Leids Universitair Medisch Centrum

Validation of prediction models in the presence of competing risks

STRATOS mini-symposium 2022

Nan van Geloven

Dept of Biomedical Data Sciences, LUMC

Leiden, the Netherlands

BMJ publication

Validation of prediction models in the presence of competing risks: a guide through modern methods

Nan van Geloven,[1] Daniele Giardiello,[1,2] Edouard F Bonneville,[1] Lucy Teece,[3] Chava L Ramspek,[4] Maarten van Smeden,[5] Kym I E Snell,[3] Ben van Calster,[1,6] Maja Pohar-Perme,[7] Richard D Riley,[3] Hein Putter,[1] Ewout Steyerberg,[1,8] on behalf of the STRATOS initiative

For numbered affiliations see end of the article

Correspondence to:
E Steyerberg
e.steyerberg@erasmusmc.nl
(ORCID 0000-0002-7787-0122)

Additional material is published online only. To view please visit the journal online.

Cite this as: BMJ 2022;377:e069249
http://dx.doi.org/10.1136/bmj-2021-069249

Accepted: 08 April 2022

Thorough validation is pivotal for any prediction model before it can be advocated for use in medical practice. For time-to-event outcomes such as breast cancer recurrence, death from other causes is a competing risk. Model performance measures must account for such competing events. In this article, we present a comprehensive yet accessible overview of performance measures for this competing event setting, including the calculation and interpretation of statistical measures for calibration, discrimination, overall prediction error, and clinical usefulness by decision curve analysis. All methods are illustrated for patients with breast cancer, with publicly available data and R code.

In these settings, prediction models should target the cumulative incidence (or absolute risk[3]) of the adverse event, which is defined as the probability of the event of interest occurring by a particular time point with no other competing event occurring earlier. In the breast cancer example, the cumulative incidence of recurrence at five years is the risk of developing a recurrence within five years, taking into account that patients who die within five years cannot develop recurrence anymore. Failing to account for competing events during model development leads to overestimation of the cumulative incidence.[4] The higher the risk of the competing event, the more pronounced the overestimation. Crucially, failure to account for competing events during validation leads to a distorted view on model performance, especially for calibration.

Such distortion was recently revealed for an internationally recommended prediction model of kidney failure, which systematically overestimated the absolute risk of kidney failure at five years in patients with advanced chronic kidney disease. The absolute overestimation by 10 percentage points on average and by 37 percentage points in the highest risk group could have resulted in overtreatment of patients, which therefore led to the conclusion that

Collaborative project of TG6 (prediction models) and TG8 (survival analysis)

joint work with:

Daniele Giardiello

Edouard F Bonneville

Lucy Teece

Chava Ramspek

Maarten van Smeden (TG6)

Kym Snell

Ben van Calster (TG6)

Maja Pohar-Perme (TG8)

Richard D Riley

Hein Putter (TG8)

Ewout Steyerberg (TG6)

Motivation and aim

Ageing population -> more comorbidities -> competing events more common

Prediction models should account for competing events, both at development and at validation

Validation guidance currently spread out over many technical papers -> low uptake in applied studies

Aim: provide **accessible overview** of performance measures for validating competing risk prediction models

Setting (1)

A prediction model has already been developed

It allows calculating estimates of absolute risk for new patients

We want to externally validate this model

Setting (2)

Interest is in the primary event occurring by a certain (or several) time point(s):

$$F_1(s \mid \mathbf{z}_i) = P(T \leq s, D = 1 \mid \mathbf{z}_i)$$

T time to first event

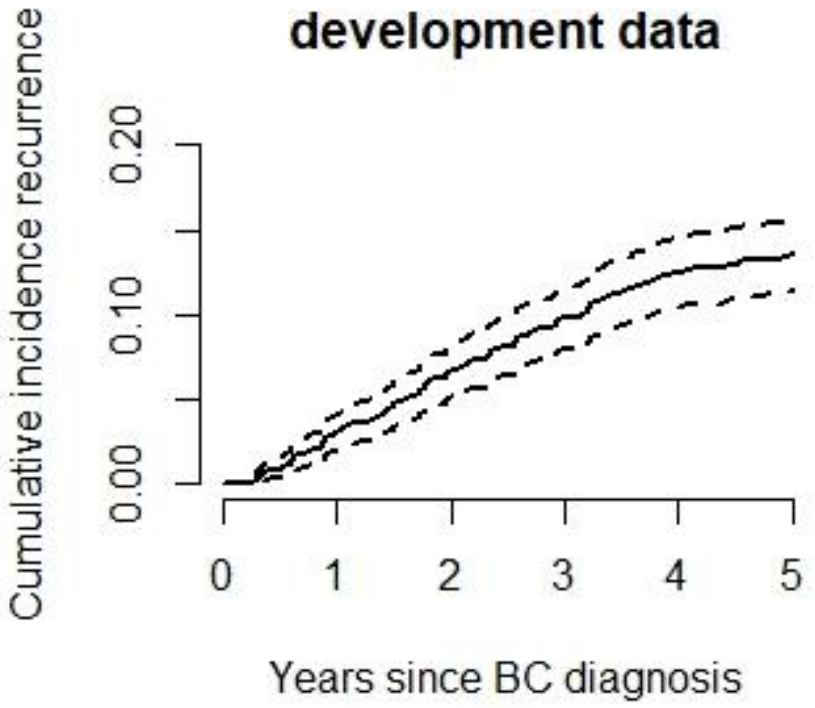
s prediction horizon

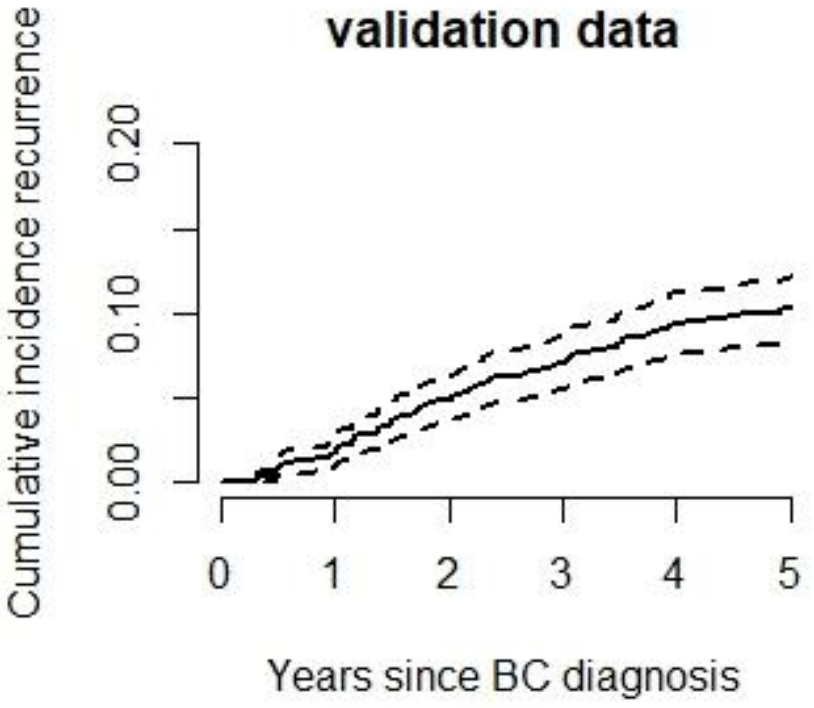
D event status (0,1,2,..)

Z covariate vector

Case study

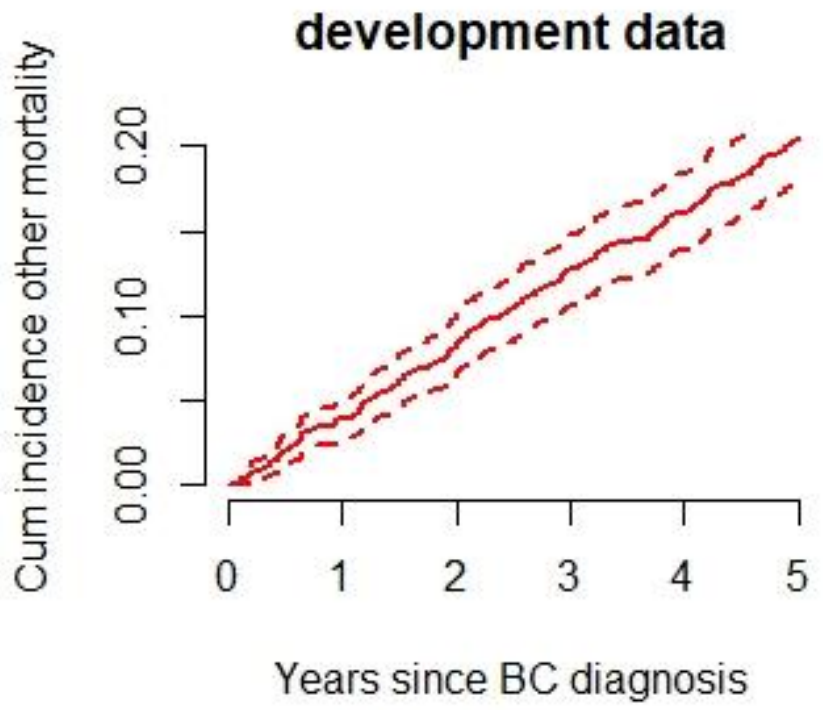
- predicting absolute risk of breast cancer recurrence 5 years after diagnosis
- mortality from other causes is a competing event
- random samples (n=1000) 2 Dutch cohorts for data sharing

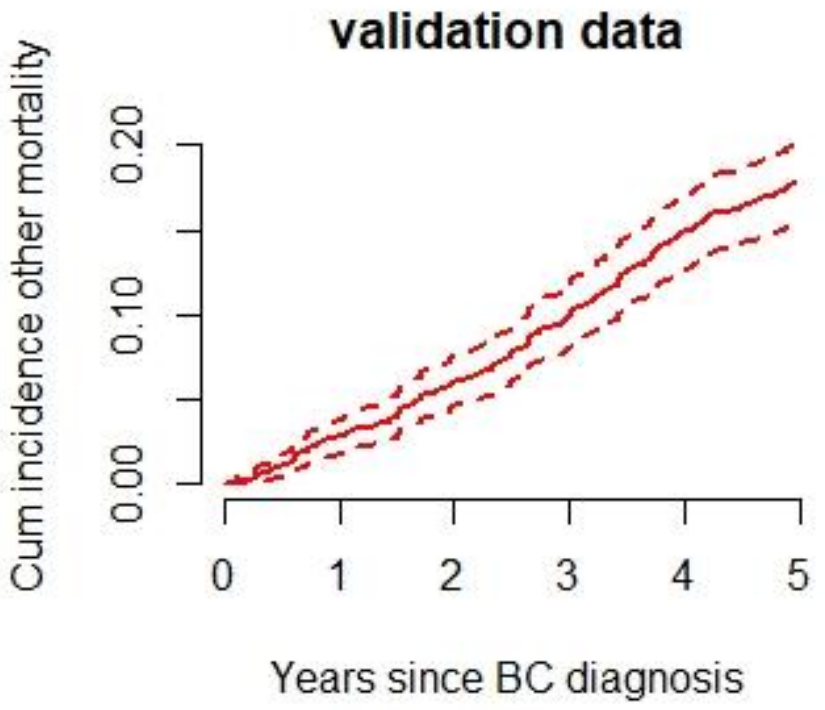




Case study

- predicting absolute risk of breast cancer recurrence 5 years after diagnosis
- mortality from other causes is a competing event
- random samples (n=1000) 2 Dutch cohorts for data sharing





Validation aspects

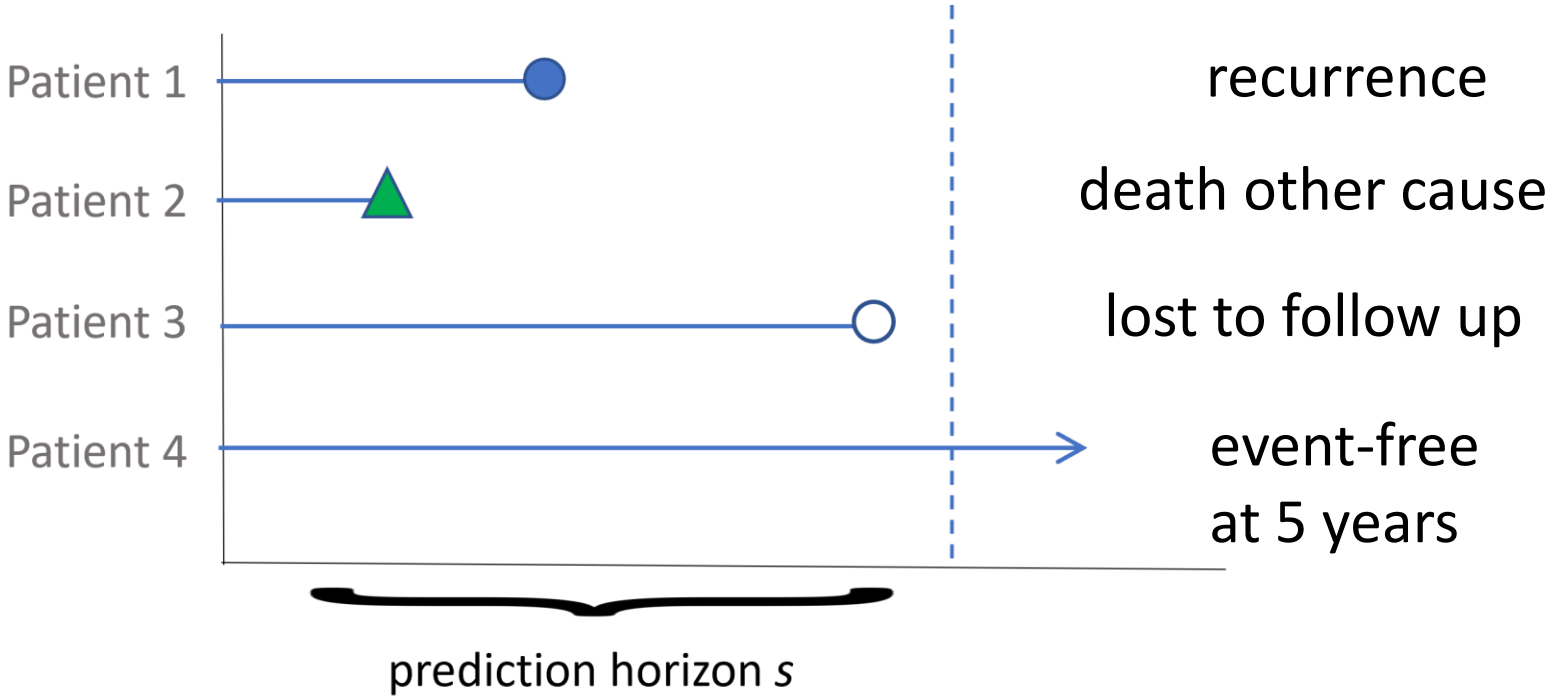
Calibration: How close are estimated risks to observed outcome proportions?

Discrimination: How well does the model separate those who experience the primary event earlier than others?

Prediction error: How close are estimated risks to the observed primary event indicators? How much closer compared to null model?

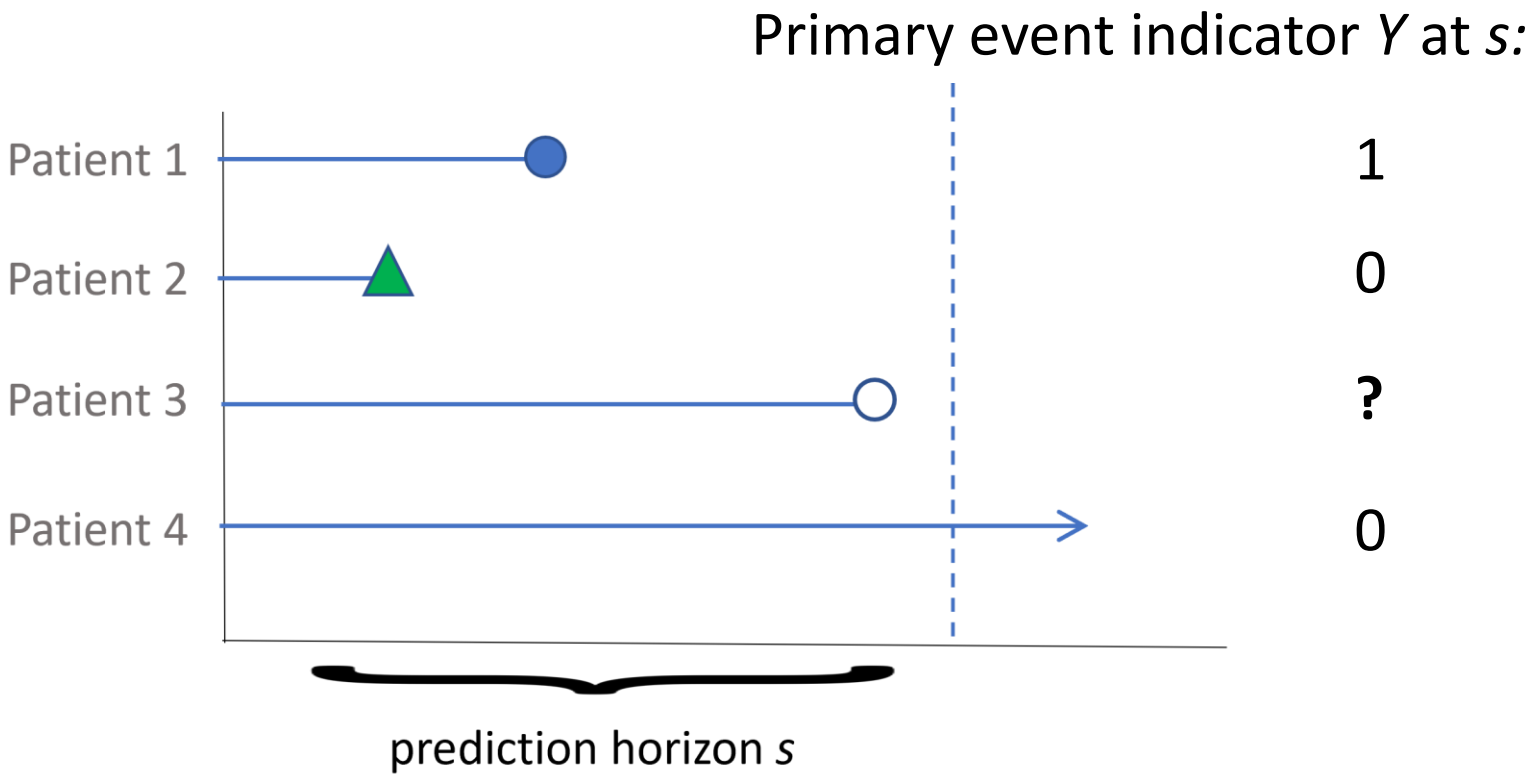
Decision curve analysis: What is the net result from correctly and falsely classified high risk patients?

Main challenges in validation data



How to incorporate competing events ▲?
How to incorporate censored observations ○?

Calibration: observed vs expected outcome proportions



Expected: average of patient specific risk estimates
Observed: Aalen-Johansen estimator

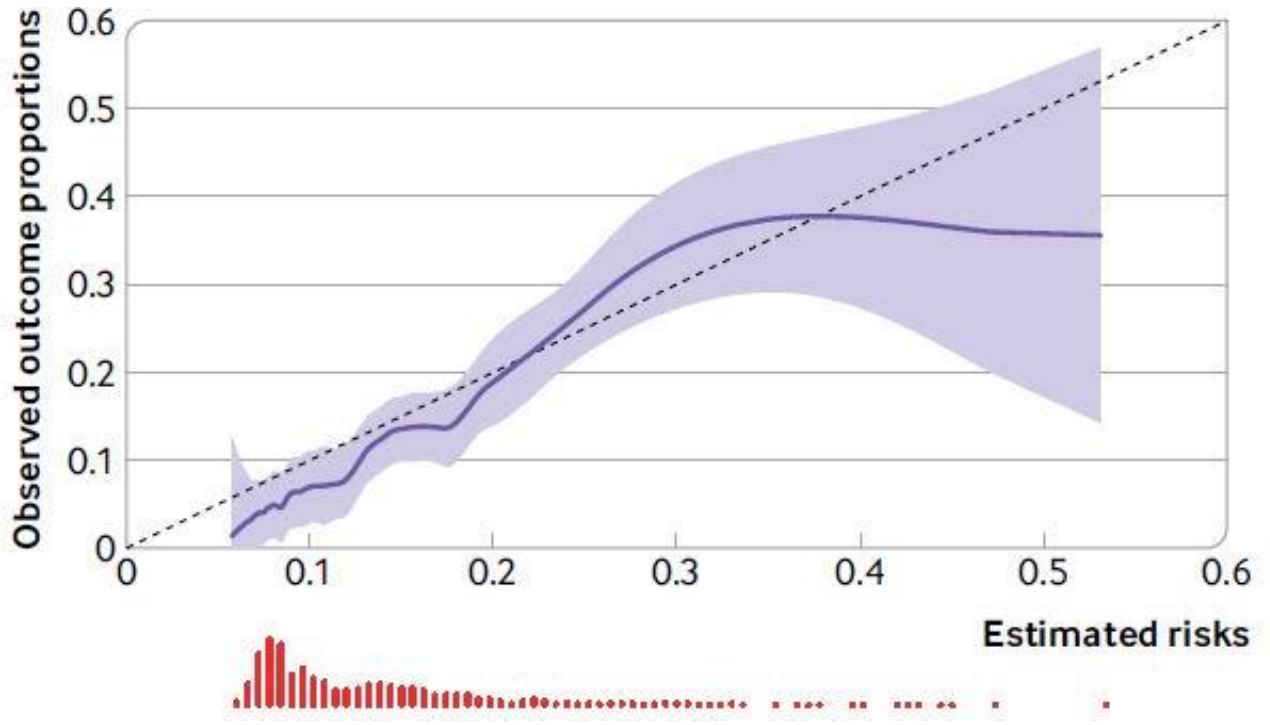
O/E ratio 0.81 (95% CI 0.62 to 0.99) -> slight overestimation

Calibration curve with pseudo-obervations

Replace primary event indicators by pseudo-observations:

$$\tilde{Y}_i(s) = n\widehat{F}_1(s) - (n-1)\widehat{F}_1^{-i}(s)$$

Draw smooth curve of pseudo-observations versus estimated risks

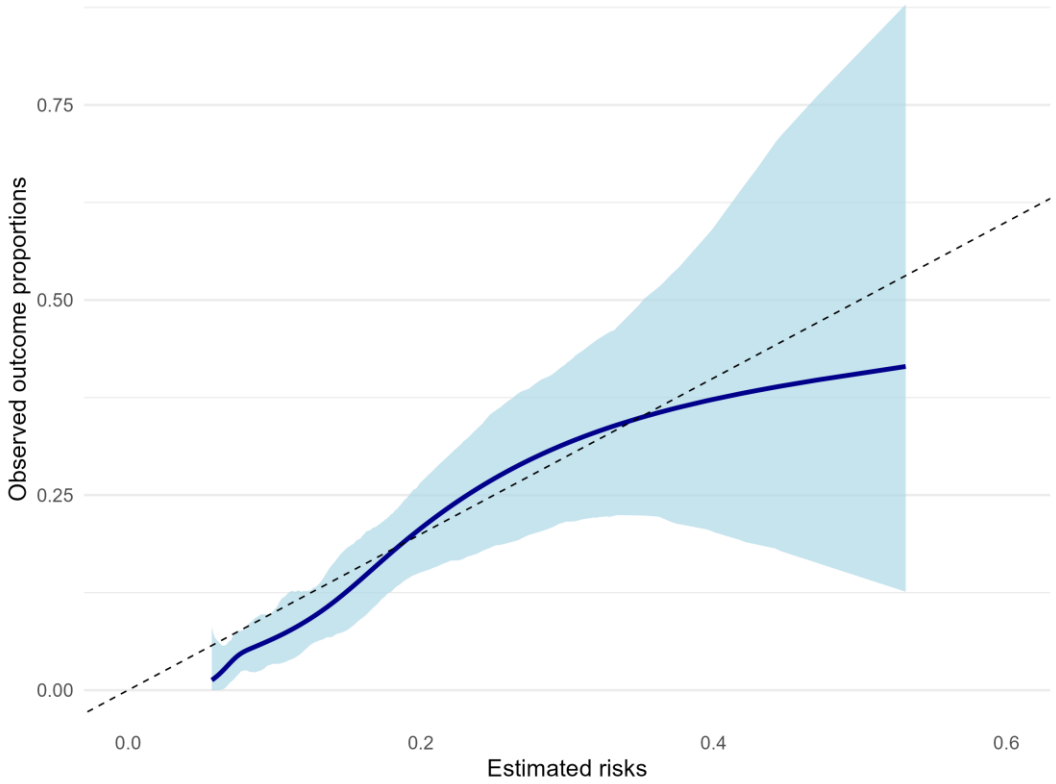


Gerds et al 2014,
Royston 2014

Calibration curve with flexible regression

Fit a new, flexible (eg spline-based), regression model to the validation data with cloglog transformed risks as covariate

Predict the observed outcome proportions at *s* from this model



Austin et al 2022

Discrimination: c-index

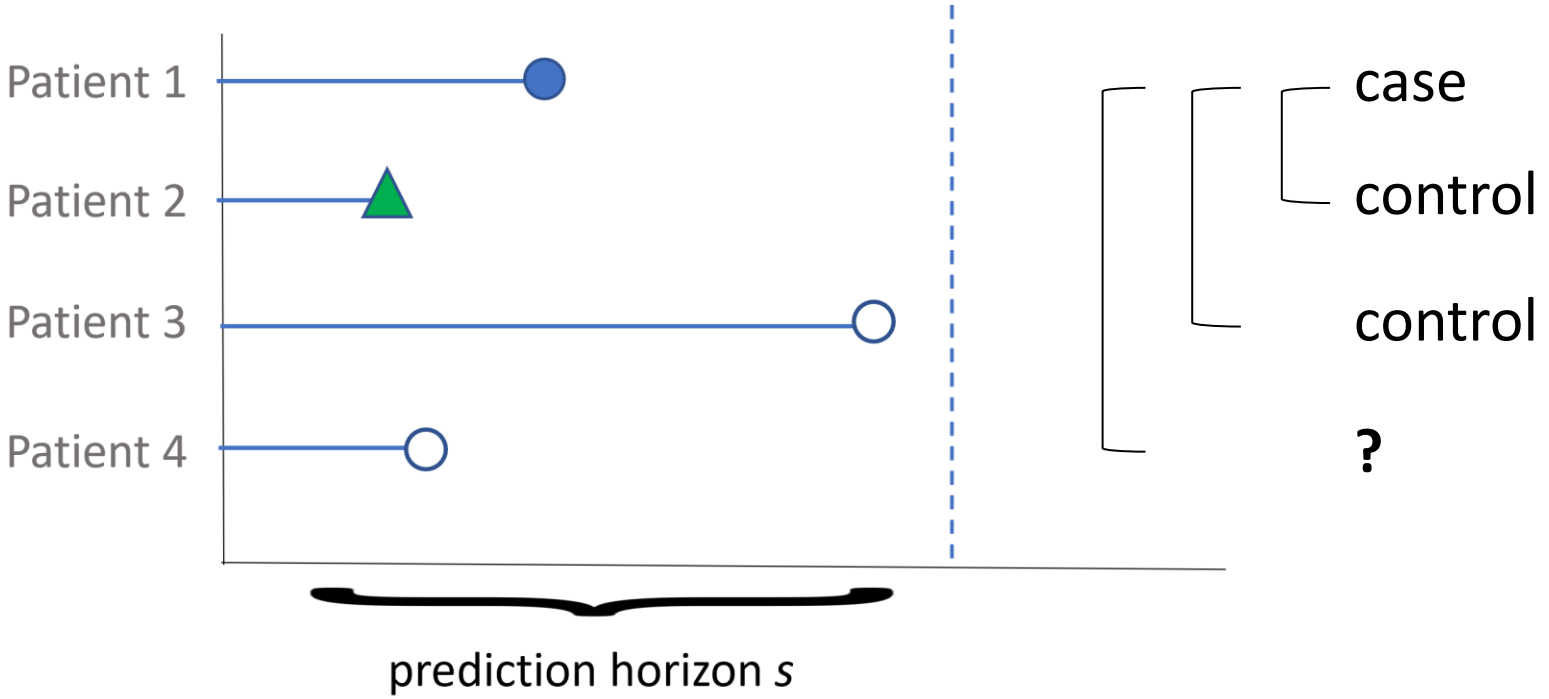
Does the model assign higher risk estimates to patients who experience the primary event earlier than others?

Cases: event of interest

Controls: event later than event of case or competing event

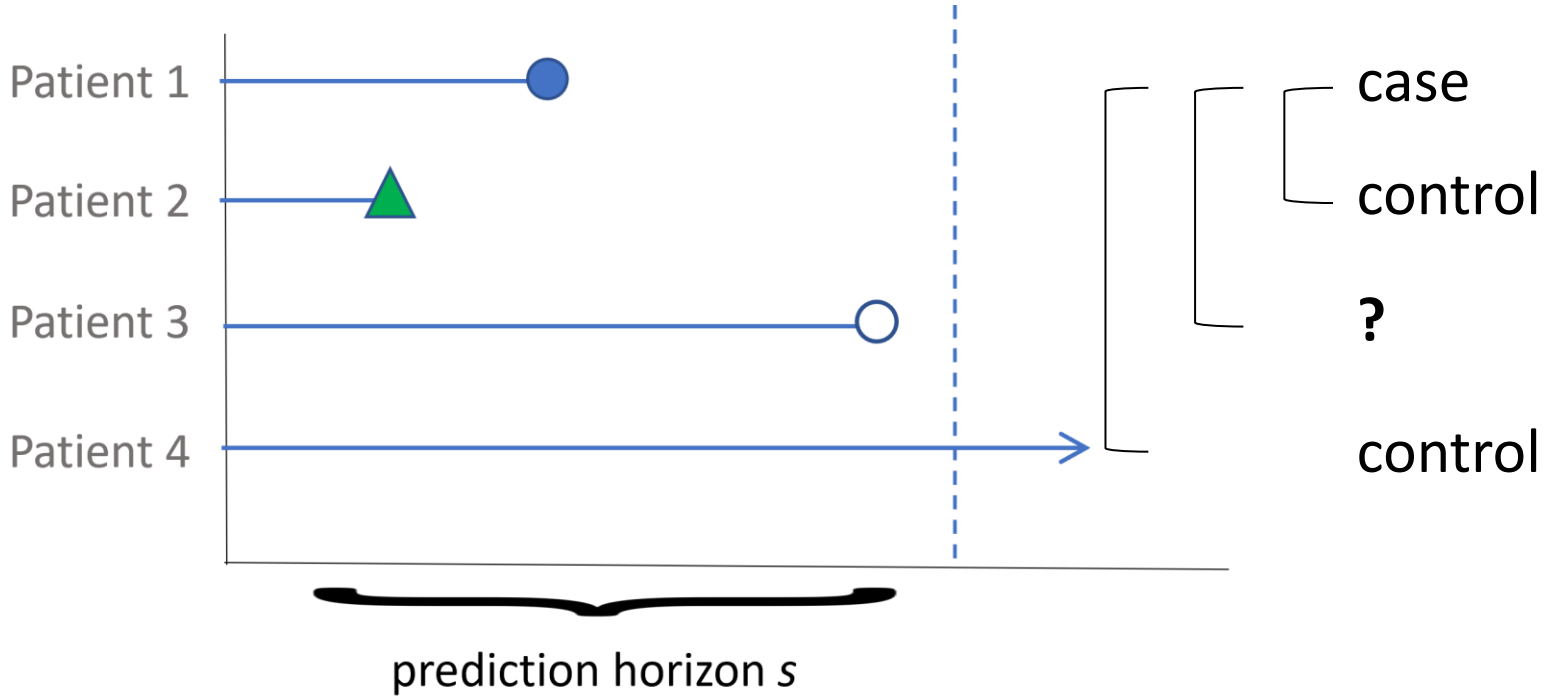
c-index: proportion of pairs where case has highest risk estimate

Discrimination: c-index censored observations



- Either ignore pair 1-4 (similar to Harrell's c)
- Or redistribute observation weights by inverse probability of censoring weighting (Wolbers et al 2014)

Discrimination: cumulative/dynamic AUCt



Also here, inverse probability weighting is proposed to account for censored patients (Blanche et al 2013)

Full list of performance measures in the paper

Calibration

- O/E ratio
- calibration plot
- squared bias/ICI/E50/E90/Emax
- calibration intercept and slope

Discrimination

- C index + C/D AUCt

Prediction error

- Brier score + scaled Brier score

Decision curve analysis

- Net benefit + Decision curve

<u>Approach to censoring</u>

- Aalen-Johansen estimator
- pseudo-observations
- secondary flexible regression model
- IPCW

Tutorial format

main text

supplement

case study

github

Tutorial format: example c-index

main text

..compare pairs where one individual has the primary event within the prediction horizon and the other either has the primary event later or experiences a competing event. Such a pair is considered concordant when the first individual has the higher estimated risk. C index is the proportion of concordant pairs.

supplement

$$\mathcal{C}_1(s) = P\{\pi_1(s \mid \mathbf{z}_i) > \pi_1(s \mid \mathbf{z}_j) \mid D_i = 1, T_i \leq s, (T_i < T_j \cup D_j \notin \{0,1\})\}$$

case study

In the breast cancer data, the c index calculated for the time range until five years of follow-up was 0.71 (95% confidence interval 0.67 to 0.76)

github

```
cindex_csh <- pec::cindex(
  object = fit_csh,
  formula = Hist(time, status_num) ~ 1,
  cause = primary_event,
  eval.times = horizon,
  data = vdata
)$AppCindex$CauseSpecificCox
```

Github

https://github.com/survival-lumc/ValidationCompRisks

The repository contains the following code:

- Prediction_CSC_minimal.R : the companion (minimal) script for the manuscript, illustrating external validation of a prediction model. The file uses a cause specific hazards prediction model. To reproduce all mean tables and figures of the manuscript, this script is sufficient.
- Prediction_CSC.md : a markdown document containing a more in-depth version script, with details on model development, descriptive tables and plots. The RMarkdown source code (.Rmd) is here.
- Additional code to alternatively develop a competing risk prediction model using the subdistribution hazard approach (Fine & Gray) is here. The Rmarkdown source code (.Rmd) is here. A more concise R source code (.R) is here.
- sharing_CSC_model.R : example/template of how to share a cause-specific hazards prediction model for external validation, without having to share the original development data.

Some reflections on writing statistical guidance for broad readership

- Collect a great group of experts from different perspectives. In our case prediction / survival / epidemiology

- Start out with a glossary

- Use a technical ‘shadow’ document / appendix

- Tailor readers with alternative ways to comprehend (text / formulas / code / case study) so they can follow their own learning path

Thank you

https://github.com/survival-lumc/ValidationCompRisks

n.van_geloven@lumc.nl