Disclosures

- Cytel Inc. provided registration fees for all presenters
- Alind Gupta is employed by Cytel Inc.
- Ryan Walton is an employee and shareholder at AstraZeneca

Presenters

- **Alind Gupta, PhD**
  - Machine learning researcher, Cytel Inc., Toronto, Canada
- **Winson Y. Cheung, MD, MPH**
  - Professor of Medicine, University of Calgary
  - Principal Director, Oncology Outcomes (O2), Calgary, Alberta, Canada
- **Ryan N. Walton, MPH**
  - Director, Medical Evidence, AstraZeneca Canada
- **Nicholas Mitsakakis, MSc, PhD, P. Stat.**
  - Adjunct Lecturer, Dalla Lana School of Public Health, University of Toronto
  - Instructor, Master of Biotechnology, University of Toronto Mississauga
  - Adjunct Professor, Rady Faculty of Health Sciences, University of Manitoba
  - Investigator, THETA Collaborative

Machine Learning Prospects and Pitfalls in Clinic

Winson Y. Cheung, MD, MPH
Professor of Medicine, University of Calgary
Principal Director, Oncology Outcomes (O2)
Calgary, Alberta, Canada

Machine Learning to Clinicians

- 5 years ago…
  - “My iPad”
  - “Another decision aid or tool”
  - “Anything and everything that improves the clinic efficiency and enhances the clinic effectiveness”
  - “Is it artificial intelligence?”
  - “You mean IBM Watson?”

Machine Learning to Clinicians

- Present day…
  - “A machine that does my day job so that I can relax”
  - “A machine that gets faster and smarter over time”
  - “A machine that sorts through lots of complex data”
  - “A machine that tells me what I need to do or should do for a specific patient”
  - “A machine that addresses my clinical questions more quickly and accurately”

Clinical Interest in Machine Learning

- RWE = Real-world evidence
  - Curation, analysis, of data from sources from outside of conventional clinical trials
  - Clinical trials remain the gold standard for evaluating efficacy, but real-world data are necessary to inform “effectiveness”
  - Messy and variable

- Biomarkers and genomics data
  - Important for treatment section and targeted therapy
  - Complex

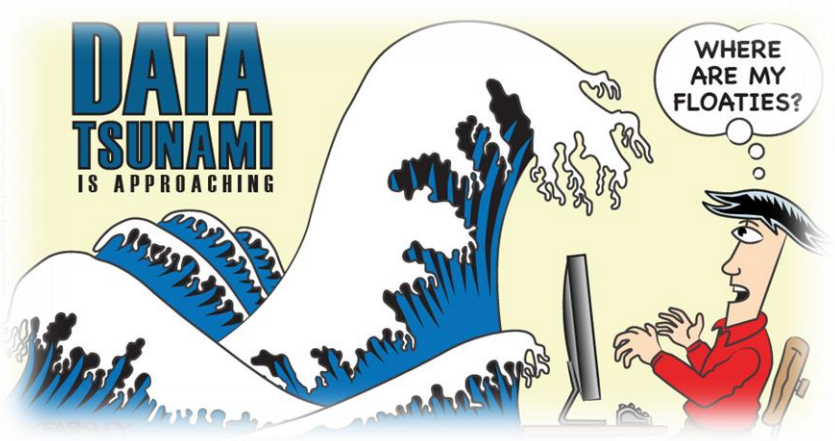
Clinical Interest in Machine Learning

- Efforts to liberate and federate of health datasets in certain jurisdictions
- Broad implementation of electronic health records in more clinical settings
- Emphasis on patient-reported outcome measures
- Digitized medical images
- Wider use of phones and tablets
- Leveraging wearable sensor technologies

⇨ More and more data

Clinical Interest in Machine Learning

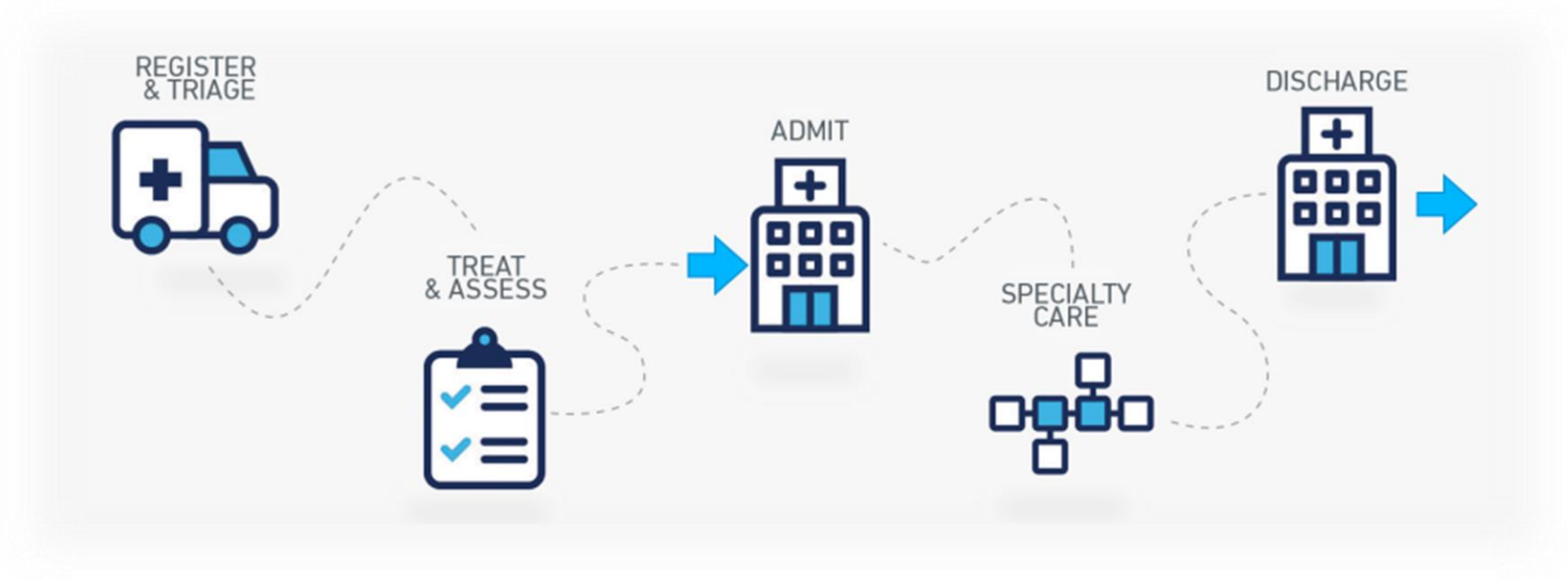
- Delivering the highest quality clinical care is increasingly complicated and challenging:
  - Improved access to medications
    - More choices
    - More providers
    - More toxicities
  - Longer survival
    - Different preferences
  - Patient engagement
    - Different information



Potential Clinical Applications

- Diagnose conditions
- Estimate risk
- Select optimal treatments

⇨ Facilitate clinical decision making



1. Diagnose Conditions

- Example:
  - Early and accurate diagnosis of suspicious skin lesions could be aided by smartphone cameras with machine-learning based apps
  - Allow patients to conveniently perform their own regular skin checks and transmit results to specialists who may not be easily accessible
  - Program is trained on dermo-scopy images of different skin conditions and tested against dermatologists using biopsy-proven examples

1. Diagnose Conditions

- Benefits
  - Rural and remote patients with care closer to home
  - Patient engagement and self-efficacy
  - Aligns with “Choosing Wisely” recommendations

2. Estimate risk

- Example:
  - Electronic health record data from previous admissions across different hospitals were used to train machine-learning algorithms to predict in-hospital mortality, long length of stay, discharge diagnoses 24 hours after admission, and 30-day unplanned readmission risk at discharge
  - Outperforms conventional risk-prediction models that rely on a minimum number of variables

2. Estimate risk

- Benefits:
  - Helps with the development of clinical care pathways and streamlines the triage process
  - Facilitates operational planning (e.g. staffing, resource allocation, bed assignments)

3. Select Optimal Treatments

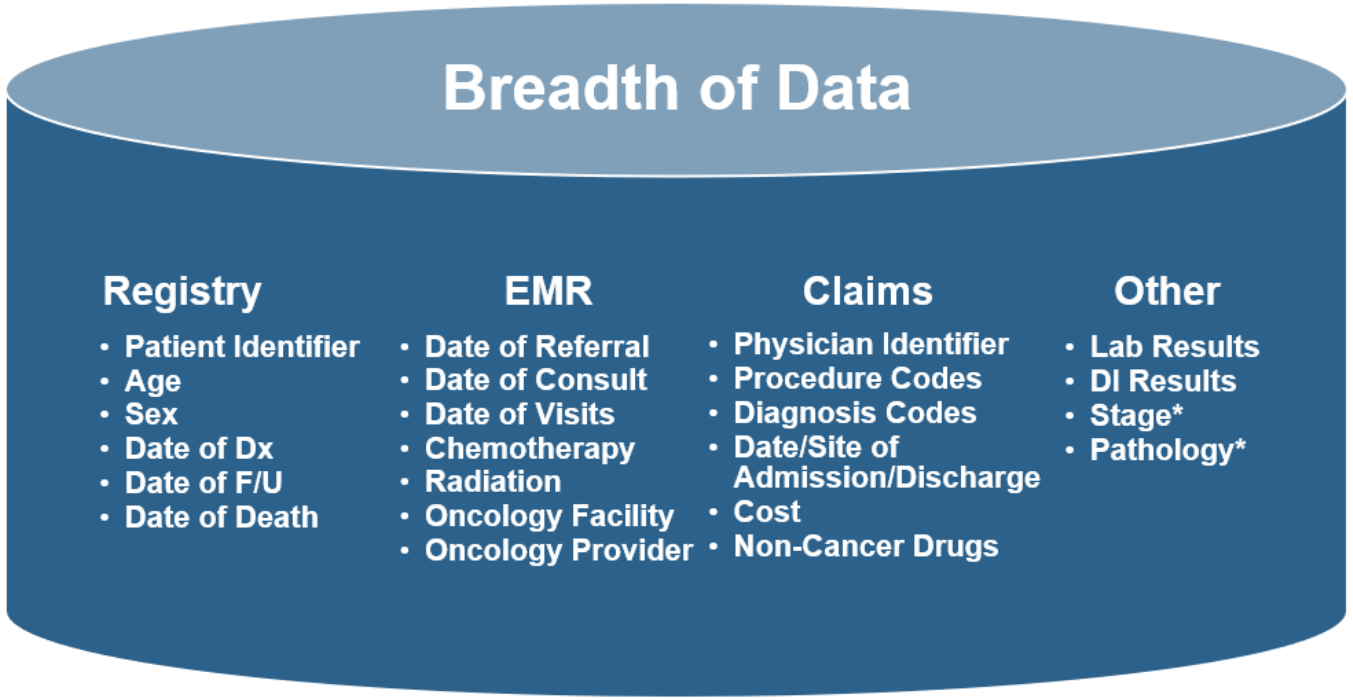
- Diagnose conditions
- Estimate risk

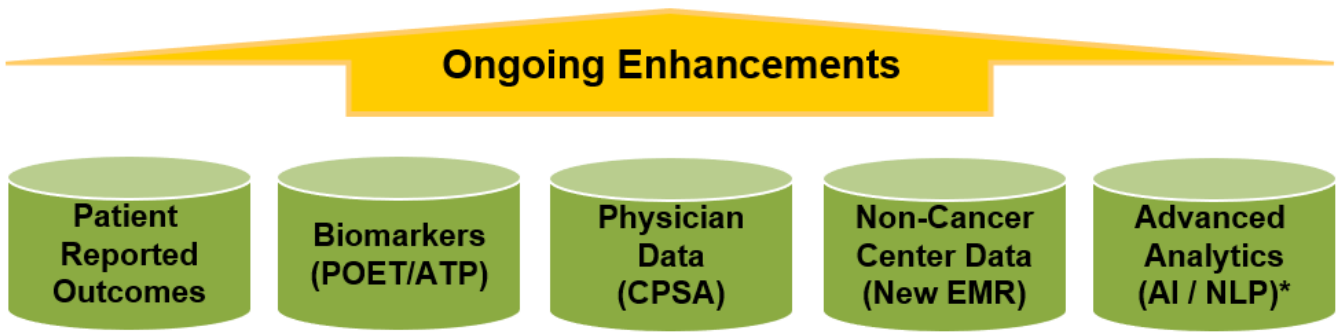
⇨Select optimal treatments
- ⇨Ideal patient?
- ⇨Ideal therapy?
- ⇨Ideal dose / duration?
- ⇨Ideal timing?
- ⇨Ideal sequencing?

Challenges

- Overpromise and underdeliver (e.g. IBM Watson)
- Technology lag in medicine (paper charts still exist!)
- Implementation in the setting of busy and overrun clinics
- Additional barriers in group practices
- Learning curve
- Cost and compensation
- Data privacy
- Clinical data are rarely complete
- Medical landscape is constantly evolving

ONCOLOGY OUTCOMES





Industry perspective on machine learning for drug development

Ryan N. Walton, MPH

Director, Medical Evidence, AstraZeneca Canada

Machine Learning (ML)

Applicability Across Spectrum Of Drug Discovery And Development

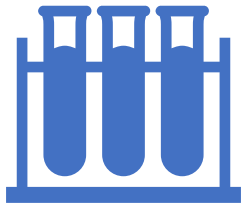
Target discovery
- Predict molecular targets related to diseases of interest

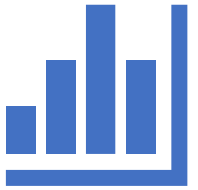
Lead identification
- Profile compounds
- Predict bioactivity
- Predict properties

Lead optimization
- Predict bioactivity
- Predict properties
- Predict formulation

Clinical development
- Trial recruiting
- Trial design and optimization
- Clinical responses

Post approval
- Drug and cost effectiveness
- Pricing models
- AE reporting

Source: adapted from Ekins et al., 2019

Target discovery

- Predict molecular targets related to diseases (specialization, high unmet need)

Application of ML for Target Discovery

- Vast genetic, transcriptomic, proteomic and metabolomic data available for healthy individuals and those with disease
- ML can be applied to several aspects of target discovery by leveraging this data


Applications of machine learning in drug discovery and development

Jessica Vamathevan[1]*, Dominic Clark[1], Paul Czodrowski[2], Ian Dunham[3], Edgardo Ferran[1], George Lee[4], Bin Li[5], Anant Madabhushi[6,7], Parantu Shah[8], Michaela Spitzer[3] and Shanrong Zhao[9]

- Assessing potential starting points (e.g., gene expression in animal models)
- Druggability of protein cavities (physiochemical and structural characteristics)
- Prediction of clinical trial success based on molecular response (e.g., disease-specific RNA expression)

Reference: Vamathevan et al., 2019

Clinical development

- Trial recruiting
- Trial design and optimization
- Clinical responses

Application of ML for Clinical Development

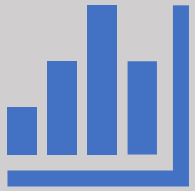
- **Identification of prodromal disease patients to facilitate clinical trial recruitment**
  - ➢ Identification of predictive traits from for Alzheimer's Disease using claims data: treatments received, healthcare visits, testing (Uspenskaya-Cadoz, 2019)

- **Leverage baseline covariate information to improve the efficiency of clinical trials (Zhang, 2019)**
  - ➢ Functional assessment data linked with response from previous trials in disease area

- **Identification of heterogeneous treatment effects**
  - ➢ Random forest analysis to identify subpopulations of interest for hypertension patients who smoke based on similar treatment response (Scarpa, 2019)

Post approval

- Drug and cost effectiveness
- Pricing models
- AE reporting



Application of ML Post-Approval

- Depends on high quality RWD
  - ➢Billing claims, registries, charts/EMRs, surveys
- In many ex-US countries, regional or national governments are key source of credible RWD

- ML methods for HEOR analyses should be considered based on nature of research question (Doupe, 2019)
  - ➢Decision tree methods to characterize risk of outcome in subpopulations
  - ➢Deep learning methods can identify complex patterns and interactions between variables
  - ➢Ensemble methods to improve predictive performance

Case example – *Using machine learning to predict remission in patients with major depressive disorder treated with desvenlafaxine (Benoit, 2019)*

Background

- Major depressive disorder (MDD) is associated with low rates of treatment success – with low remission rates and frequent treatment switches
- Study used baseline clinical data to predict remission after desvenlafaxine

Methods

- Applied ML algorithms to data from 3776 patients from nine phase III/IV studies to produce a model predicting remissions; trained on random 90%

Outcomes

- A trained linear support vector machine (SVM) generated (using 26 features) to predict remission without having to wait 8 weeks to determine treatment outcome

Use of ML for healthcare decision-making faces implementation challenges.

- Many challenges may be agnostic of ML methods applied
  - ➢Upfront costs
  - ➢Clinician acceptance
  - ➢Changes to clinician-patient relationships
  - ➢Privacy, social and ethical considerations

- “Decision transparency” and interpretability are largely method-dependent
  - ➢Documented as key implementation issue for AI, ML, NLP by regulatory and HTA decision-making bodies
  - ➢Apprehension about “black box” models

Reference: CADTH, 2018

Canadian Agency for Drugs and Technologies in Health

...pported by deep learning cannot express reasons ...ular conclusion was made and clinicians cannot ...uts were used to come to a decision.[4,164] Achieving ...w decisions are reached is a challenge, not only for ...t also for developers of AI systems.[165] AI that uses ...rning algorithms based on artificial neural networks may be almost impossible to understand as far as why or how algorithms reach conclusions. Machine learning based on decision trees or Bayesian networks are more transparent to inspection.[166] Within a health care context, AI tools must be able to provide evidence as to how they arrive at specific conclusions, allowing physicians to confirm that the conclusion makes sense and course correct if necessary.[155] Abstract, hidden layers may also create challenges in comparing and assessing the performance of different AI systems.[148]

Reference: CADTH, 2018

There are opportunities to expand the impact of ML for drug development.

- Current approaches to ML in drug development risk siloed decision-making at each stage
  - **Integrated end-to-end (E2E) approach** to ML could provide more seamless process to predict clinical viability (Ekins, 2019)
- Outcomes based agreements (OBA) may have increasingly complex data needs – can ML play a role in optimizing value?
  - **Hypothetical OBA models have highlighted opportunities for optimization** (Brown, 2018)
- **Government-academic-industry partnerships** can help to unlock data-driven discovery
  - Canadian federal government award to Digital Health and Discovery Platform in 2019

References

- Ekins, S., Puhl, A.C., Zorn, K.M. *et al.* Exploiting machine learning for end-to-end drug discovery and development. *Nat. Mater.* **18,** 435–441 (2019). https://doi.org/10.1038/s41563-019-0338-z
- Vamathevan, J., Clark, D., Czodrowski, P. *et al.* Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov* **18,** 463–477 (2019). https://doi.org/10.1038/s41573-019-0024-5
- Uspenskaya-Cadoz, O., Alamuri, C., Wang, L. *et al.* Machine Learning Algorithm Helps Identify Non-Diagnosed Prodromal Alzheimer's Disease Patients in the General Population. *J Prev Alzheimers Dis* **6,** 185–191 (2019). https://doi.org/10.14283/jpad.2019.10
- Zhang, Z., Li, W. and Zhang, H. Efficient Estimation of Mann–Whitney-Type Effect Measures for Right-Censored Survival Outcomes in Randomized Clinical Trials, *Statistics in Biosciences*, 10.1007/s12561-019-09246-2, (2019).
- Scarpa J, Bruzelius E, Doupe P, Le M, Faghmous J, Baum A. Assessment of Risk of Harm Associated With Intensive Blood Pressure Management Among Patients With Hypertension Who Smoke: A Secondary Analysis of the Systolic Blood Pressure Intervention Trial. *JAMA Netw Open.* 2019;2(3):e190005. doi:10.1001/jamanetworkopen.2019.0005
- Doupe, P., Faghmous J, Basu S. Machine learning for health services researchers. *Value in Health* **22,** 808-815 (2019). https://www.sciencedirect.com/science/article/pii/S1098301519301469?via%3Dihub
- Benoit, J.R.A., Dursun, S.M., Greiner, R., *et al.* Using Machine Learning to Predict Remission in Patients with Major Depressive Disorder Treated with Desvenlafaxine (March 15, 2019). Available at SSRN: http://dx.doi.org/10.2139/ssrn.3353387
- An Overview of Clinical Applications of Artificial Intelligence. Ottawa: CADTH; 2018 Sep. (CADTH issues in emerging health technologies; issue 174)
- Brown, J.D., Sheer, R., Pasquale, M. *et al.* Payer and pharmaceutical manufacturer considerations for outcomes-based agreements in the United States. *Value Health.* **21**, 33-40 (2018). https://www.sciencedirect.com/science/article/pii/S1098301517303315

Machine Learning in HEOR

Nicholas Mitsakakis
Adjunct Lecturer, Dalla Lana School of Public Health, University of Toronto
Adjunct Professor, Rady Faculty of Health Sciences, University of Manitoba
Investigator, THETA Collaborative

Types of HEOR studies

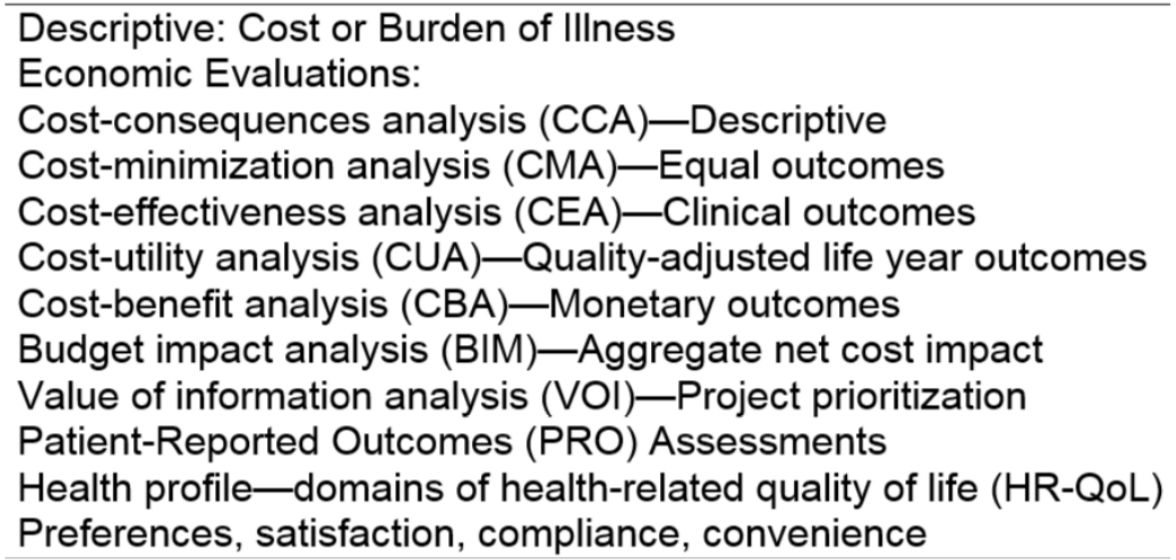


Figure 1. Types of health economics and outcomes research (HEOR) studies.

Chen Y *et al.* Journal of personalized medicine (2016)

Large observational health data

- RWD/RWE
  - Billing claims (administrative data)
  - Surveys e.g. Canadian Community Health Survey (CCHS)
  - Patient registries
  - Lab tests
  - Clinical records (EMR)
- Heterogeneity, variability, dependence on jurisdiction

Typical goals of analysis

- Estimation of mean outcome given the values of predictors
  - E.g. Mean health utility or mean cost per 30 days, conditional on predictors
- Estimation of *treatment effect* on outcomes
- Estimation of *attributable* or *net* costs for a particular condition or disease
- Estimation of transition probabilities/intensities (for parameterizing decision analytic models)

How ML can help?

- Powerful in identifying patterns and relationships in data that are not explicitly described
- Complexity, high level of interactions
- Non-linearity
- No need for explicit functional form specification
- Involving a very large number of heterogeneous features
- Sparseness

Structured vs. unstructured data

- The majority of health data are *structured*
- Typical “data table”:
  - Rows are the observations (subjects, visits, measurements, episodes of care etc.)
  - Columns are the features (variables)
- Some health data are *unstructured*
  - Physician’s notes
  - Speech
  - Medical imaging

Natural Language Processing

- It can be used to extract features from free text (*information extraction* task)
- These can then be used from traditional methods for further analysis, e.g. classification methods
- “Computational phenotyping”: algorithmic methods used to identify patients with specific disease or condition in health data (Wong *et al*, 2018)
- Similar approaches can be used for other types of unstructured data (e.g. medical imaging)

Causal inference and machine learning

- Average Treatment Effect (ATE):

$$E_{\mathbf{X}}[E(Y|T=1,\mathbf{X}) - E(Y|T=0,\mathbf{X})]$$

- Estimation approaches:
  - (Generalized) Propensity Score
  - *G*-Computation
  - Targeted Maximum Likelihood Estimation (TMLE)
- ML methods can be incorporated for the estimation of
  - the propensity score, $P(T=1|\mathbf{X})$
  - the expected outcome, $E(Y|T,\mathbf{X})$
- E.g. Kreif *et al*, 2015; Schuler and Rose, 2016

Predicting costs over the first year after renal transplantation: a case study

- Fu R, Mitsakakis N, Coyte P. *A healthcare cost calculator for older patients over the first year after renal transplantation.* ITU Kaleidoscope (2019)
- Kidney transplantation (KT) is considered the preferred option for patients with end-stage renal disease (ESRD)
- However it can be extremely costly
- It is important to be able to predict the cost of KT and to identify important drivers of the cost
- **Objective:** To develop a “cost calculator” for 1-yr costs after KT for older patients with ESRD, using machine learning

Materials and Methods

- Data:
  - Canadian Organ Replacement Registry (only Ontario transplantation centers)
  - Linked to administrative data at the Institute of Clinical Evaluative Sciences (ICES)
- Outcome:
  - 1-yr cumulative total cost (log transformed)
  - Calculated by existing micro-costing algorithm at ICES

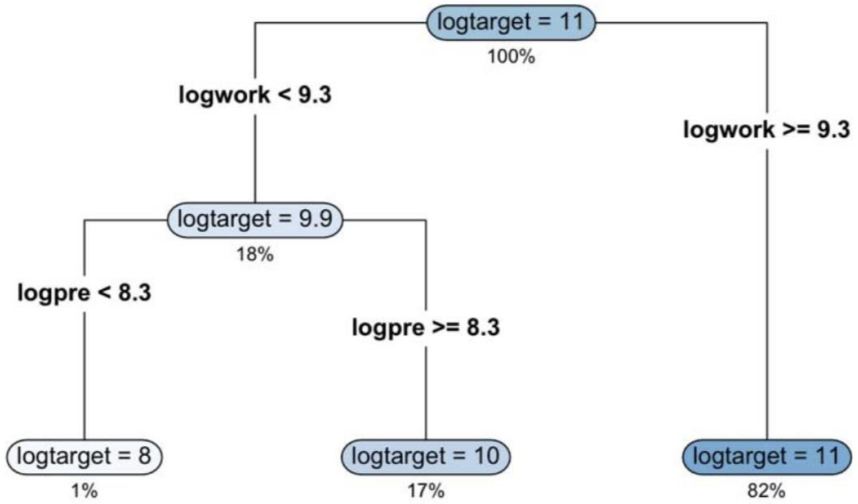
Materials and Methods (cont.)

- Predictors
  - Chosen based on previously identified association
  - Guided by previous publications with similar objectives
  - Demographics, co-morbidities, clinical characteristics, transplant-information, pre-transplant healthcare use

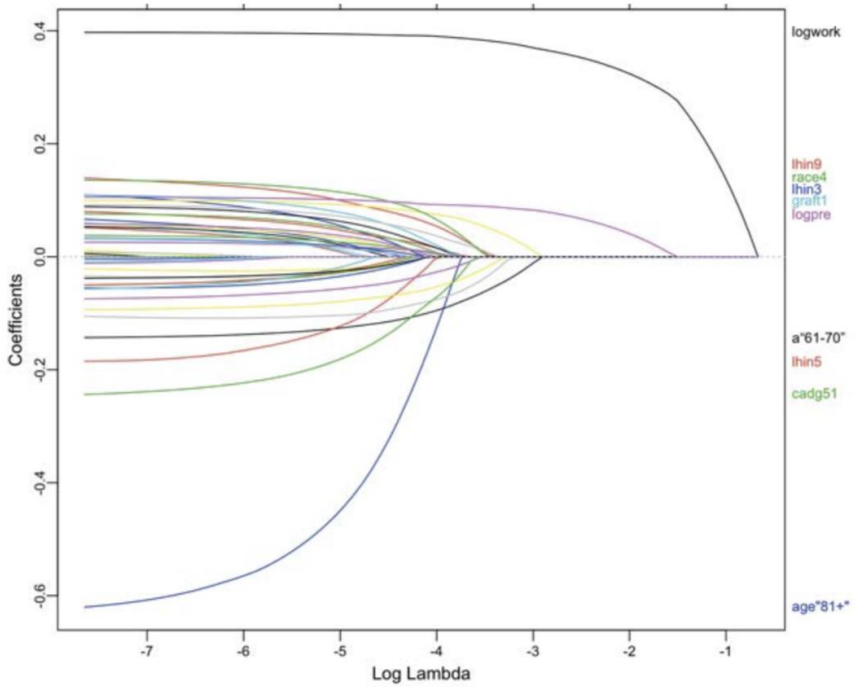
- Methods
  - Linear regression with regularization (ridge regression, lasso, elastic net)
  - Regression tree with pruning
  - 10-fold cross-validation based on MSE

Results

Regression tree



Lasso



Models	RMSE	R^2
Ridge regression	0.618	0.255
Lasso regression	0.604	0.258
Elastic net regression	0.610	0.251
Regression tree	0.630	0.0101

N = 1328
Caucasian: 75.2%
Male: 67.3%
61-70yrs: 81.4%

Important predictors

- Lasso identified workup costs (6m pre tx), pre-workup costs (12m before that) and age 80+ yrs as the most important predictors
- Regression tree identified workup and pre-workup costs as the only predictors selected in the tree

Discussion

Strengths

- Data linkage allowed individual patient level costing, even prior to tx; availability of clinical variables
- Methods accommodated successfully a large number of predictors battling overfitting
- Consistency in the selection of important predictors
- Interpretability (chosen methods)

Limitations

- Missing values
- Censoring
- Interpretability (log-transformed costs)
- Donor’s information is missing

Other considerations

- ML methods cannot correct problems due to data (e.g. missingness, censoring, measurement error etc.)
- Sampling bias (“observations missing”)
- Selection bias (variables missing)
- Results need to be interpreted properly
- Generalizability
- Equity

References

- Chen Y, Guzauskas GF, Gu C, Wang B, Furnback WE, Xie G, Dong P, Garrison LP. Precision health economics and outcomes research to support precision medicine: big data meets patient heterogeneity on the road to value. Journal of personalized medicine. 2016 Dec;6(4):20.
- Wong J, Horwitz MM, Zhou L, Toh S. Using machine learning to identify health outcomes from electronic health record data. Current epidemiology reports. 2018 Dec 1;5(4):331-42.
- Kreif N, Grieve R, Díaz I, Harrison D. Evaluation of the effect of a continuous treatment: a machine learning approach with an application to treatment for traumatic brain injury. Health economics. 2015 Sep;24(9):1213-28.
- Schuler MS, Rose S. Targeted maximum likelihood estimation for causal inference in observational studies. American journal of epidemiology. 2017 Jan 1;185(1):65-73.
- Fu R, Mitsakakis N, Coyte PC. A Healthcare Cost Calculator for Older Patients Over the First Year After Renal Transplantation. In2019 ITU Kaleidoscope: ICT for Health: Networks, Standards and Innovation (ITU K) 2019 Dec 4 (pp. 1-9). IEEE.

Interpretable machine learning and Bayesian networks

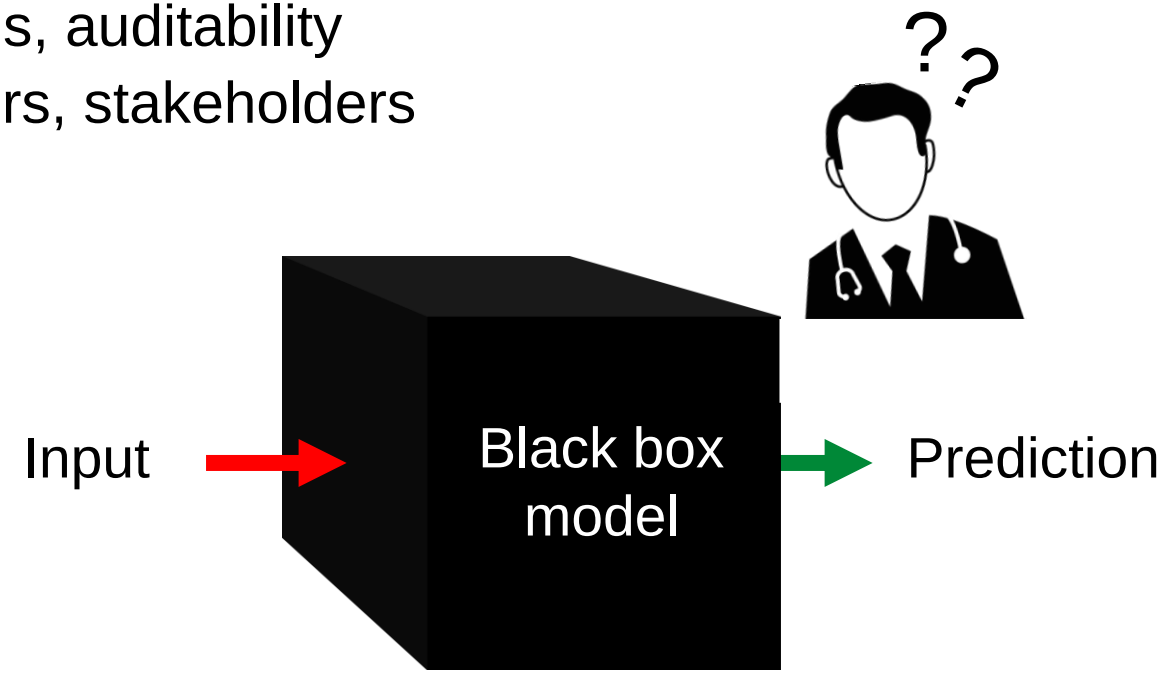
Alind Gupta

Real-world and advanced analytics (RWA^2)

Cytel Inc., Toronto, Canada

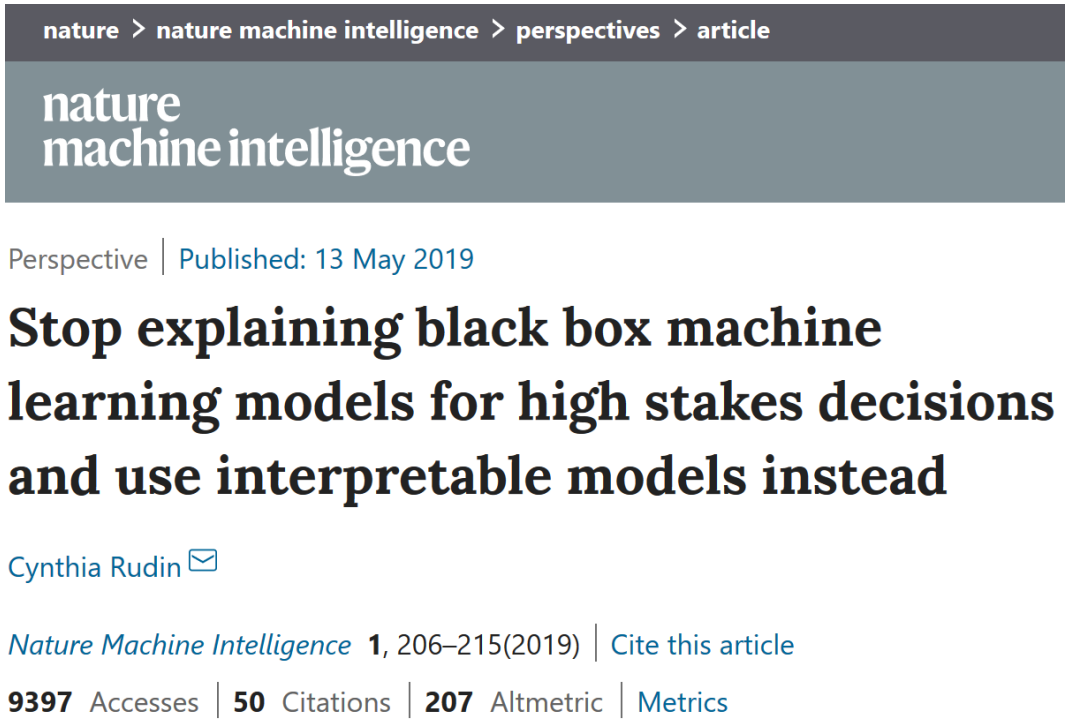
Machine learning as a black-box

- Trade off interpretability for capacity to learn highly non-linear relationships
- Problem – accuracy on given sample is not guarantee of generalizability, robustness or fairness
  - Insight from domain experts
  - Bias, limitations, auditability
  - Trust from users, stakeholders



Transparency is important

- **EU GDPR**
  - Individual's right to explanation of an automated decision regarding them
- **FDA** guidance for good machine learning practices
  - Appropriate validation and transparency
  - Need for continued safety and efficacy
  - "Clinicians in the loop" where necessary



Cytel

What are models really learning?

“Panda”
(57%)

+

=

“Gibbon”
(99%)

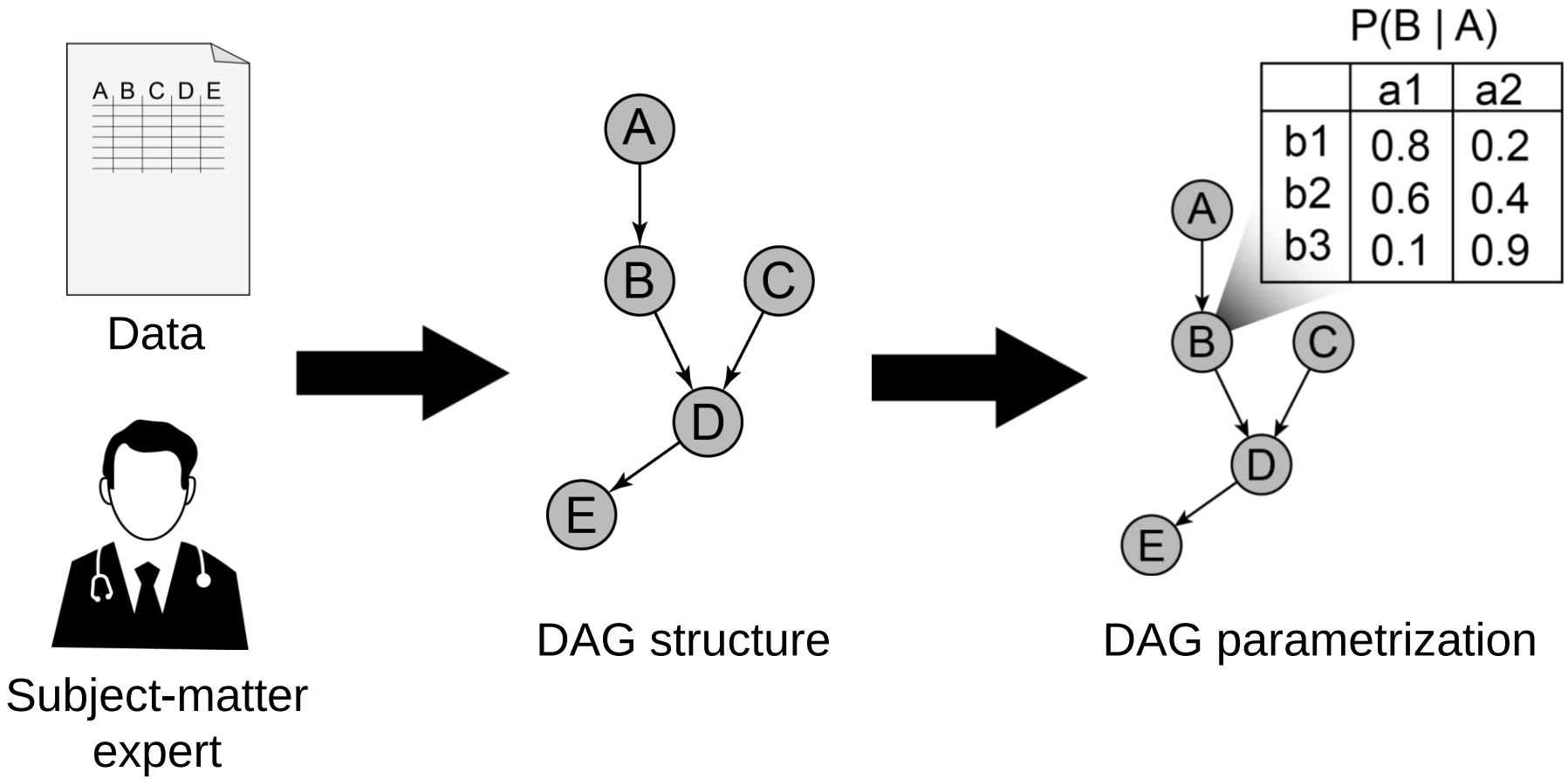
- IBM Watson for Oncology
- COMPAS – unvetted model for predicting recidivism
- Commercial face detection software – gender bias

Strickland, E. (2019). IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. IEEE Spectrum, 56(4), 24-31.
Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. Science advances, 4(1), eaao5580.
Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency (pp. 77-91).
Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., & Swami, A. (2017, April). Practical black-box attacks against machine learning. In Proceedings of the 2017 ACM on Asia conference on computer and communications security (pp. 506-519).

Cytel

Bayesian networks

- Computations over a DAG



Graphical models

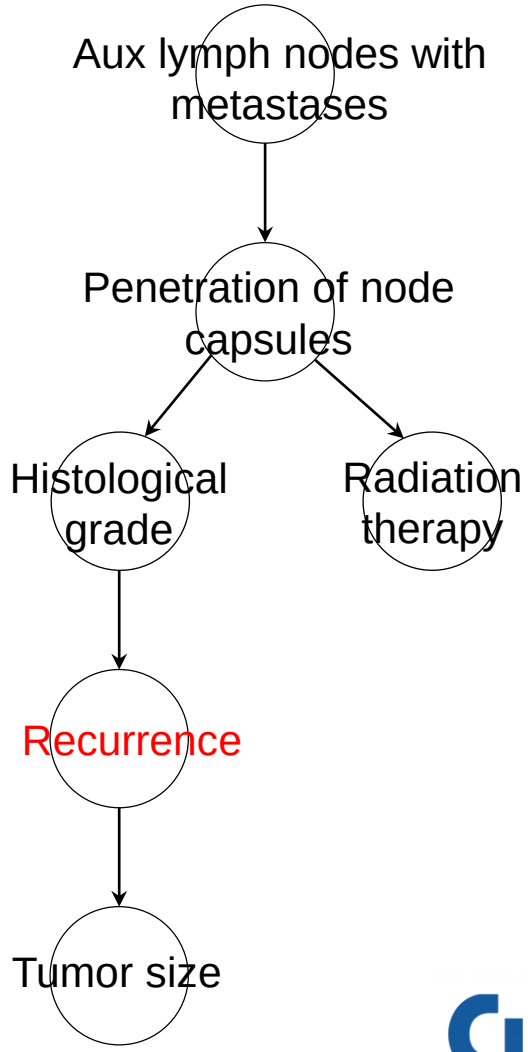
- Bayesian inference
- Generative models for NLP and computer vision
  - Latent Dirichlet allocation
  - GANs, variational autoencoders
- Causal inference
  - Confounders and bias
  - Estimating causal effects from observational data
- Prediction

Cytel

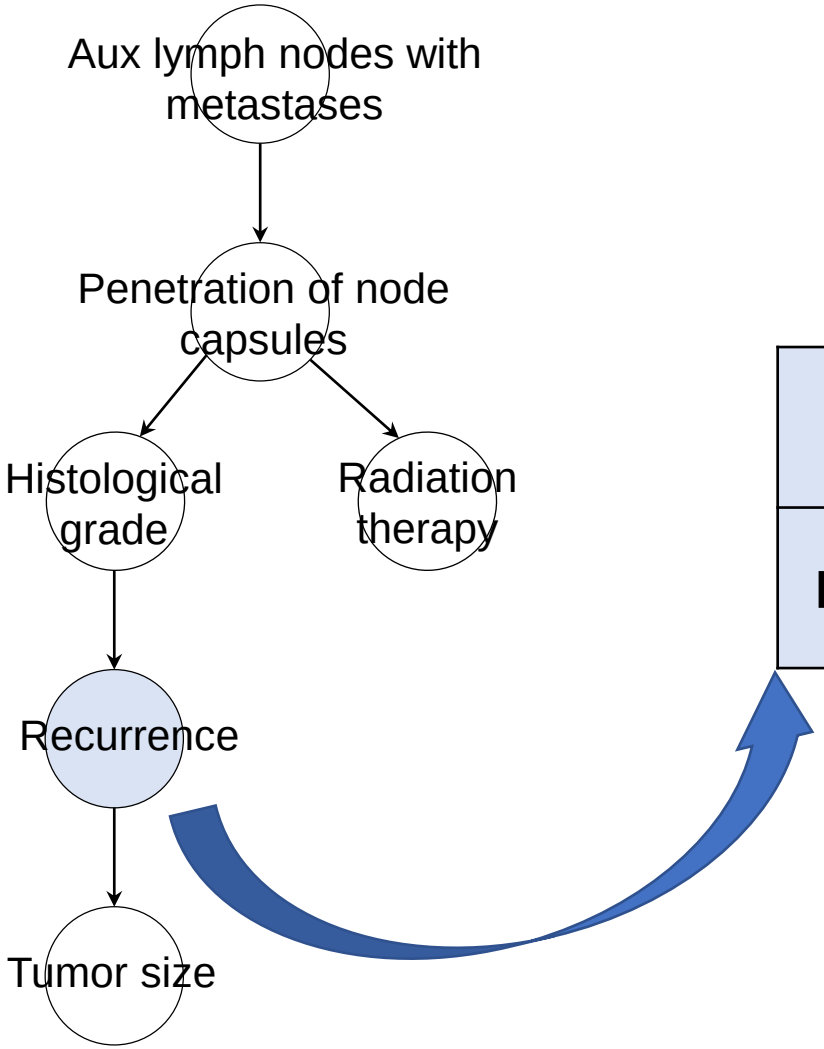
Step 1 – Identify a model structure

- Breast cancer recurrence data from UCI ML repository (*n*=285)
- 7 variables
  - *Recurrence*
  - Malignancy grade
  - Tumour size
  - Penetration to lymph node capsules
  - Number of auxiliary lymph nodes with metastases
  - Irradiation

- Maximize a likelihood function
- Conditional independence tests (e.g. Chi-square)
- Expert elicited



Step 2 – Parametrize model



***P*(Recurrence|Histological grade)**

		Histological grade		
		1	**2**	**3**
Recurrence	**No**	0.83	0.78	0.46
	Yes	0.17	0.22	0.53

Cytel

Step 3 – Run queries

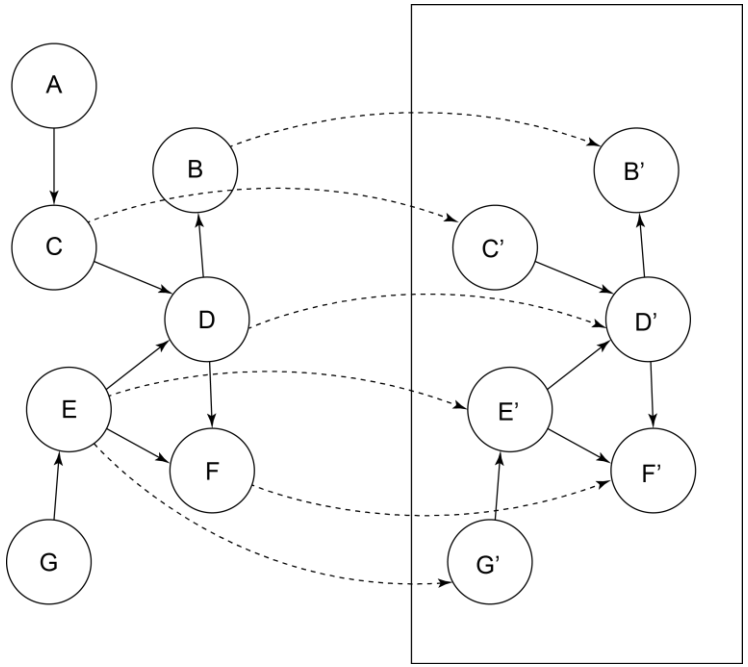
- Prediction
  - P(outcome|predictors)
- Multivariate prediction
  - P(outcome_1, outcome_2, …, outcome_N|predictors)
- “Backward” prediction
  - P(predictors|outcome)
- Anomalous data
  - P(observations)
- Interventions and counterfactuals (causal DAGs)
  - P(outcome|*do*(intervention))

(All queries are implicitly conditioned on a trained model)

Cytel

Extensions of Bayesian nets

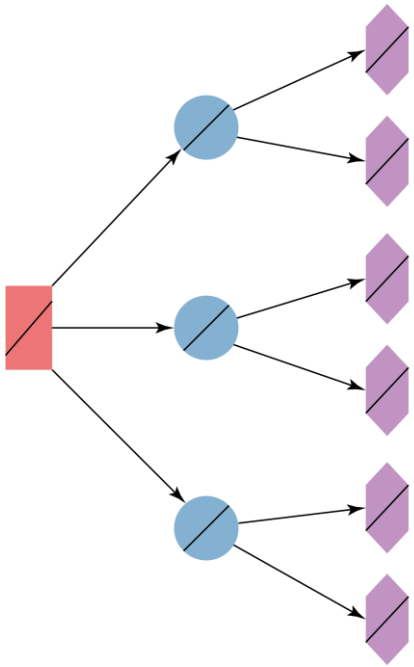
Time modelling with Dynamic Bayes nets



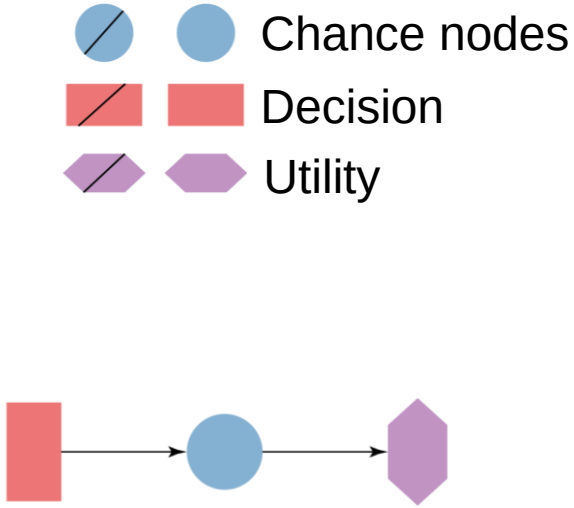
Initial distribution

Time replication

Decision theoretic modelling with decision nets/influence diagrams



Decision tree



Influence diagram

Cytel

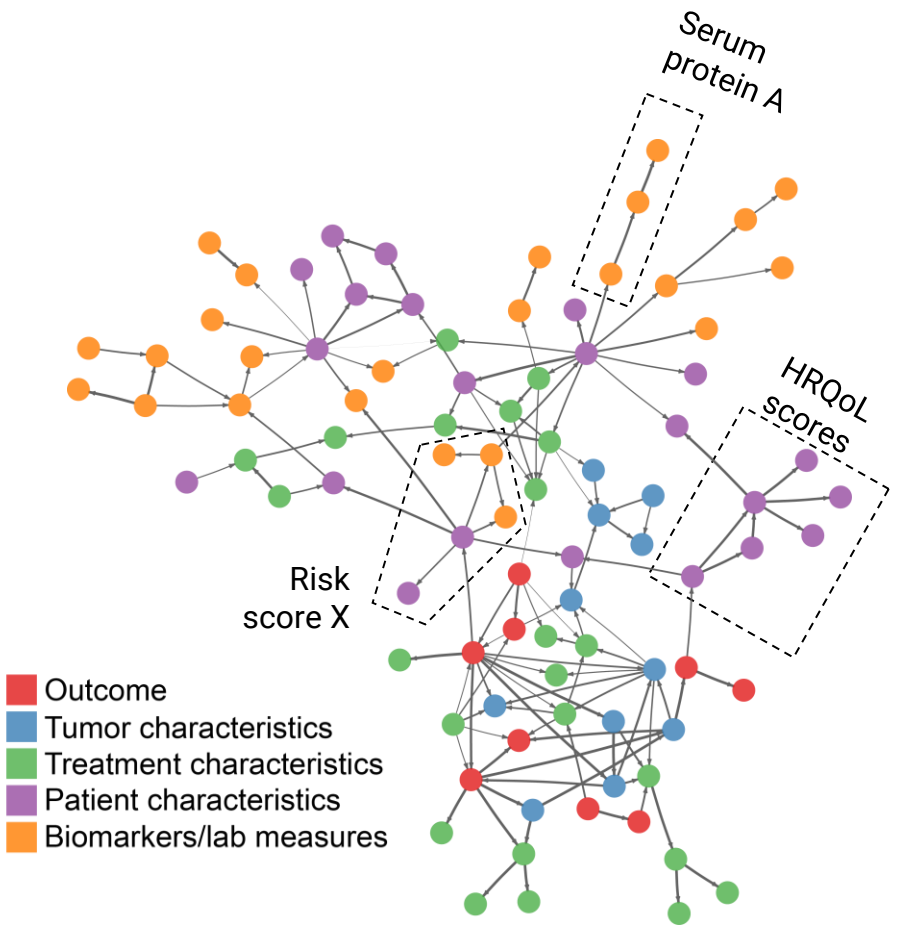
Case study: Immunotherapy for advanced cancer

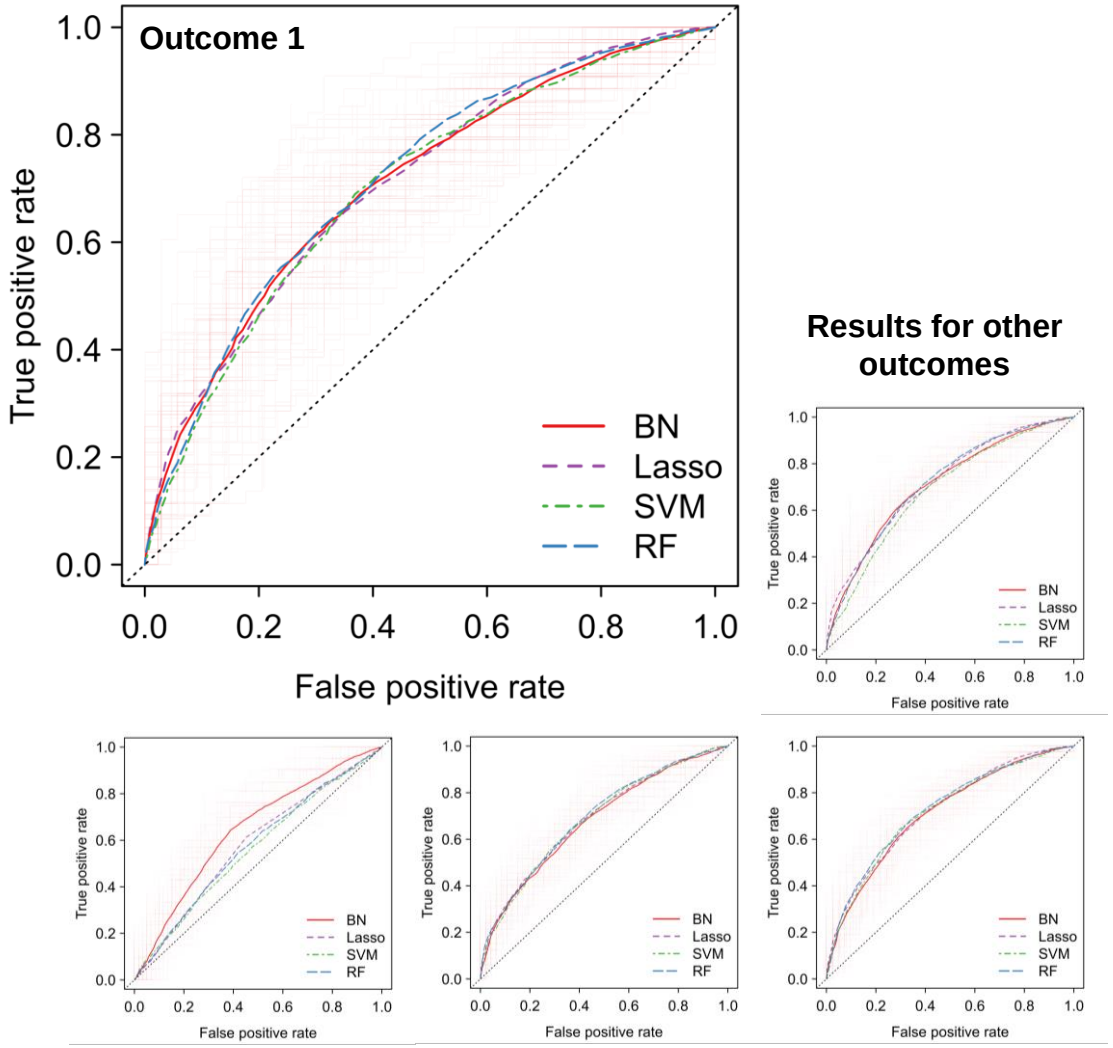
- **Challenges**
  - High individual-level heterogeneity in response to treatment
  - Subsets show durable response, severe adverse events
  - Short follow-up early
  - Multiple outcomes of interest

- **Applications of machine learning**
  - Identifying prognostic variables
  - Long-term projections for health economic evaluations for HTA
  - Informing future trial design, surrogate endpoints
  - Patient simulation with time-varying interventions

Cytel

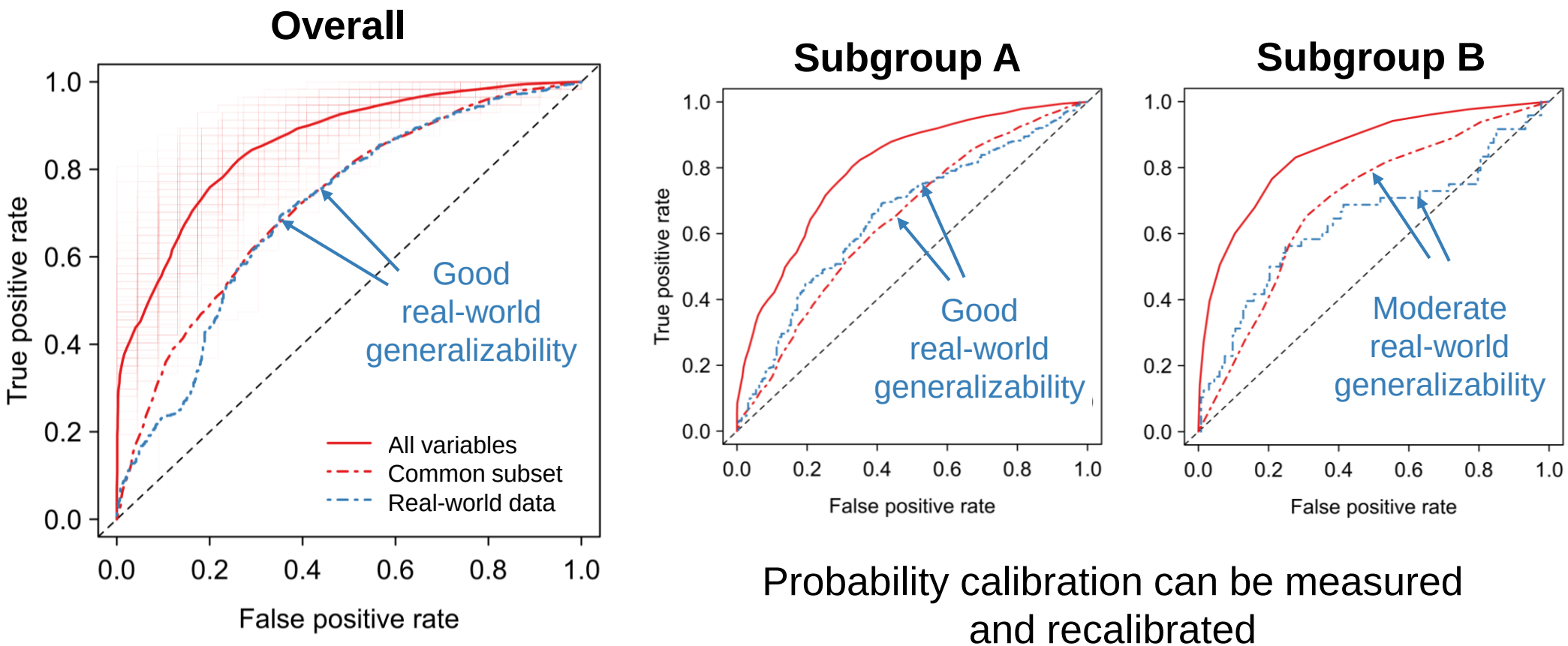
Multivariate prediction network





- 35% causal edges

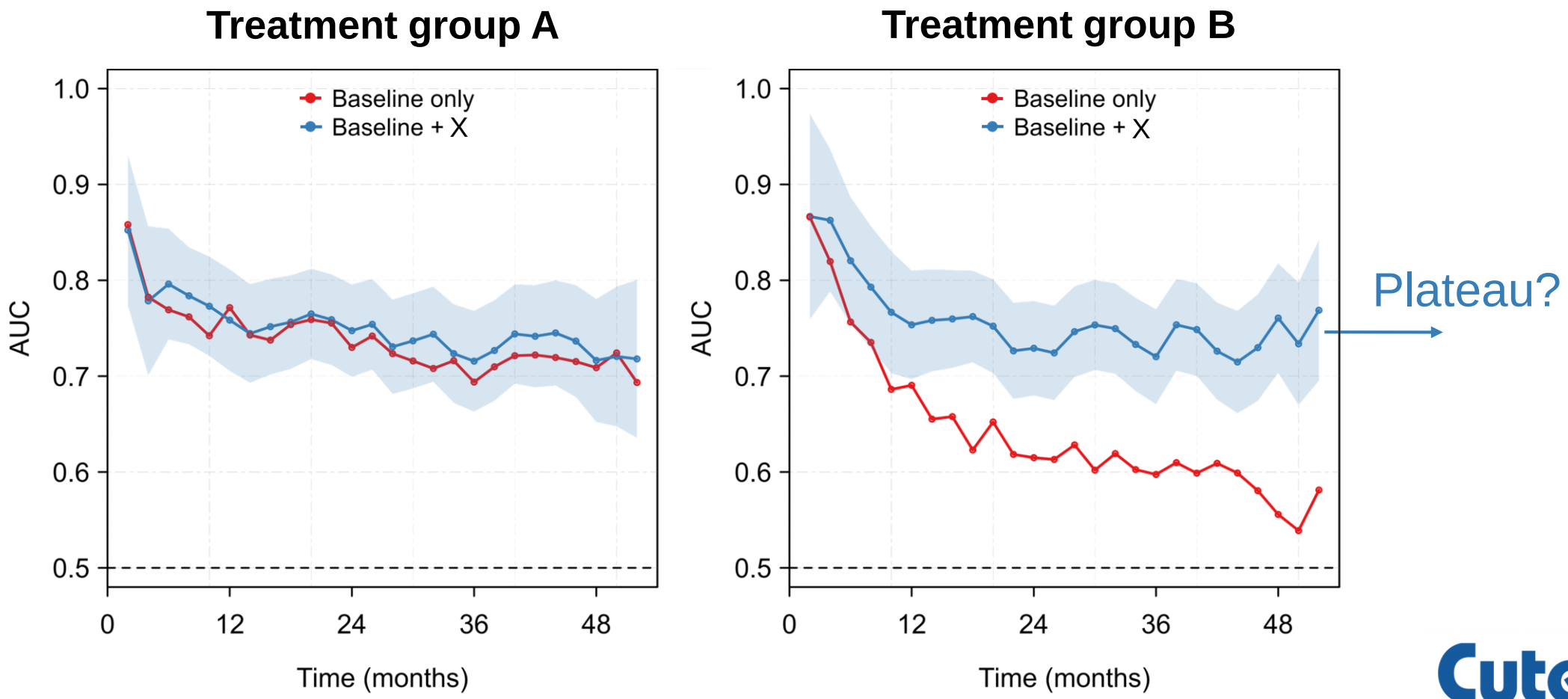
External validation using real-world data



Probability calibration can be measured and recalibrated

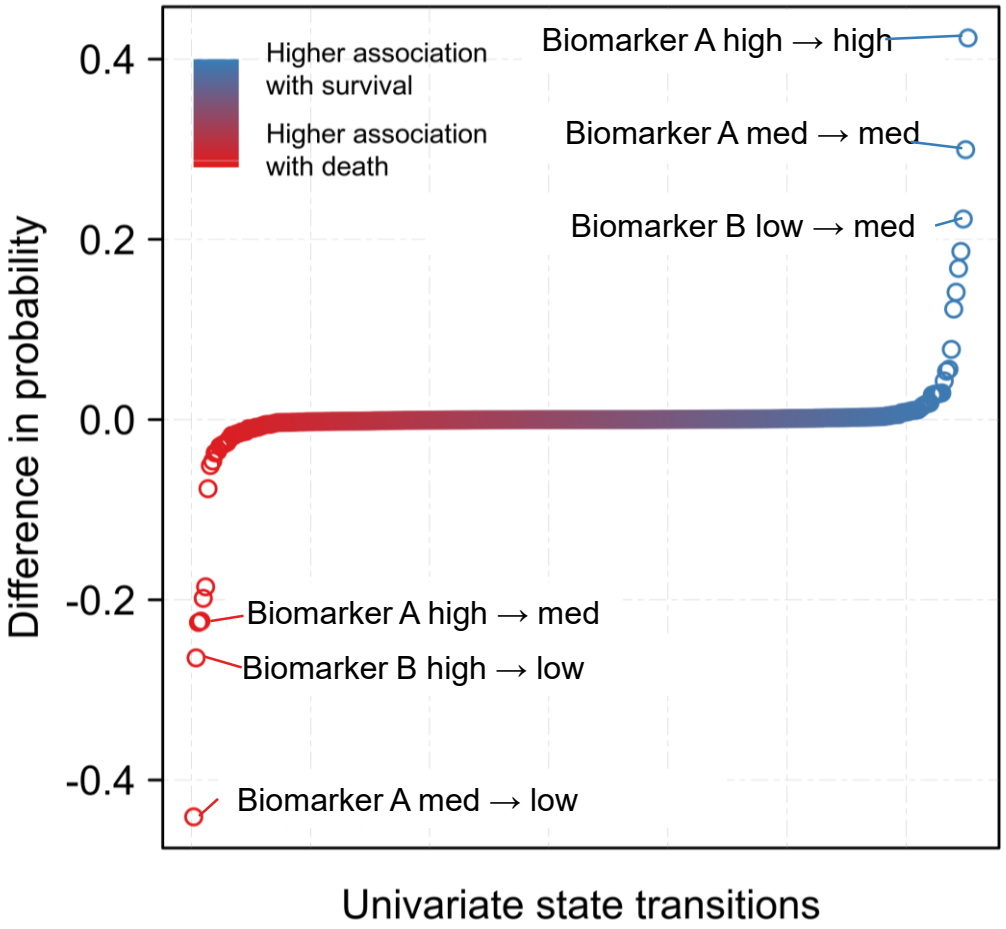
Prediction over time

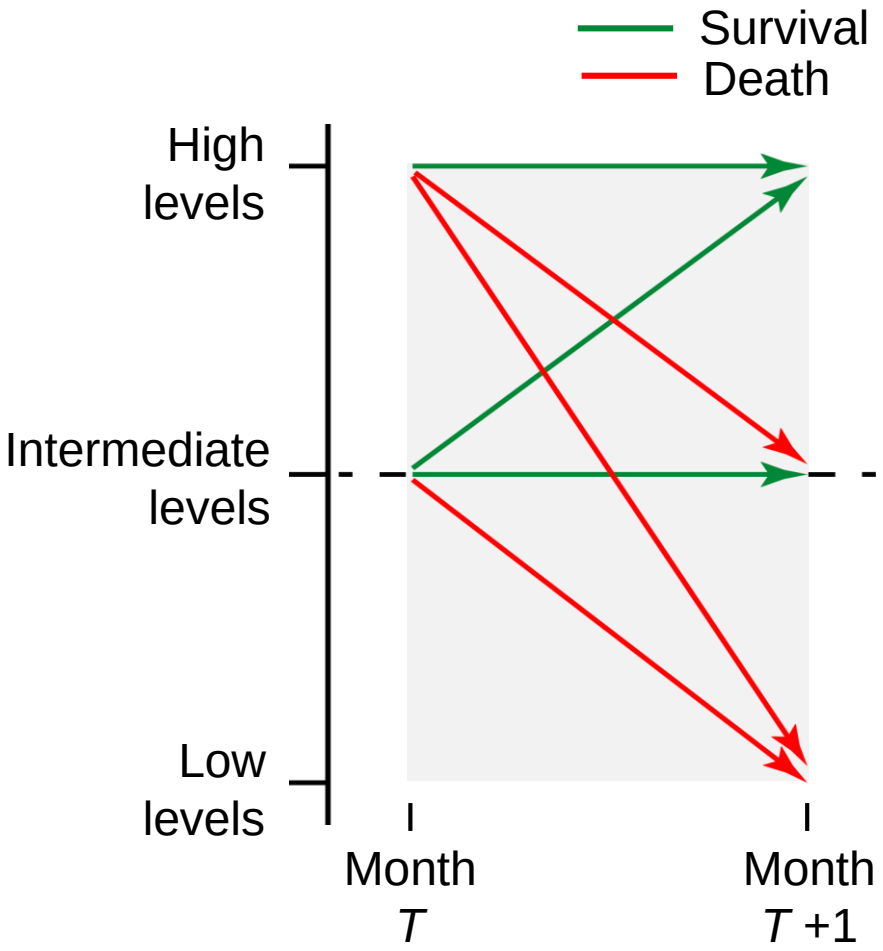
$P(OS_t|Y_{baseline})$ and $P(OS_t|Y_{baseline}, X)$



Prognostic changes in variables

$P(X|OS_{t-1}=\text{“alive”}, OS_t=\text{“alive”}) - P(X|OS_{t-1}=\text{“alive”}, OS_t=\text{“dead”})$





Conclusion

- Bayesian networks are useful for
  - Small data
  - Interpretable models
  - Multivariate prediction
  - Missing data problems
  - Incorporating subject matter expertise

Cytel

Contacts

- Alind Gupta (alind.gupta@cytel.com)
- Winson Y. Cheung (winson.cheung@ahs.ca)
- Ryan N. Walton (ryan.walton@astrazeneca.com)
- Nicholas Mitsakakis (n.mitsakakis@theta.utoronto.ca)