causation [14].

- BNs support levels 1–2 (*Association* and *Intervention*).
- SCMs support also support level 3 (*Counterfactuals*).
- SCMs can express latent confounding by linking exogenous variables with multiple endogenous variables. BNs cannot do that well, ad-hoc hacks are needed [8].

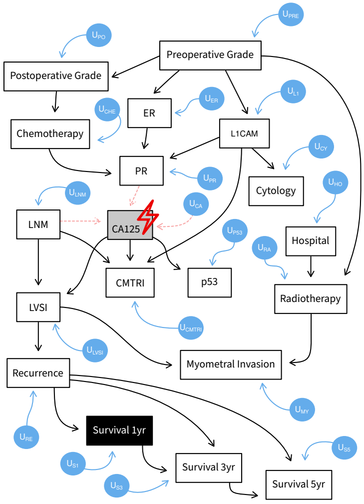
INTERVENTIONS



Just observing

$$P(\text{Survival 1y} \mid \text{CA125} = c) \neq P(\text{Survival 1y} \mid \text{do}(\text{CA125} = c))$$

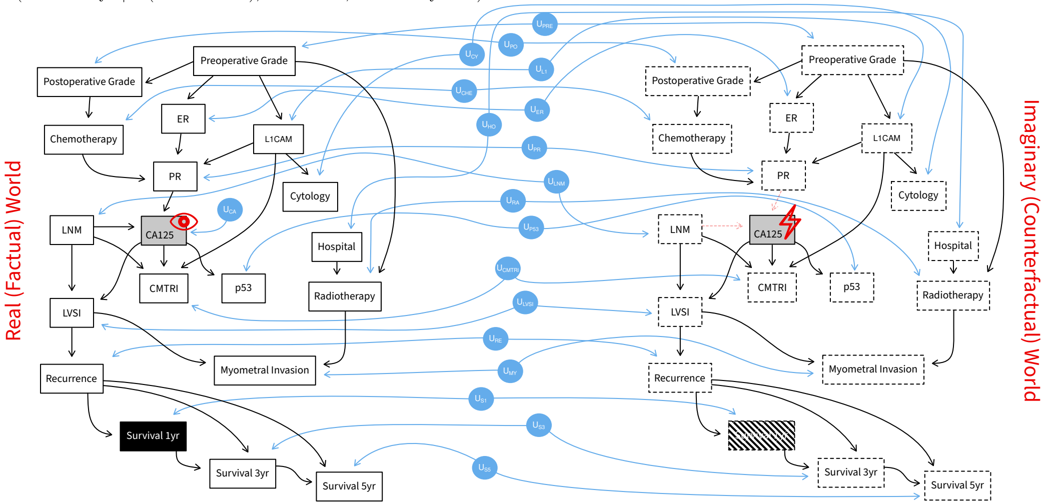
Actively acting



Manipulate variables to set them to a single value ("hard") or a distribution ("soft").

COUNTERFACTUAL

$$P(\text{Survival 1yr}' \,|\, do(\text{CA125}' = c'), \text{CA125} = c, \text{Survival 1yr} = s)$$



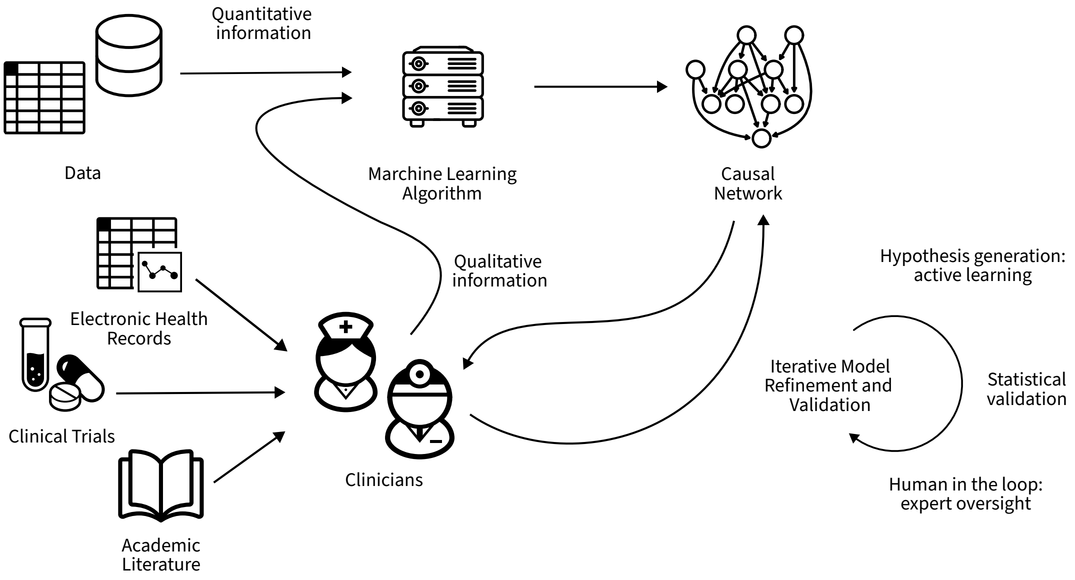
Imagine a world that did not actually occur, but could have: what would have been different?

Bayesian Network Inference

Bayesian network inference automates quantitative assessments of probabilistic and causal inference statements in BNs and SCMs. You ask queries on some event given some evidence you have observed, and you get the answer.

Exact Inference	Approximate Inference
Based on the junction tree algorithm.	Based on MC or MCMC sampling.
Heavier computational cost.	More scalable.
Akin to symbolic computations. They provide exact answers.	Like posterior inference, answers have simulation noise.

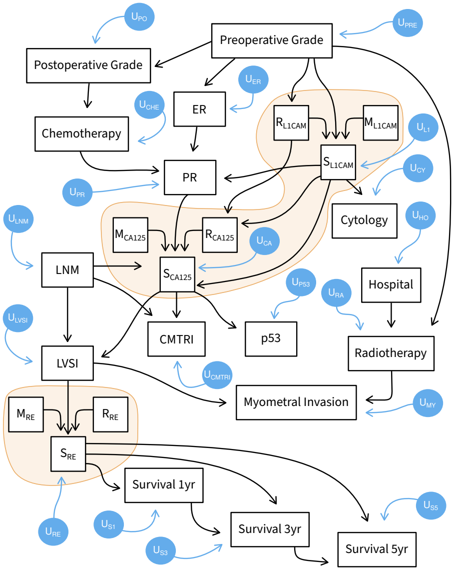
Inference for Refinement and Validation



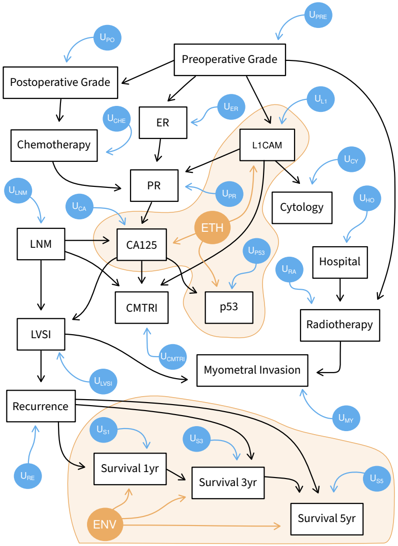
Structural Causal Models vs Experimental Design

SCMs	Trial Design	Nuance
DAGs as prior specification.	Trial design as prior.	DAG encodes prior causal knowledge; informs trial design.
Intervention.	Treatment assignment.	Bayesian belief updates post-trial.
Backdoor adjustment.	Randomisation.	Information on known confounders.
Colliders.	Stratification.	Collider bias, sub-populations.
Causal Effects.	Treatment Effects.	Bayesian effect estimates.
Transportability.	External validity.	Generalising to different populations.

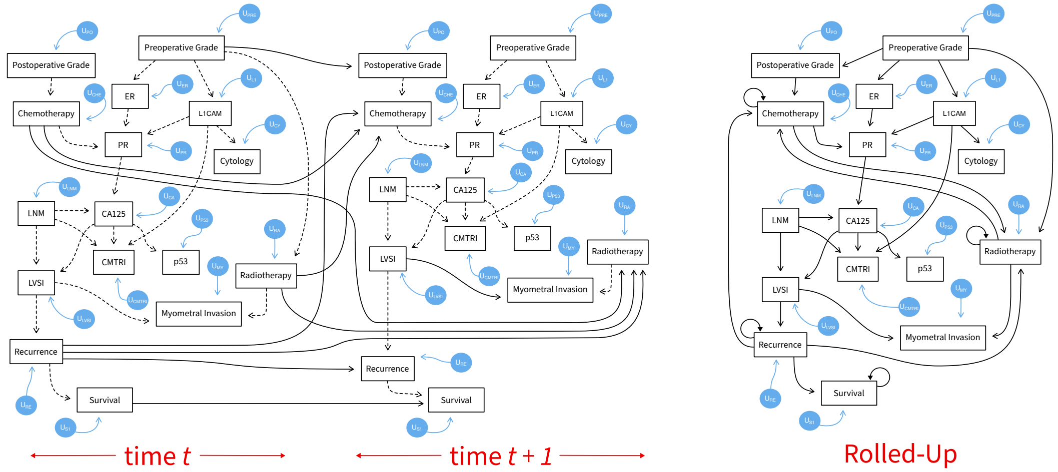
- All missing data patterns identified by Rubin [12] have a graphical representation in augmented SCMs called missingness graphs [13].
- Each partially-observed X_i is split into three nodes: M_{X_i} (complete, but latent), S_{X_i} (incomplete, but observed) and R_{X_i} (binary missingness indicators). $R_{X_i} \rightarrow S_{X_i} \leftarrow M_{X_i}$.
- The parents of R_i determine the pattern: MCAR (no parents), MAR (only fully-observed X_js), MNAR (partially-observed or latent variables).
- Expectation-Maximisation is most common [16].



- Sampling bias has a graphical representation in augmented SCMs called selection diagrams [10].
- Treated like a latent confounder: each source of bias is represented as an exogenous variable liked to several endogenous variables.
- G-Transportability of study findings from a source population to a different target population, especially when the study is an RCT and the target population is observational data. Example: Breast cancer + CVD [4, 3].



Modelling Time



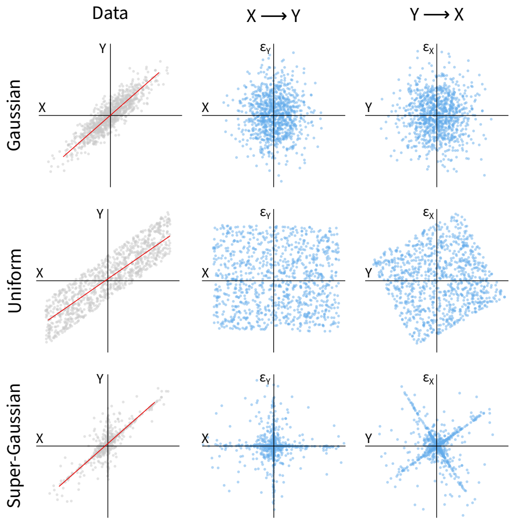
Vector auto-regressive time series, Kalman filters, hidden Markov models can all be cast as dynamic SCMs [16]. Continuous-time SCMs are also possible [5].

Class	Algorithms	Approach
Constraint-Based	PC [7] RFCI [8]	Conditional independence tests to prune the search space into an equivalence class.
Score-Based	LiNGAM [19] GES [6]	Navigating the DAG space to find the optimal SCM, leveraging the asymmetries in residuals and/or regularisation.
Differentiable	NOTEARS [22] DAGMA [2]	Minimising residual variance via gradient descent, with constraints for SCM acyclicity and sparsity.

Many assumptions ensure that causal directions are correctly identified:

- Data from multiple environments [10].
- Data with different interventions [11]
- Non-Gaussianity [19].
- Heteroscedasticity [20].
- Spatial correlations, non-IID data [18].

Randomisation also works as usual.



That's all!

Happy to discuss in more detail.

scutari@bnlearn.com

References I

[1] E. Bareinboim, J. D. Correa, D. Ibeling, and T. Icard.
On Pearl's Hierarchy and the Foundations of Causal Inference, pages 509–56.
Association for Computing Machinery, 2021.

[2] K. Bello, B. Aragam, and P. Ravikumar.
DAGMA: Learning Dags via M-Matrices and a Log-Determinant Acyclicity Characterization.
Advances in Neural Information Processing Systems, 35:8226–8239, 2022.

[3] A. Bernasconi, A. Zanga, P. J. F. Lucas, M. Scutari, S. Di Cosimo, M. C. De Santis, E. La Rocca, P. Baili, I. Cavallo, P. Verderio, C. M. Ciniselli, S. Pizzamiglio, A.Blanda, P. Perego, P. Vallerio, F. Stella, A. Trama, and the Ada Working Group.
From Real-World Data to Causally Interpretable Models: A Bayesian Network to Predict Cardiovascular Diseases in Adolescents and Young Adults with Breast Cancer.
Cancers, 16(21):3634, 2024.

[4] A. Bernasconi, A. Zanga, P. J. F. Lucas, M. Scutari, A. Trama, and F. Stella.
A Causal Network Model to Estimate the Cardiotoxic Effect of Oncological Treatments in Young Breast Cancer Survivors.
Progress in Artificial Intelligence, 2024.

References II

[5] A. Bregoli, M. Scutari, and F. Stella.
A Constraint-Based Algorithm for the Structural Learning of Continuous-Time Bayesian Networks.
International Journal of Approximate Reasoning, 138:105–122, 2021.

[6] D. M. Chickering.
Optimal Structure Identification with Greedy Search.
Journal of Machine Learning Research, 3:507–554, 2002.

[7] D. Colombo and M. H. Maathuis.
Order-Independent Constraint-Based Causal Structure Learning.
Journal of Machine Learning Research, 15:3921–3962, 2014.

[8] D. Colombo, M. H. Maathuis, M. Kalisch, and T. S. Richardson.
Learning High-Dimensional Directed Acyclic Graphs with Latent and Selection Variables.
The Annals of Statistics, 40(1):294–321, 2012.

[9] G. Kratzer, F. Lewis, A. Comin, M. Pittavino, and R. Furrer.
Additive Bayesian Network Modeling with the R Package Abn.
Journal of Statistical Software, 105(8), 2023.

- [10] 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.
- [11] A. Li, A. Jaber, and E. Bareinboim.
  Causal Discovery from Observational and Interventional Data Across Multiple Environments.
  In *Advances in Neural Information Processing Systems*, volume 36, pages 16942–16956, 2023.
- [12] R. J. A. Little and D. B. Rubin.
  Statistical Analysis with Missing Data.
  Wiley, 3rd edition, 2019.
- [13] K. Mohan and J. Pearl.
  Graphical Models for Processing Missing Data.
  Journal of the American Statistical Association, 116(534):1023–1037, 2018.
- [14] J. Pearl and D. Mackenzie.
  The Book of Why: the New Science of Cause and Effect.
  Basic Books, 2018.

References IV

[15] J. Peters, J.M. Mooij, D. Janzing, and B. Schölkopf.
Causal Discovery with Continuous Additive Noise Models.
Journal of Machine Learning Research, 15(1):2009–2053, 2014.

[16] M. Scutari.
Bayesian Network Models for Incomplete and Dynamic Data.
Statistica Neerlandica, 74(3):397–419, 2020.

[17] M. Scutari and J.-B. Denis.
Bayesian Networks with Examples in R.
Chapman & Hall, 2nd edition, 2021.

[18] M. Scutari, D. Kerob, J. Krutmann, and S. Salah.
Causal Networks of Infodemiological Data: Modelling Dermatitis.
In *Proceedings of the 23rd International Conference on Artificial Intelligence in Medicine (AIME25)*, pages 397–407, 2025.

[19] S. Shimizu, P. O. Hoyer, A. Hyvärinena, and ' A. Kerminen.
A Linear Non-Gaussian Acyclic Model for Causal Discovery.
J. Mach. Learn. Res., 7(72):2003–2030, 2006.

[20] S. Xu, O. A. Mian, A. Marx, and J. Vreeken.
Inferring Cause and Effect in the Presence of Heteroscedastic Noise.
Proceedings of Machine Learning Research (ICML), 162:24615–24630, 2022.

[21] 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 *Proceedings of the 21st International Conference on Artificial Intelligence in Medicine (AIME23)*, pages 40–44, 2023.

[22] X. Zheng, B. Aragam, P. K. Ravikumar, and E. P. Xing.
DAGs with No Tears: Continuous Optimization for Structure Learning.
Advances in Neural Information Processing Systems, 31:0472–9483, 2018.

Large Scale Causal Modelling to Identify Adults at Risk for Combined and Common Variable Immunodeficiencies

npj Digital Medicine

Marco Scutari
scutari@bnlearn.com

Dalle Molle Institute for Artificial Intelligence

October 22, 2025

npj | digital medicine

Article

https://doi.org/10.1038/s41746-025-01761-5

Large scale causal modeling to identify adults at risk for combined and common variable immunodeficiencies

Giorgos Papanastasiou[1,8], Marco Scutari[2,8], Raffi Tachdjian[3,4], Vivian Hernandez-Trujillo[5], Jason Raasch[6], Kaylyn Billmeyer[1,7], Nikolay V. Vasilyev[1] & Vladimir Ivanov[1]

npj Digital Medicine | (2025)8:361

Notebooks: https://www.bnlearn.com/research/pfizer24

Pfizer

- Goal: unravelling the interplay between clinical diagnosis codes linked to combined immunodeficiencies (CID) and common variable immunodeficiencies (CVID).
- Main variables: clinical history with ICD codes from 4 different cohorts, each with its own inclusion/exclusion criteria. ICD codes are transformed into phenotypes.
- Possible confounders: hopefully none, given we condition on all clinical history and we use US nation-wide data. Matched cases/controls using propensity scores from demographics.
- Size: $\approx 800/800/2.3k/20k$ observations and $\approx 550/550/400/300$ variables in cohorts 1–4.
- Missing values: none!

Following up on a previous effort based on a custom deep-learning architecture [1].

Preprocessing: Data and ICD Codes

Discovery

Cohort 1 (Raw Data)
797 CID cases with pneumonia
797 controls with pneumonia

Replication

Cohort 2 (Raw Data)
797 CID cases with pneumonia
797 random controls, with or without pneumonia

Cohort 3 (Raw Data)
2,312 CID cases with (N=797) or without pneumonia
2,312 random controls, with or without pneumonia

Cohort 4 (Raw Data)
19,924 CID/CVID cases with (N=2,350) or without pneumonia
19,924 random controls, with or without pneumonia

Data Preparation

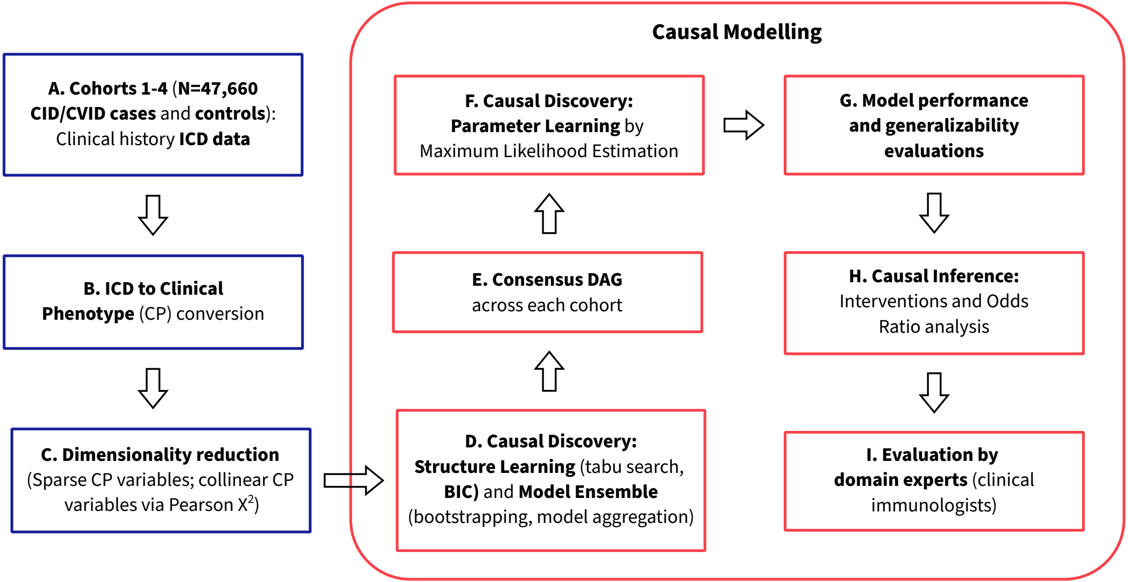
Feature (ICD code) Selection
- Optum data Table combinations
- Patient demographics and medical claims indentification

Data Engineering
- ICD-9 to ICD-10 code extraction and initial mapping
- Hierarchical ICD code mapping

Futher Pre-Processing
- Cleaning confounders
- Missingness assessment
- One-hot encoding/Embeddings

Propensity Score Matching
- Matched controls across all cohorts

DATA ANALYSIS



The Role of Clinical Immunologists

- Telling phenotypes apart.
  - With finite sample sizes and so many variables, some end up being numerically equivalent (pair-wise association p-value ≈ 0).
  - They explain CID/CVID equally well.
  - Which ones make the most (clinical) sense to keep?
- Validating the causal networks:
  - Reviewing the consensus causal networks from each cohort.
  - Confirming that the networks really are dense.
  - Assessing the Markov blankets of CID/CVID, which identify direct precursors of CID/CVID diagnoses as parents.

We consulted them independently, effectively handling their observations as an ensemble expert model.

That's all!

Happy to discuss in more detail.

scutari@bnlearn.com

G. Papanastasiou, G. Yang, D. I. Fotiadis, N. Dikaios, C. Wang, A. Huda, L. Sobolevsky, J. Raasch, E. Perez, G. Sidhu, and D. Palumbo.
Large-Scale Deep Learning Analysis to Identify Adult Patients at Risk for Combined and Common Variable Immunodeficiencies.
Communications Medicine, 3(1):189, 2023.

Causal Networks of Infodemiological Data: Modelling Dermatitis

International Conference on Artificial Intelligence in Medicine

Marco Scutari
scutari@bnlearn.com

Dalle Molle Institute for Artificial Intelligence

October 22, 2025

Causal Networks of Infodemiological Data: Modelling Dermatitis

Marco Scutari[1(✉)], Samir Salah[2], Delphine Kerob[2,3], and Jean Krutmann[4,5]

R. Bellazzi et al. (Eds.): AIME 2025, LNAI 15734, pp. 397–407, 2025.
https://doi.org/10.1007/978-3-031-95838-0_39

Notebooks: http://www.bnlearn.com/research/aime25

IUF
LEIBNIZ-INSTITUT FÜR UMWELT-MEDIZINISCHE FORSCHUNG

LA ROCHE-POSAY
LABORATOIRE PHARMACEUTIQUE

- Goal: understanding the effect of pollution and changing weather patterns on mental conditions and dermatitis, and the cascading effect of mental conditions on dermatitis.
- Main variables: 3 pollutants (NO_2, SO_2, $PM_{2.5}$), 3 mental conditions (anxiety, depression, sleep disorders), obesity, dermatitis, weather patterns (temperatures, wind speed, precipitations; both mean and spread).
- Possible confounders: education level, unemployment, income, household size and population density.
- Size: $\approx$53k observations over $\approx$500 US counties and 134 weeks.
- Missing values: between 0% (the conditions) and 55% (pollutants).

Following up on a previous infodemiology study [3].

Data Sources: Google Trends, NOAA, EPA, US Census

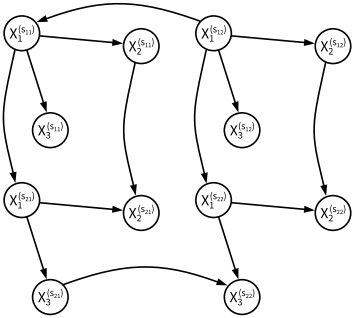
Google COVID-19 Open Data: 400 health conditions, 4 countries (county-level in the US), weekly search frequencies for 2020-2023 normalised by NLP.

Weather stations
in 1652 counties with and satellite images.

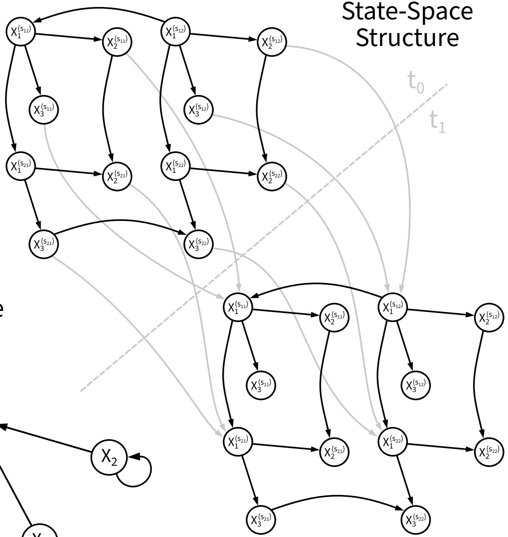
Monitoring stations
in 1470 counties with hourly measurements of NOx, SOx, O3, PMx.

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

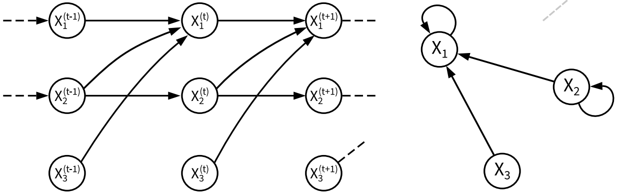
Spatial Structure



State-Space Structure



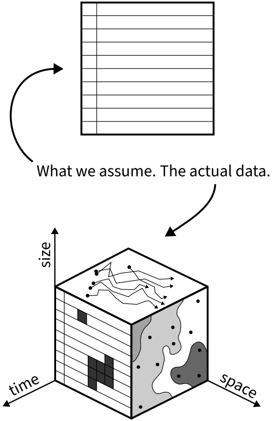
Temporal Structure (dynamic BNs)



Learned a dynamic network encoding a first-order VAR process:

$$X_{it} = f_i(\Pi_{X_{it}} \boldsymbol{\beta}_{it}) + \varepsilon_{it}, \qquad \mathrm{COV}(\varepsilon_{it}) = \mathbf{w}_{it}^{\mathrm{T}} \boldsymbol{\Sigma}_i(\mathbf{L}; \xi_i) \mathbf{w}_{it}.$$

- $\boldsymbol{\Sigma}_i(\mathbf{L}; \xi_i)$ models spatial correlation via generalised least squares (GLS); location $\mathbf{L}$ and correlation decay ξ_i.
- The $\mathbf{w}_{it}$ handle
  - heteroscedasticity, via iteratively reweighted least squares (IRLS);
  - missing values, with 0-1 weights like the PNAL score [1] (if MCAR) or inverse-probability weights like HC-aIPW [2] (if MAR or MNAR).



Model averaging: bagging with data-driven threshold [4].

Causal Inference: What Conclusions Can We Draw?

- What is the relative impact of the direct risk factors?

  ANX (0.574), NO_2 (0.339), OBE (0.077), $PM_{2.5}$, RANGETEMP, SO_2 (0.01).

- What proportion of pollution effects is mediated?

  $PM_{2.5}$, NO_2 and SO_2 change by 0.54x, 0.93x and 0.56x.

- What proportion of weather effects is mediated?

  TEMP/RANGETEMP, WIND/RANGEWIND, RAIN change by 0.29x, 0.38x, 0.02x

- What would be the impact of tightening environmental regulations?

  $PM_{2.5}$ $12 \rightarrow 9\mu g/m^3$ for 1 year: -18% DER. $PM_{2.5}$ $12 \rightarrow 8\mu g/m^3$: -21% DER.

- How long must a cold spell last before dermatitis increases?

  DER +5% after 4 weeks.

Data Analysis

1. Data fusion and preprocessing.
2. Causal discovery assuming IID data.
3. Statistical validation of the residuals.
4. Causal discovery with a spatial correlation structure.
5. Check the residuals, Bayes factors.
6. Causal discovery with spatial correlation + heterogeneity.
7. Check the residuals, Bayes factors, imposed sparsity level.
8. Predictive accuracy assessment.

That's all!

Happy to discuss in more detail.

scutari@bnlearn.com

References I

- T. Bodewes and M. Scutari.
  Learning Bayesian Networks from Incomplete Data with the Node-Averaged Likelihood.
  International Journal of Approximate Reasoning, 138:145–160, 2021.
- Y. Liu and A. C. Constantinou.
  Greedy Structure Learning From Data That Contain Systematic Missing Values.
  Machine Learning, 111(10):3867–3896, 2022.
- M. Scutari, D. Kerob, and S. Salah.
  Inferring Skin-Brain-Skin Connections from Infodemiology Data Using Dynamic Bayesian Networks.
  Scientific Reports, 14:10266, 2024.
- M. Scutari and R. Nagarajan.
  On Identifying Significant Edges in Graphical Models of Molecular Networks.
  Artificial Intelligence in Medicine, 57(3):207–217, 2013.