Prediction models in medicine

Michael C Sachs

SMARTbiomed
Pioneer Centre

Why prediction?

Context: We wish to know an unknown or future event

- Underlying disease state (diagnosis/classification)
- Future disease outcome (prognosis)
- Response to treatment

Step 1: Form predictions based on observations

- Medical tests, Questionnaires
- Genetic mutations/expression, register data

Step 2: Assess the value of the prediction model

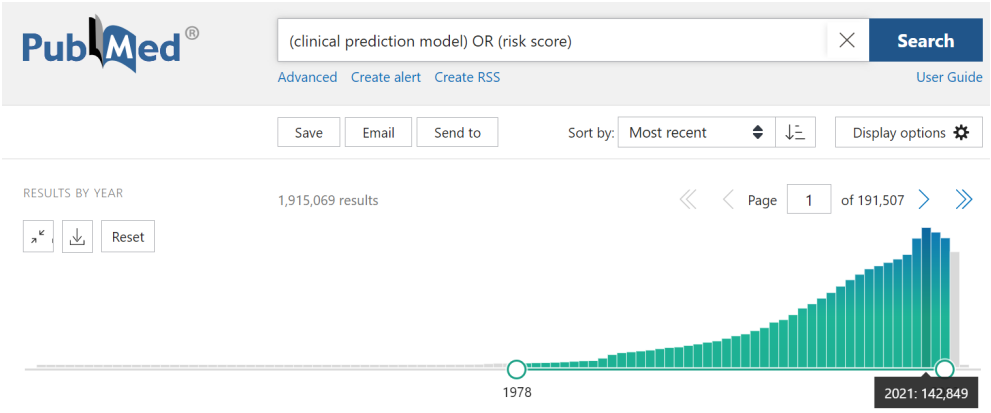
- Accuracy?
- Utility for determining treatment?

The value of a prediction model

What is the right way to assess the value of a prediction model?
Answer: It depends. What is the intended use?

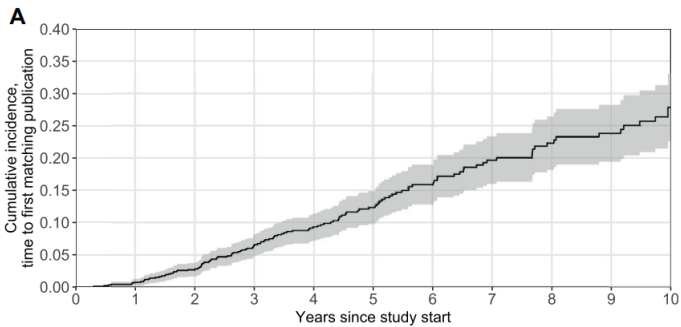
- Implement a new policy or screening program on a population level
- Guide treatments for individual patients
- Allocate funds for further research and development?

The prediction model business is booming ...



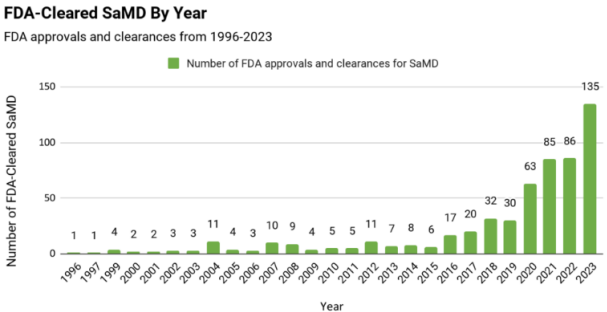
... with many more beneath the surface ...

- Out of about 1000 prediction model studies registered on ClinicalTrials.gov
- Less than 1/3 of them were published after 10 years [White et al., 2024]



... and almost none of these models are used in practice.

- Graph from a company called Orthogonal, which pulled the data from FDA public documents. SaMD = Software as a medical device
- About 2/3 of these are image processing tools in Radiology



Use cases for clinical prediction models

Intended Use	**Examples**
Diagnose	eGFR, cardiac monitors
Determine treatment	HER2, Mammaprint, OncotypeDX
Inform decisions	Framingham, SCORE2
Research only	CCI, Inflammatory burden score [Axelrad et al., 2020]

- High stakes settings require a high level of rigor and high quality evidence that using the model benefits patients, on average, compared to the standard of care (clinical utility, effectiveness).

Biomarker signature/risk score/...

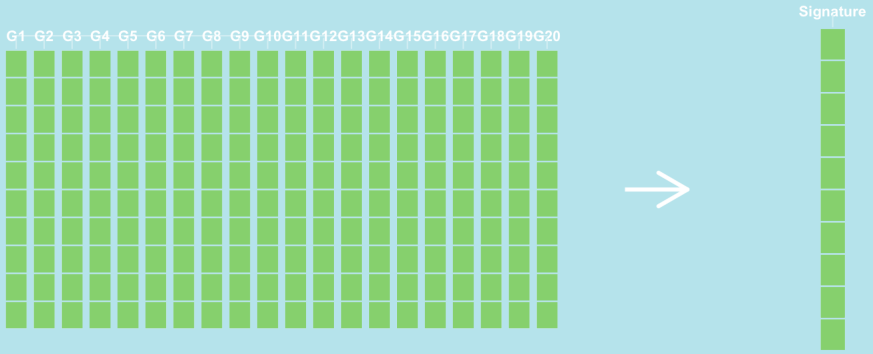


Figure: A set of 20 predictors measured on a number of subjects is translated into a one-dimensional prediction.

A risk prediction model

> *A **risk prediction model** is a transformation of multiple individual features to a one-dimensional space that coincides with the outcome space.*

- A model that *reliably predicts* an outcome may be useful for treatment selection or prognosis.
- A model that *discriminates* between groups that would be treated differently may be clinically useful.
- A signature may be continuous, binary, or take multiple discrete values.

The performance of the signature is evaluated with a variety of measures, but ...

Key point

The development of a model must be cleanly separated from its evaluation

Example in cardiovascular disease

Example: QRISK3 cardiovascular disease risk prediction model [Hippisley-Cox et al., 2008]. [`https://qrisk.org`]

- What are the *predictors*?
- What is the result of the calculator?
- How is it used?

Example in cardiovascular disease

Example: QRISK3 cardiovascular disease risk prediction model [Hippisley-Cox et al., 2008]. [`https://qrisk.org`]

- What are the *predictors*?

> Age, sex, clinical information (treatments, conditions, BP and cholesterol).

- What is the result of the calculator?

> The risk/probability of developing cardiovascular disease within the next 10 years.

- How is it used?

> "to identify people at high risk of developing CVD who need to be recalled and assessed in more detail to reduce their risk of developing CVD."

How are we using risk scores?

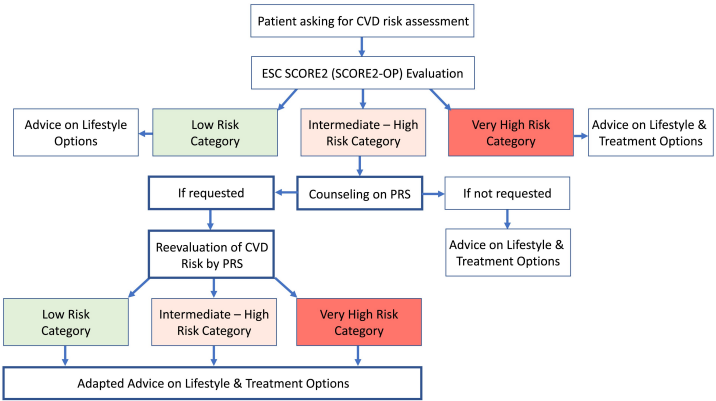


Figure: Clinical utility and implementation of polygenic risk scores for predicting cardiovascular disease: A clinical consensus statement [Schunkert et al., 2025]

The impact of reporting risk scores

Some randomized studies have shown that

- reporting risk score to patients increases their awareness and information seeking [Kullo et al., 2016]
- ... but had little to no effect on their actions that are known to reduce risk of bad outcomes [Silarova et al., 2019]

People may respond better when presented with answers to "what if" questions: e.g., *what if* I start taking statins? $\leftrightarrow$ what is $P\{Y(A = 1) = 1|X = x\}$?

Impact of using risk scores in guidelines

Instead of simply reporting risk scores, use them in guidelines/formal decision rules.

How do we decide who to treat with what?
Often, rules are: `if high risk` $\Rightarrow$ `use most aggressive treatment`

Problems with this:

1. The highest risk might include people who are doomed, where there is no treatment effect.
2. The same treatment might be the best for everyone

You need to consider the treatment effect given the risk score, or given the covariates going into the risk score

Summary and looking ahead

- We will start with the basics of developing risk predictions and assessing their accuracy
- Causal thinking in prediction modeling can help us ask the right questions
- Close collaboration with practitioners is needed so that we can get the information we need.

The Common Task Framework

- Take a subset of the sample, and hide it, ignore it, pretend the outcome doesn't exist. Give it to an "honest broker"
- Use rest of the sample to develop the model, multiple investigators can do it, trying different things
- Groups provide specifications of the models or just predictions to the honest broker
- Broker forms predictions and evaluates performance

Netflix prize, Kaggle, DREAM challenges, **Our lab this afternoon!**

References I

Jordan E Axelrad, Michael C Sachs, Jonas F Ludvigsson, Ola Olén, and SWIBREG Study Group. A novel method for quantifying intestinal inflammatory burden in inflammatory bowel disease using register data. *Clinical Epidemiology*, pages 1059–1072, 2020.

Julia Hippisley-Cox, Carol Coupland, Yana Vinogradova, John Robson, and P Brindle. Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: a validation study. *Heart*, 94(1):34–39, 2008.

Iftikhar J Kullo, Hayan Jouni, Erin E Austin, Sherry-Ann Brown, Teresa M Kruisselbrink, Iyad N Isseh, Raad A Haddad, Tariq S Marroush, Khader Shameer, Janet E Olson, et al. Incorporating a genetic risk score into coronary heart disease risk estimates: effect on low-density lipoprotein cholesterol levels (the mi-genes clinical trial). *Circulation*, 133(12):1181–1188, 2016.

References II

Heribert Schunkert, Emanuele Di Angelantonio, Michael Inouye, Riyaz S Patel, Samuli Ripatti, Elisabeth Widen, Saskia C Sanderson, Juan Pablo Kaski, John W McEvoy, Panos Vardas, Angela Wood, Victor Aboyans, Vassilios S Vassiliou, Frank L J Visseren, Luis R Lopes, Perry Elliott, and Maryam Kavousi. Clinical utility and implementation of polygenic risk scores for predicting cardiovascular disease: A clinical consensus statement of the esc council on cardiovascular genomics, the esc cardiovascular risk collaboration, and the european association of preventive cardiology. *European Heart Journal*, 46(15):1372–1383, 02 2025. ISSN 0195-668X. doi: 10.1093/eurheartj/ehae649. URL `https://doi.org/10.1093/eurheartj/ehae649`.

Barbora Silarova, Stephen Sharp, Juliet A Usher-Smith, Joanne Lucas, Rupert A Payne, Guy Shefer, Carmel Moore, Christine Girling, Kathryn Lawrence, Zoe Tolkien, et al. Effect of communicating phenotypic and genetic risk of coronary heart disease alongside web-based lifestyle advice: the inform randomised controlled trial. *Heart*, 105(13):982–989, 2019.

Nicole White, Rex Parsons, David Borg, Gary Collins, and Adrian Barnett. Planned but ever published? a retrospective analysis of clinical prediction model studies registered on clinicaltrials. gov since 2000. *Journal of Clinical Epidemiology*, page 111433, 2024.