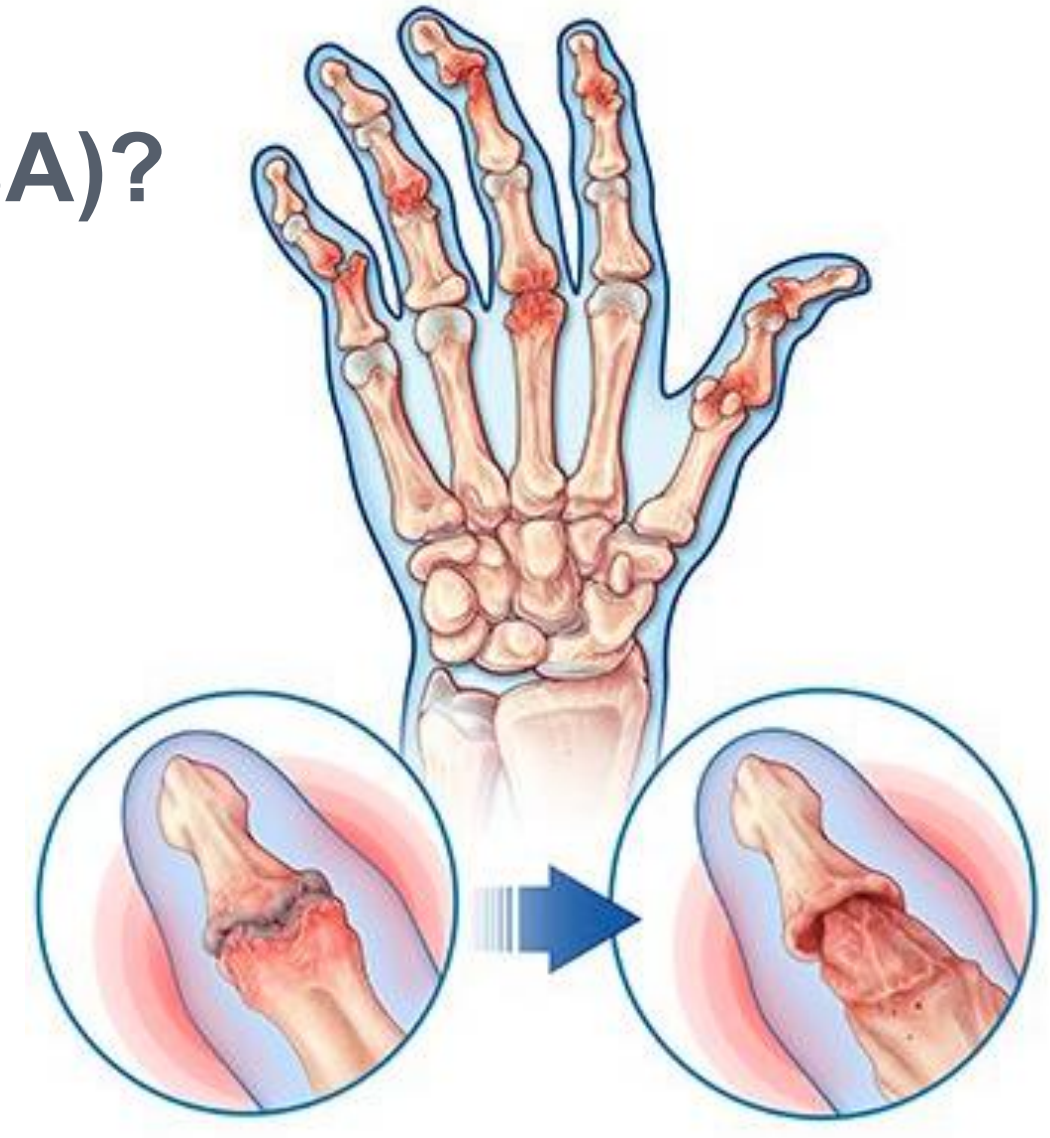
Interpretable ML to detect psoriatic arthritis with EHRs

Dr Theresa Smith
Senior Lecturer in Statistics
University of Bath

What is psoriatic arthritis (PsA)?

- Progressive inflammatory arthritis associated with destruction of joints.
- Uncommon in the general population (<1%) but common in people with the skin disease psoriasis (PsO).
- PsO often treated in primary care. Need to support GPs in identifying PsA early.
- Delays to diagnosis are common and even short delays (6m) can lead to poorer long-term outcomes.



https://www.stelarainfo.com/psoriatic-arthritis/what-is-psoriatic-arthritis/psoriatic-arthritis-symptoms/

Trying to detect PsA is hard!

Symptoms of PsA are very generic: joint pain, swelling in one or more joints, joint stiffness, fatigue. Lots of joints can be affected. Can be misdiagnosed as osteoarthritis.

PRESTO-PsA: 1-year Risk of Developing Psoriatic Arthritis

Sex	○ Male ○ Female
Age	▾ years
Family history of psoriasis	○ Yes ○ No
Morning back stiffness	○ Yes ○ No
Nail pitting	○ Yes ○ No
Stiffness level (0 - none; 10 - severe)	▾
Use of biological therapy	○ Yes ○ No
Patient global health	○ Good/Very good ○ Fair/Poor/Very poor
Pain level	○ Mild/Moderate/Severe ○ None
Duration of psoriasis (number of years)	▾ years

Your 1-year risk is _____ [Show My Risk] [Clear All] or [Back to Home Page]

We have some screening tools to identify undiagnosed cases of PsA.

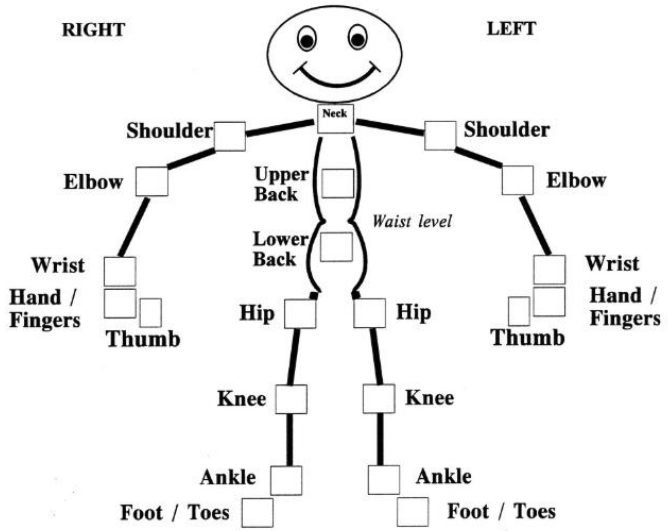
PRESTO

There is some recent work (e.g., Eder et al. 2023) on risk calculators estimating future risk of developing PsA.

PEST

	NO	YES
Have you ever had a swollen joint (or joints)?		
Has a doctor ever told you that you have arthritis?		
Do your finger nails or toe nails have holes or pits?		
Have you had pain in your heel?		
Have you had a finger or toe that was completely swollen and painful for no apparent reason?		

In the drawing below, please tick the joints that have caused you discomfort (i.e., stiff, swollen, or painful joints).



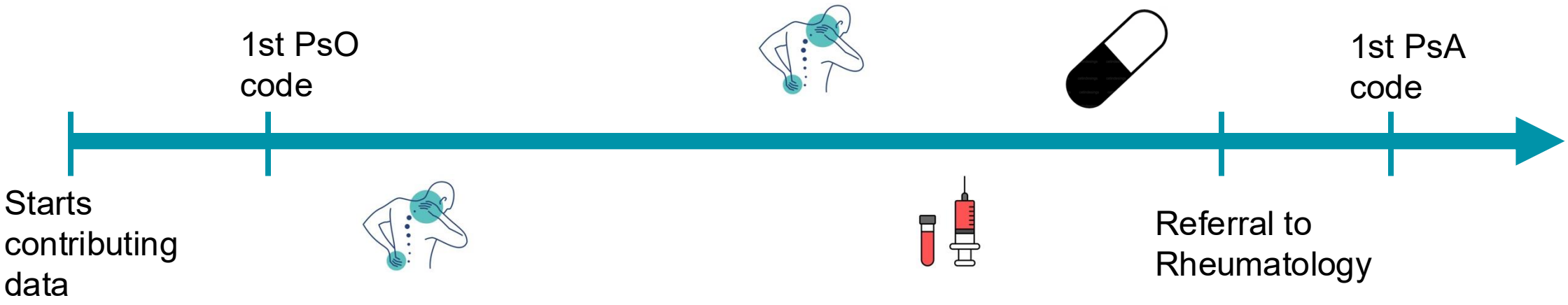
The PrediPsA Project

Building on a previous NIHR project (PROMPT) focused on early detection, use a large cohort of psoriasis patients from electronic health records to

- Develop and compare different machine learning approaches that predict development of PsA using data routinely recorded in primary care including occurrence of a wide range of musculoskeletal symptoms, blood tests and medications. (Focus of today's talk)
- Develop a dynamic model that can produce updated estimates of the risk of developing PsA as symptoms, blood test results and medications accrue over time. (Ongoing)
- Funded by the pharmaceutical company, UCB.

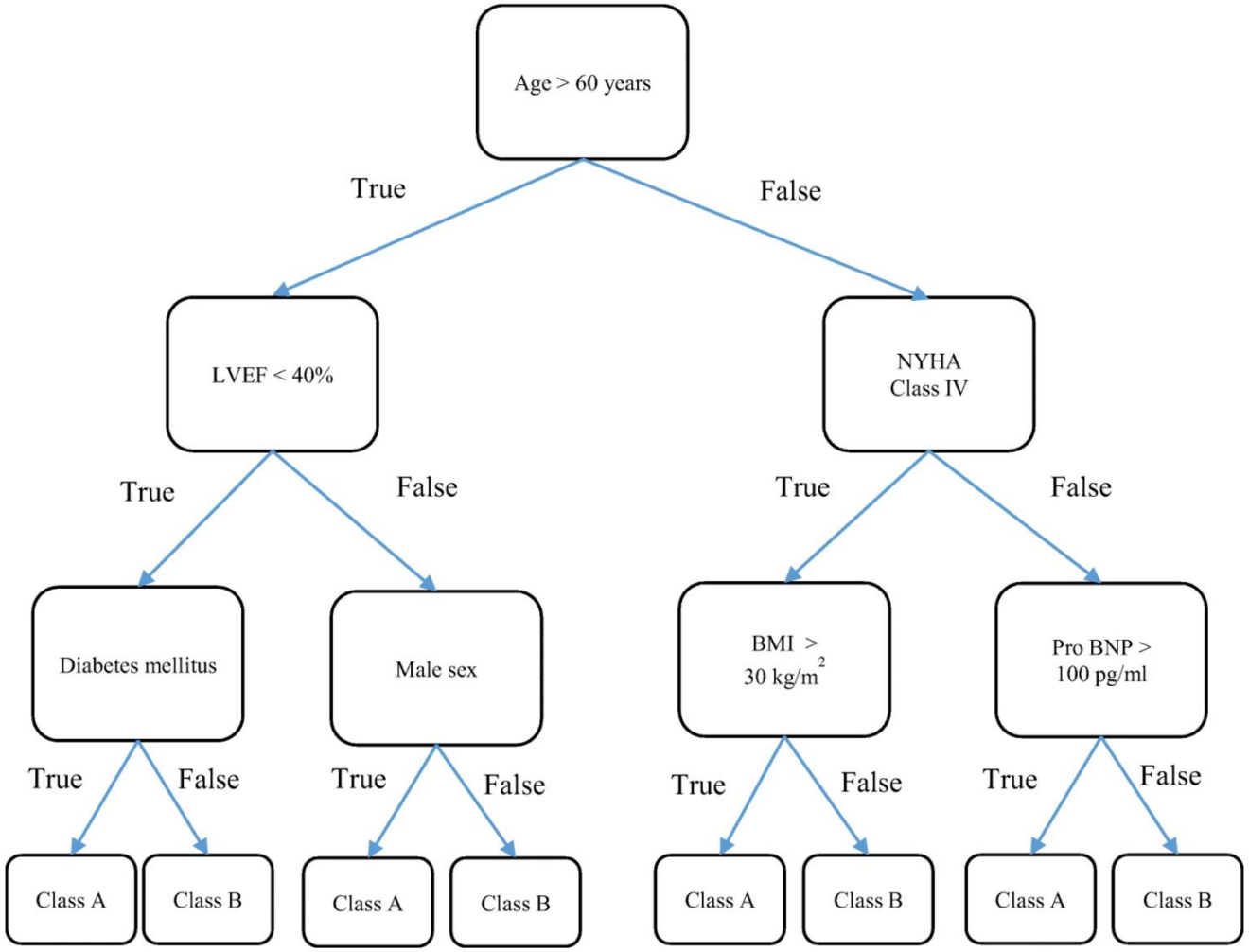
Clinical Practice Research Datalink

- Records from GP practices from across the UK, includes clinical observations, prescriptions, blood tests, referrals to secondary care.
- We identified a cohort of 122,330 patients with newly-identified psoriasis, and of these 2,460 patients developed PsA.
- PsA codes were validated with other evidence like referrals and drugs.



Random Forests

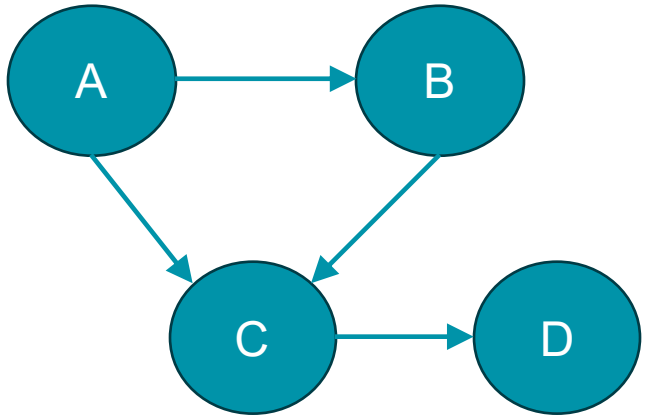
- Popular machine learning approach based on collections of *decision trees.*
- Decision trees algorithmically find split points in the features and then partitions the data into small subgroups with similar values on the endpoint of interest.
- Good predictive performance and easily capture interactions and non-linearities.



Example from Mpanya et al (2020)
doi: 10.1007/s10741-020-10052-y

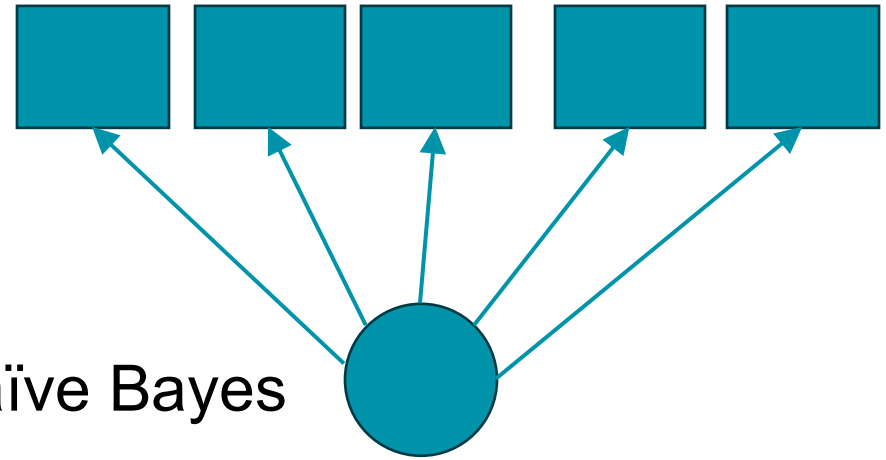
Bayesian Networks

- Probabilistic graphical model which represents dependencies between variables using a directed acyclic graph.
- Graph tells us how to factorize a joint distribution and can help clinicians interpret what the model is doing.
- Can either specify the graph from expert opinion or learn it from data.
- See McLachlan et al. 2020 for a review of their use in healthcare.



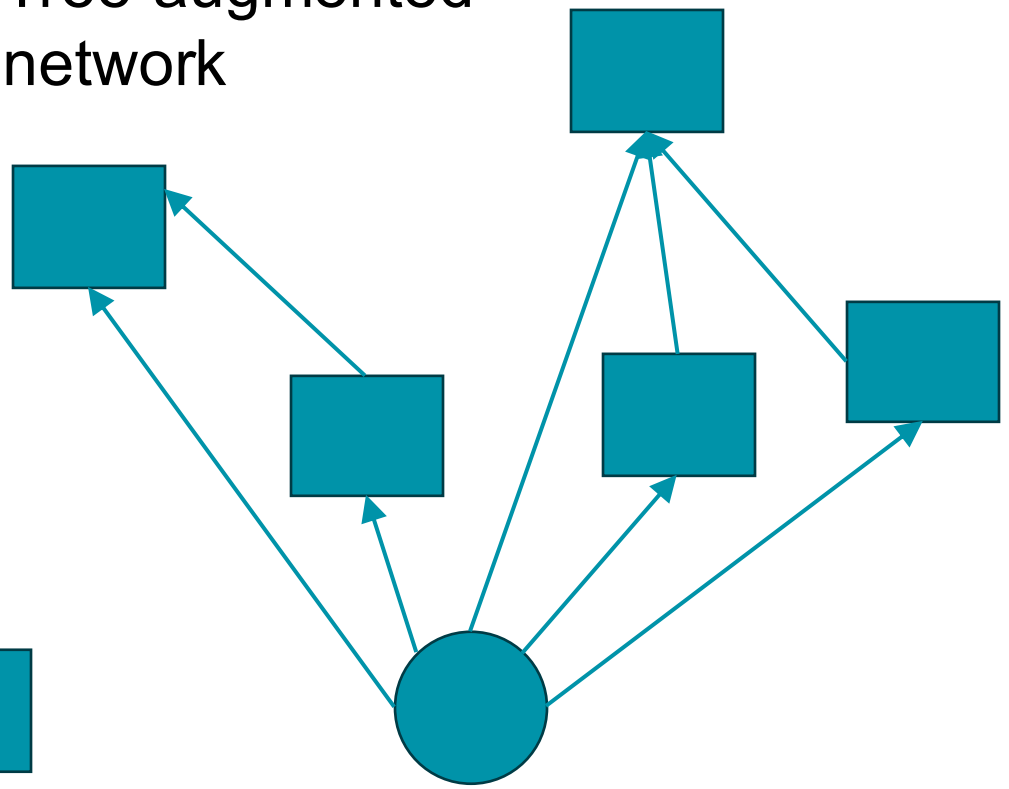
$$\mathrm{P}(A, B, C, D) = P(A) \times P(B \mid A) \times P(C \mid A, B) \times P(D \mid C)$$

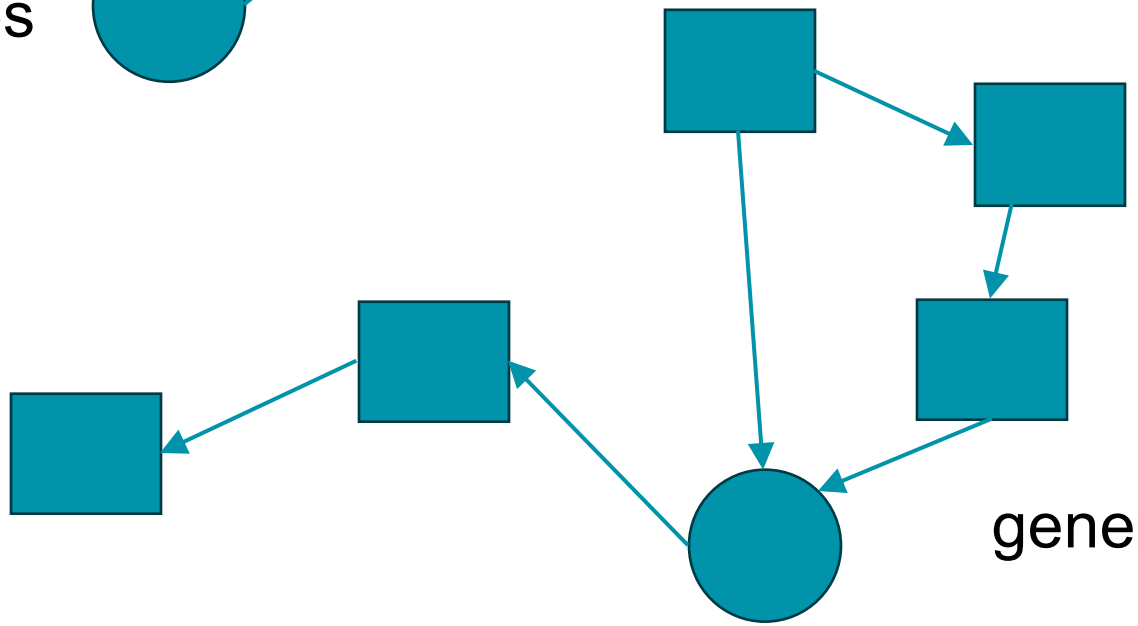
Types of networks



Naïve Bayes

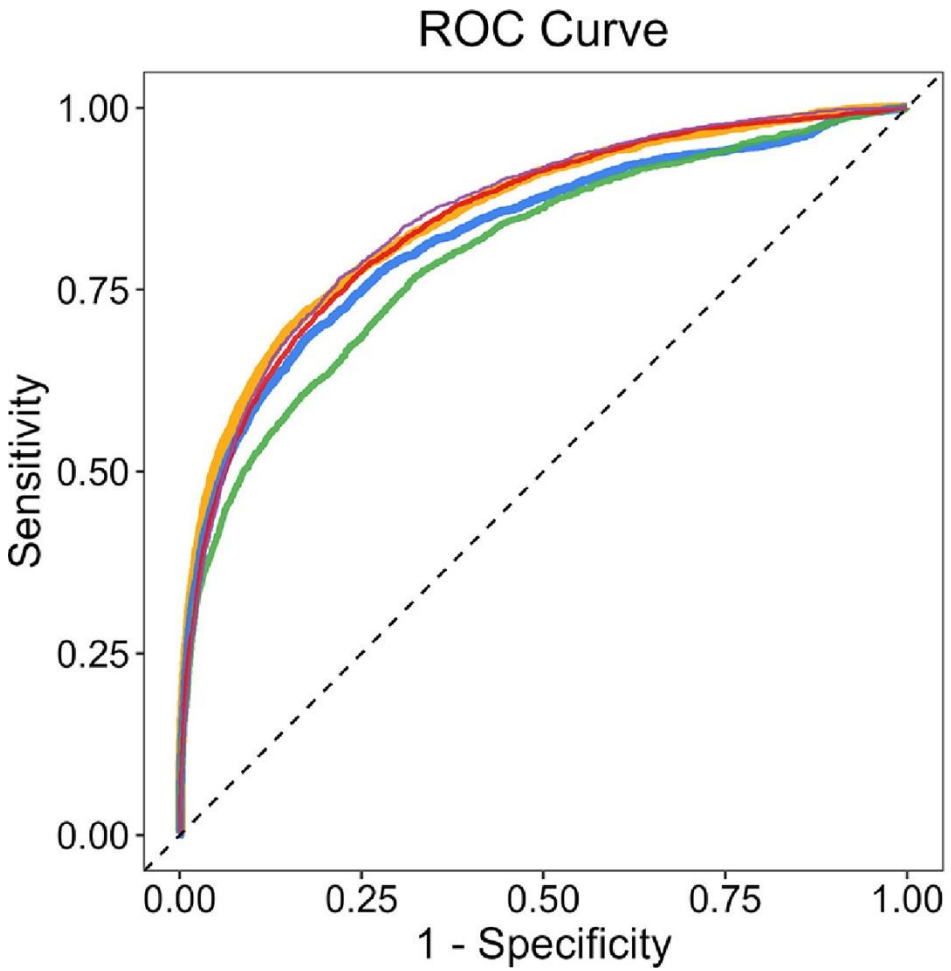
Tree-augmented network

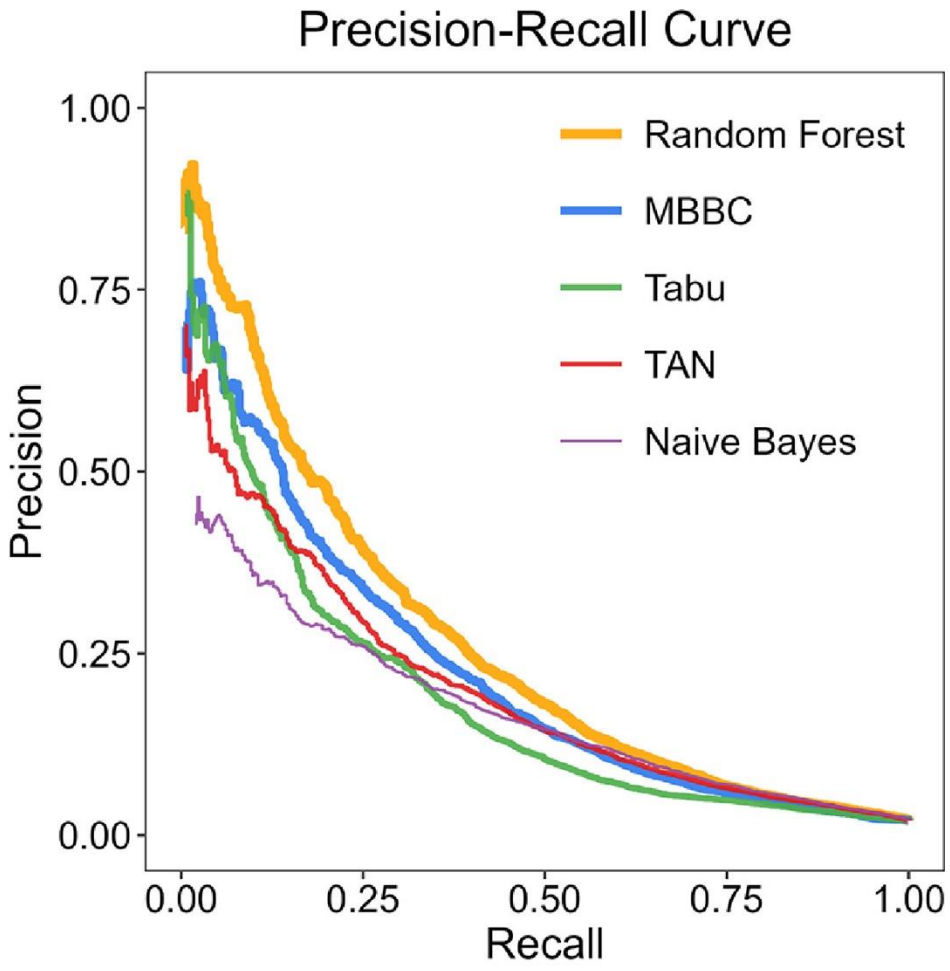




general Bayesian network

Predictive Performance





AUC:

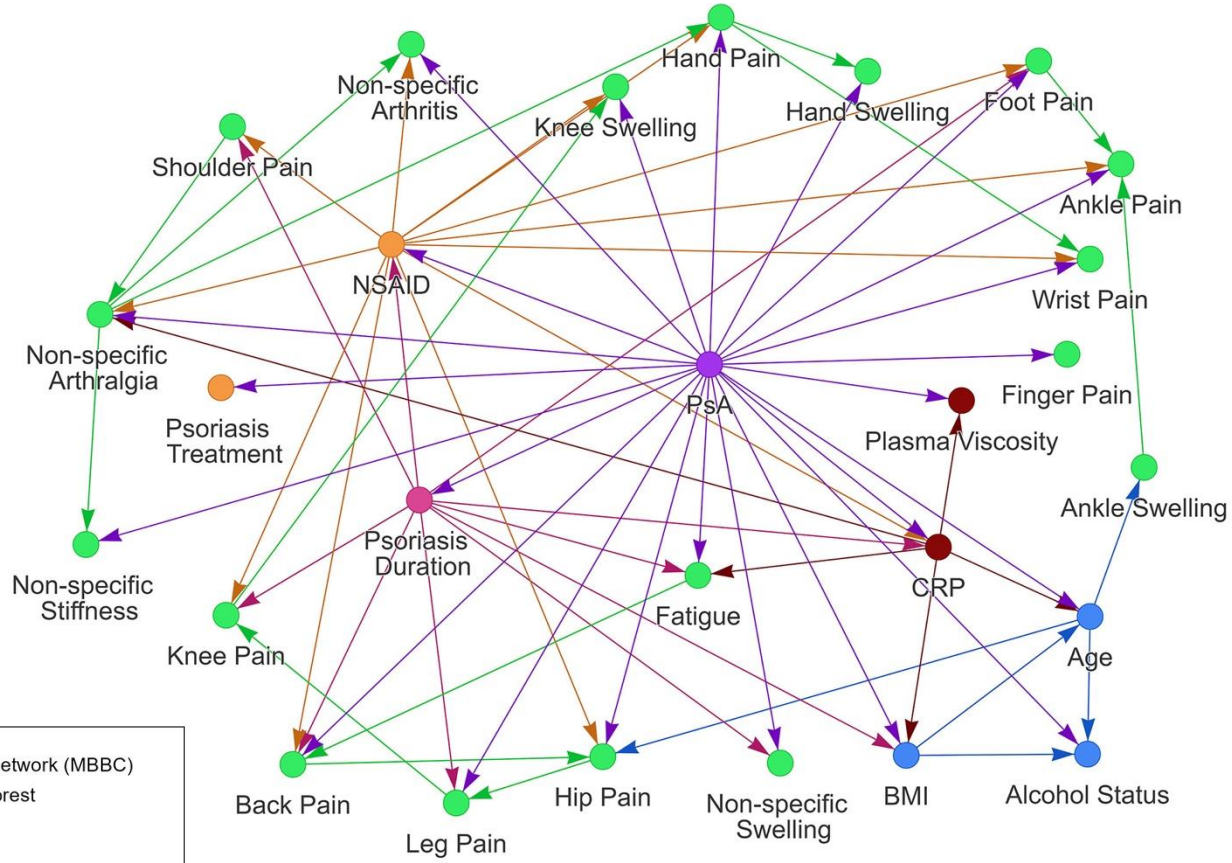
- RF: 0.851
- MBBC: 0.823

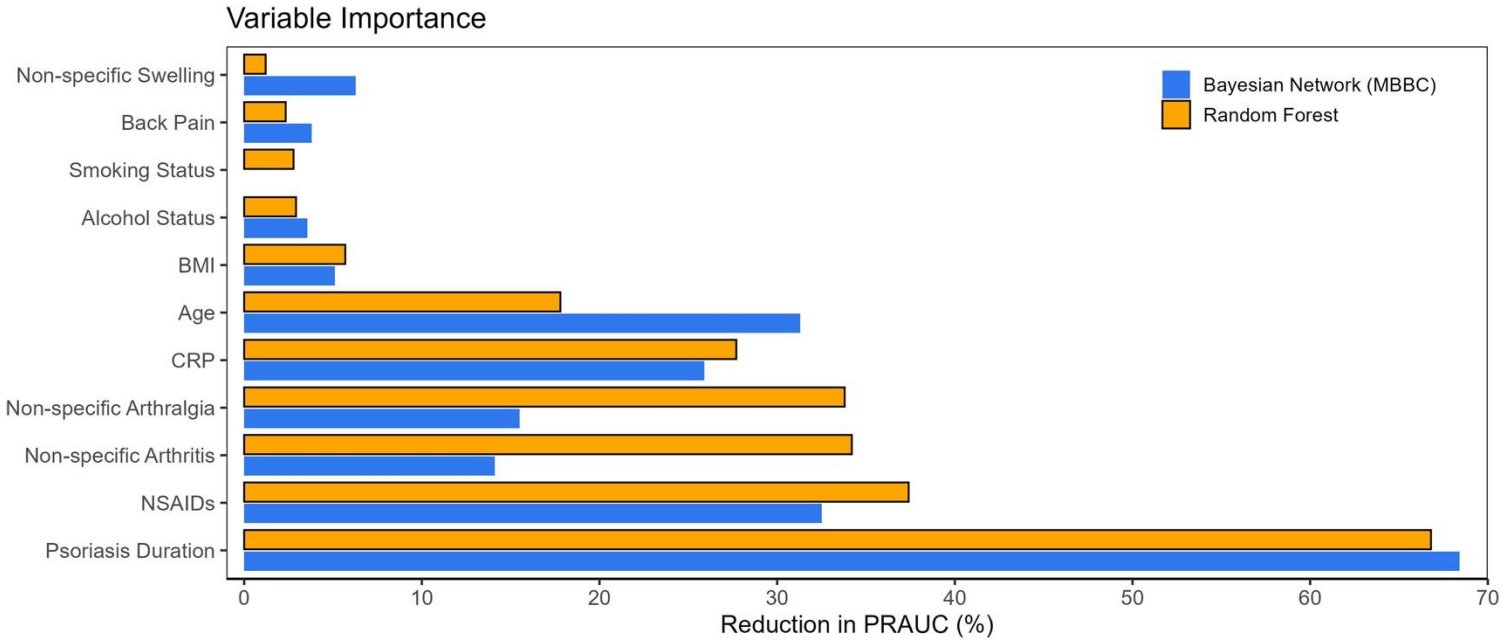
PRAUC

- RF: 0.261
- MBBC: 0.221

Useful clinical features

- Confirmed the importance of a wide range of MSK symptoms from our previous work (Green et al. 2021).
- Found increased risk in patients with a CRP blood test, even if the result was normal (with highest risk if test high).
- Found increased risk in patients with a prescription for a non-steroidal anti-inflammatory (NSAID).





Permutation variable importance:

- Randomly scramble one of the predictors.
- Refit the model with other predictors unchanged.
- Calculate change in performance metrics.
- Bigger drops in performance mean the feature is more important.

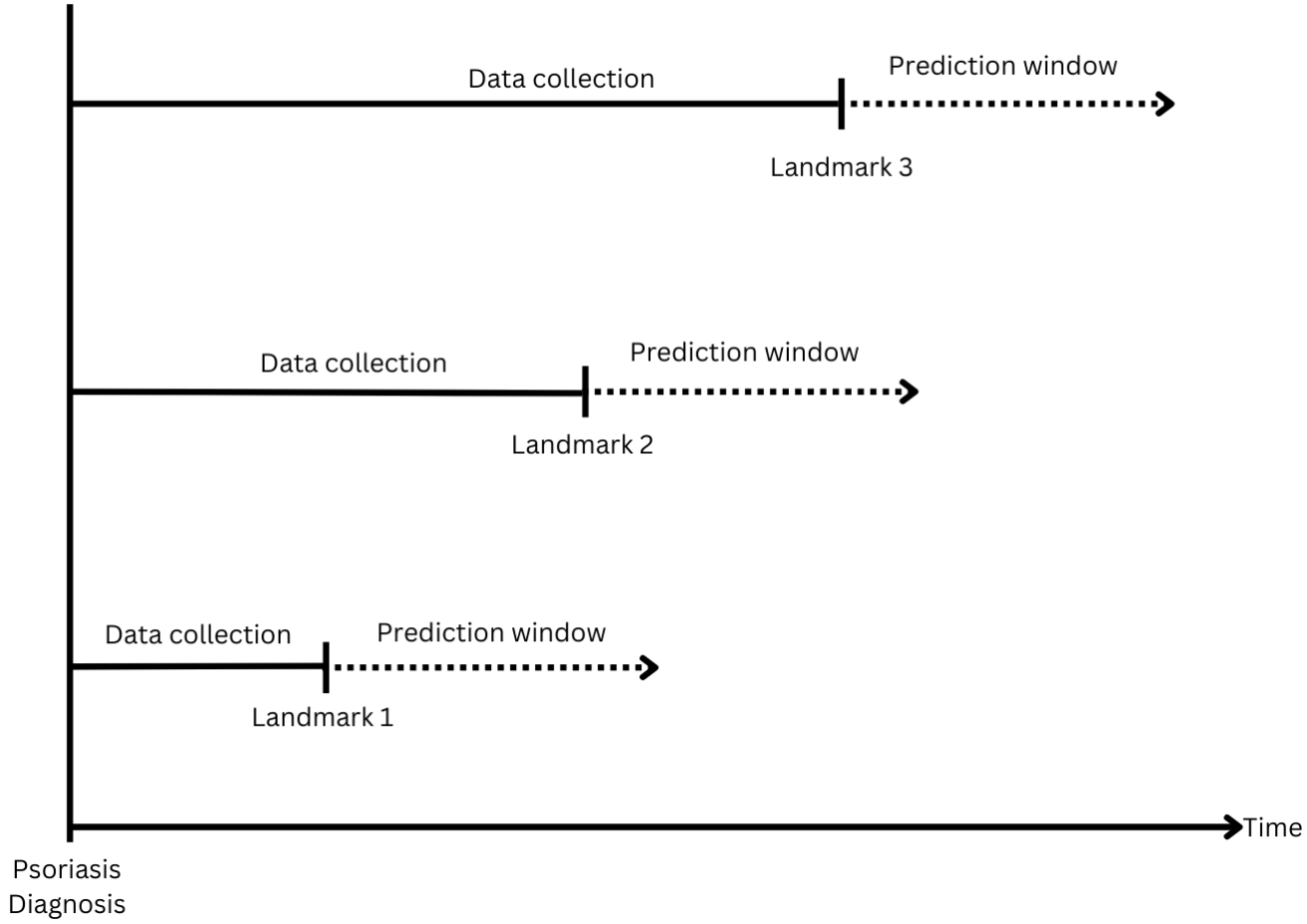
Key Take-aways

- Use of large-scale EHRs allowed us to fit more complex models than other recent PsA prediction tools.
- Both random forests and Bayesian networks can capture complex relationships between features but have different tradeoffs in terms of explainability.
- Our results could be the starting point of a clinical risk calculator, but additional validation work would be needed.
- Future research on screening tools for PsA should include indicators for nonsteroidal anti-inflammatory drug prescriptions and CRP blood tests.

Next Steps: Dynamic Model

- The analysis presented today ignores the timing of the symptoms.
  - E.g., back pain 6 months after PsO diagnosis has same impact as back pain 6 years after PsO diagnosis.

- We want to update predictions over time using the most recent covariate information.
- Landmarking is a popular framework for doing this (cf, Paige et al. 2018), but we need to modify BNs for censored data.



Questions?

See our paper recently published in ARD:
doi: 10.1016/j.ard.2025.01.051

email me: t.r.smith@bath.ac.uk

AI4CI
AI for Collective Intelligence