UNIVERSITÄT BERN

Prognostische Modelle in der klinischen Forschung – Ein Blick in die Glaskugel

DCR-CTU lecture

André Moser, Senior Statistician, CTU Bern

andre.moser@ctu.unibe.ch

27. September 2023

Aims

In this DCR-CTU lecture we dare a look into the crystal ball and discuss **some key elements** of prognostic modelling from a **regression modeling perspective.**

No (or hopefully less) crystal ball anymore

Premise of prognostic models in clinical research

UNIVERSITÄT BERN

Prognostic models are tools

- (shared) **decision making** between patients and clinicians
- a better understanding of **disease determinants**
- **health economic evaluation** and **benchmarking**
- which are **communicable, implementable** and **reproducible**

Instruction manual of the crystal ball

UNIVERSITÄT BERN

Development phase
Prognosis aim, study outcomes, study design, candidate predictor selection, statistical model

Validation phase
Performance assessment and internal validation
Performance in new patients (external validation)

Impact phase
Does the model improves decision making and patient outcomes?

Development phase

Study aim for clinical prediction models

ELSEVIER

Journal of Clinical Epidemiology 142 (2022) 230–241

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Mortality prediction in intensive care units including premorbid functional status improved performance and internal validity

André Moser[a,*], Matti Reinikainen[b], Stephan M. Jakob[c], Tuomas Selander[d], Ville Pettilä[e], Olli Kiiski[f], Tero Varpula[g], Rahul Raj[h,1], Jukka Takala[c,1]

[a] *CTU Bern, University of Bern, Mittelstrasse 43, 3012 Bern, Switzerland*
[b] *Department of Anaesthesiology and Intensive Care, Kuopio University Hospital and University of Eastern Finland, Kuopio, Finland*
[c] *Department of Intensive Care Medicine, Bern University Hospital, University of Bern, Bern, Switzerland*
[d] *Science Service Center, Kuopio University Hospital, Kuopio, Finland*
[e] *Division of Intensive Care, University of Helsinki and Helsinki University Hospital, Helsinki, Finland*
[f] *Health and Care, Benchmarking Services, TietoEvry, Helsinki, Finland*
[g] *Division of Intensive Care, University of Helsinki and Helsinki University Hospital, Helsinki, Finland*
[h] *Department of Neurosurgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland*

Accepted 17 November 2021; Available online 22 November 2021

Study aim for clinical prediction models

UNIVERSITÄT BERN

Clinical practice: Diagnosis, clinical decision making

Economic evaluation: Benchmarking

Public health: Preventive interventions

Research: Inclusion in RCTs

Study population, outcomes and candidate predictors are tailored to the prognosis aim.

Study design

UNIVERSITÄT BERN

Design	Example	Pros	Cons
Prospective designs	Multicenter RCTs	Data quality	Generalizability
Retrospective design	Patient records	Simple, costs	Patient selection
Registries	Cancer / Insurance	Large, coverage	Outcome assessment
Case-control	Rare diseases	Simple, costs	Choice of controls

Sample size

Dear Statistician.

I want to develop a prediction model, how much patients do I need?

Sample size

UNIVERSITÄT BERN

- Often used **rule of thumb** «10 events per variable (EPV)» is too simple (doi: 10.1177/0962280218784726)

- Sample size depends on more: For example, outcome proportion and expected predictive performance

- Guidelines

Riley et al. Calculating the sample size required for developing a clinical prediction model (doi: 10.1136/bmj.m441)

Or ask your statistician of trust at CTU Bern for support!

Candidate predictor selection

Candidate predictor selection

UNIVERSITÄT BERN

«Ideally, candidate predictors are selected without studying the predictor-outcome relationsship in the data under study.»

Steyerberg, Clinical Prediction Models, Springer

Good starting point:
Choose 5-20 predictors based on literature and expert knowledge

‘Automated’ predictor selection

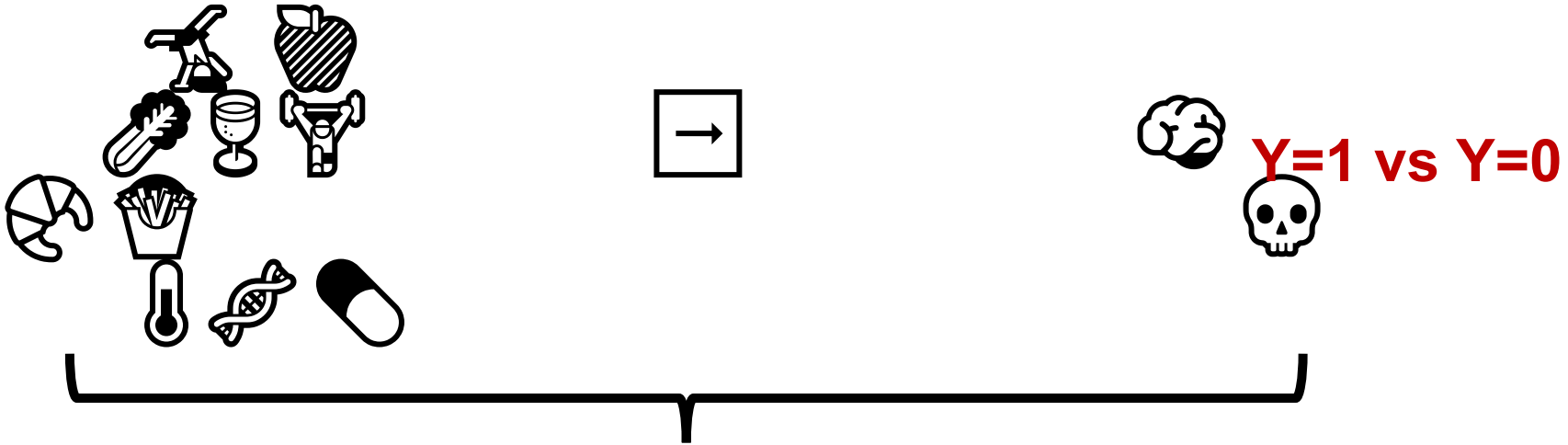
‘Automated’ predictor selection

’Statistical’ selection procedures: **Doable, but…**

- introduce ‘**variable selection bias**’ based on arbitrary cut-offs ($p<0.05$)
- might lead to **biased coefficients**
- moves towards an **estimation** problem for exploration, not a prediction problem

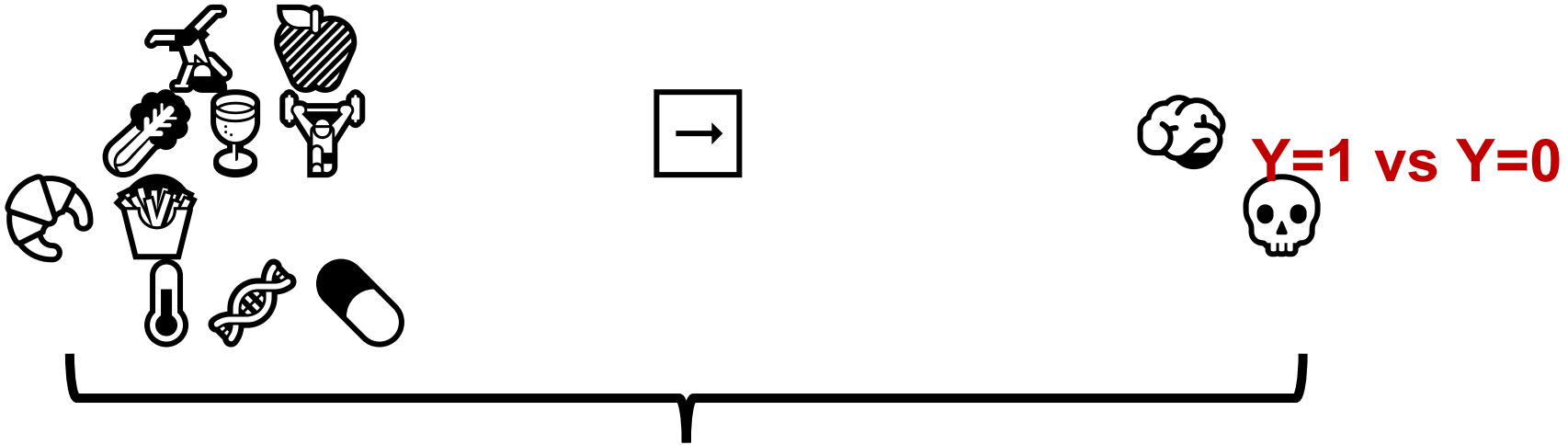
Depends study aim and context (e.g. large studies)!

Statistical model



Multivariable prediction model (here: binary outcome)

Statistical model



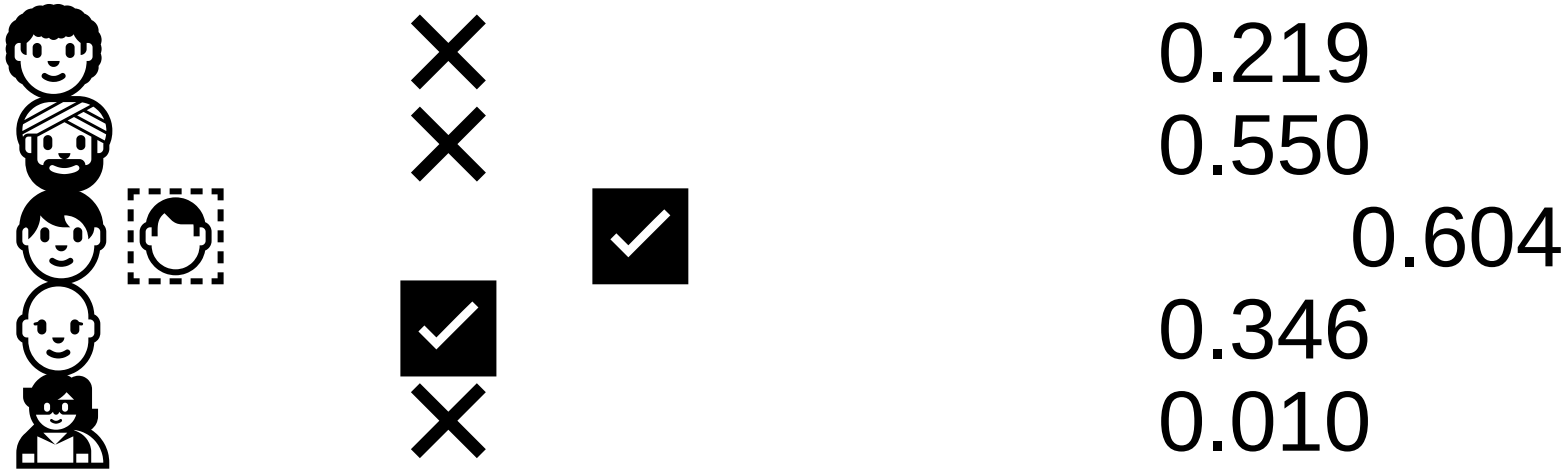
Multivariable prediction model (here: binary outcome)

Statisticians at work!

Output from statistical model

Observed outcome

P(Y=1)

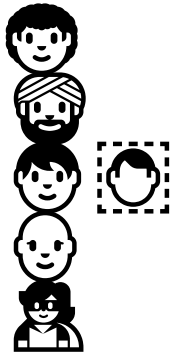


0.219
0.550
0.604
0.346
0.010

Validation phase

Validation

Observed death

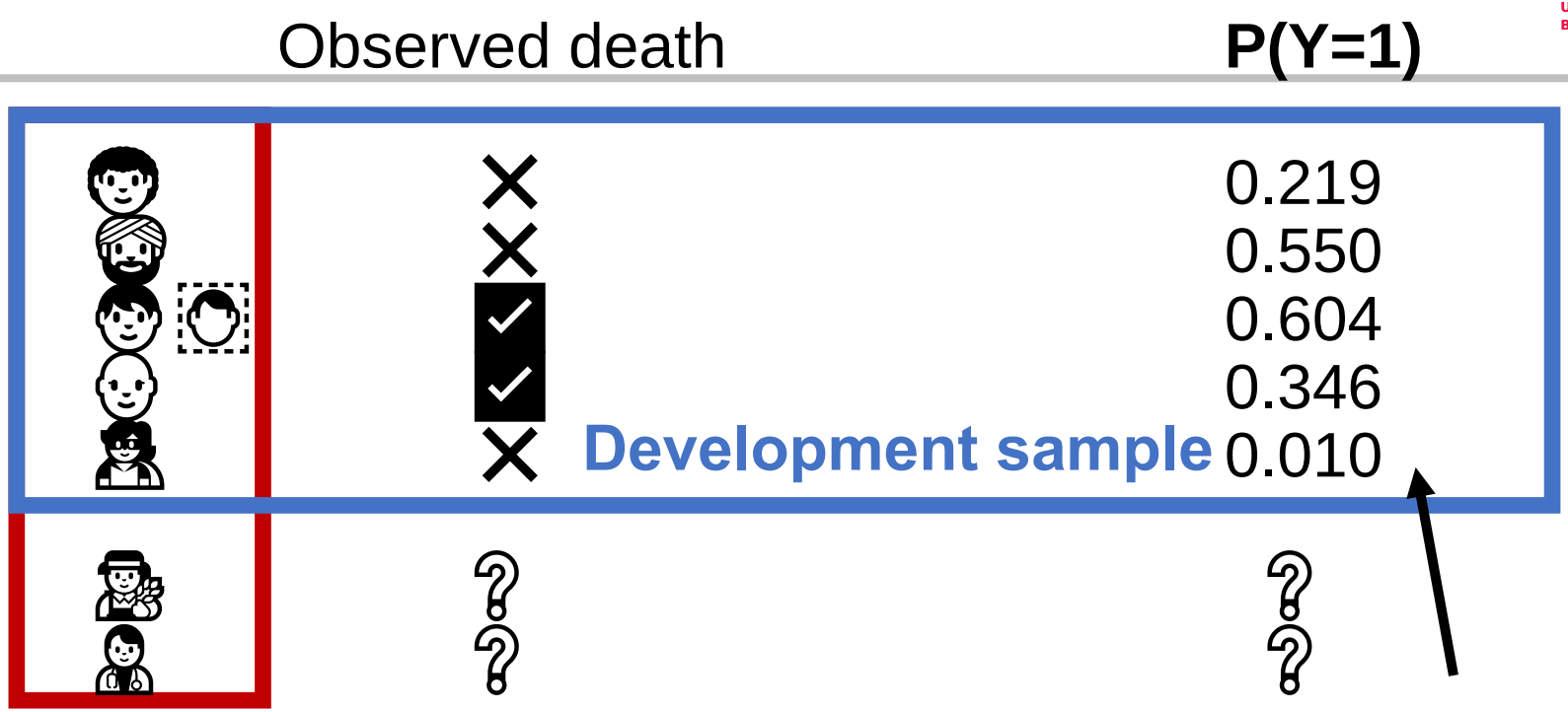


Good/poor?

P(Y=1)

0.219
0.550
0.604
0.346
0.010

Terminology: Optimism



True population

Too optimistic estimates?

“What you see may not be what you get.”

Optimism and overfitting

UNIVERSITÄT BERN

- **Optimism**: True performance vs development performance (**internal validation process**)

- Overfitting is a statistical problem which occurs when **too many variables** are fitted in a model with **too few events**

- Overfitting leads to optimism

- **Internal validation** and “**shrinkage methods**” try to minimize optimism

Validation process

UNIVERSITÄT BERN

Validation phase

Internal validation:

Optimism
Bootstrapping
Cross validation
Shrinkage

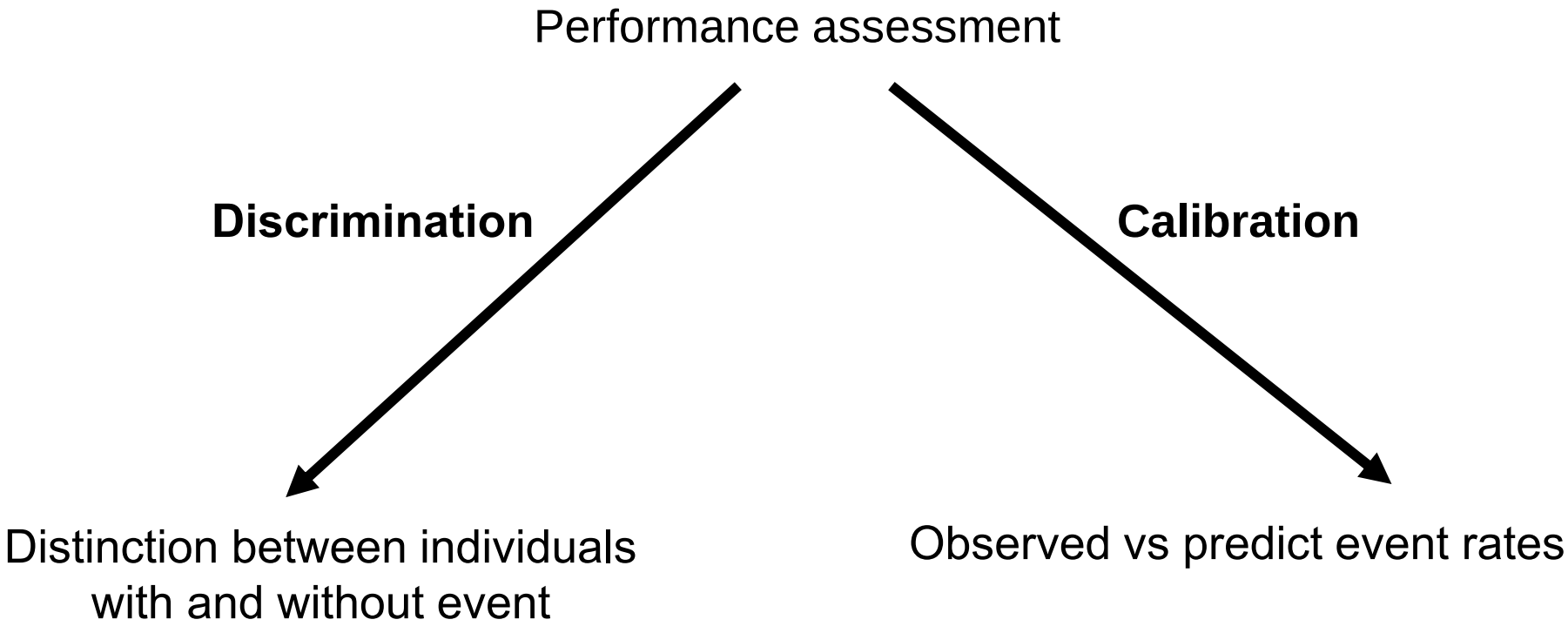
Performance assessment:

Overall performance
Discrimination
Calibration

External validation:

Performance in new patients?
Performance in different centers/periods?

Performance assessment

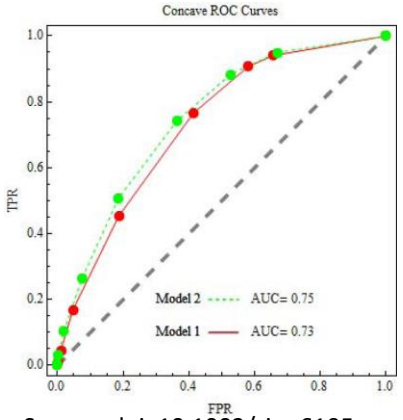


Discrimination

- AUC=**area under receiver operating characteristics curve**

- Binary outcomes: **c-statistic=AUC**

- A value between **0.5** (uninformative model, coin flip) and **1** (perfect discrimination)



Source: doi: 10.1002/sim.6195

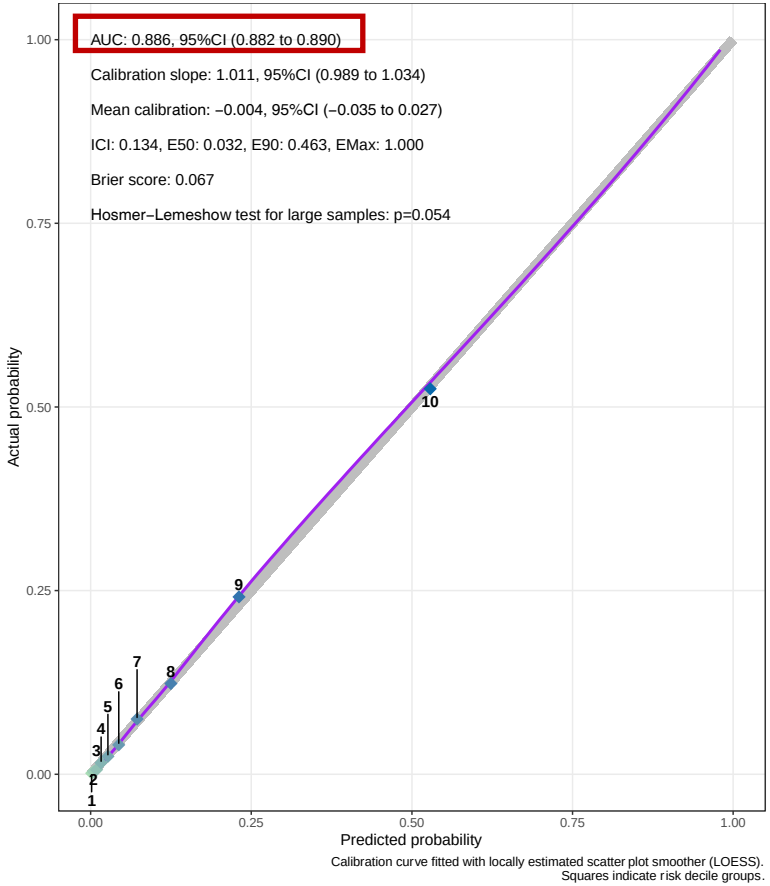
Levels of calibration

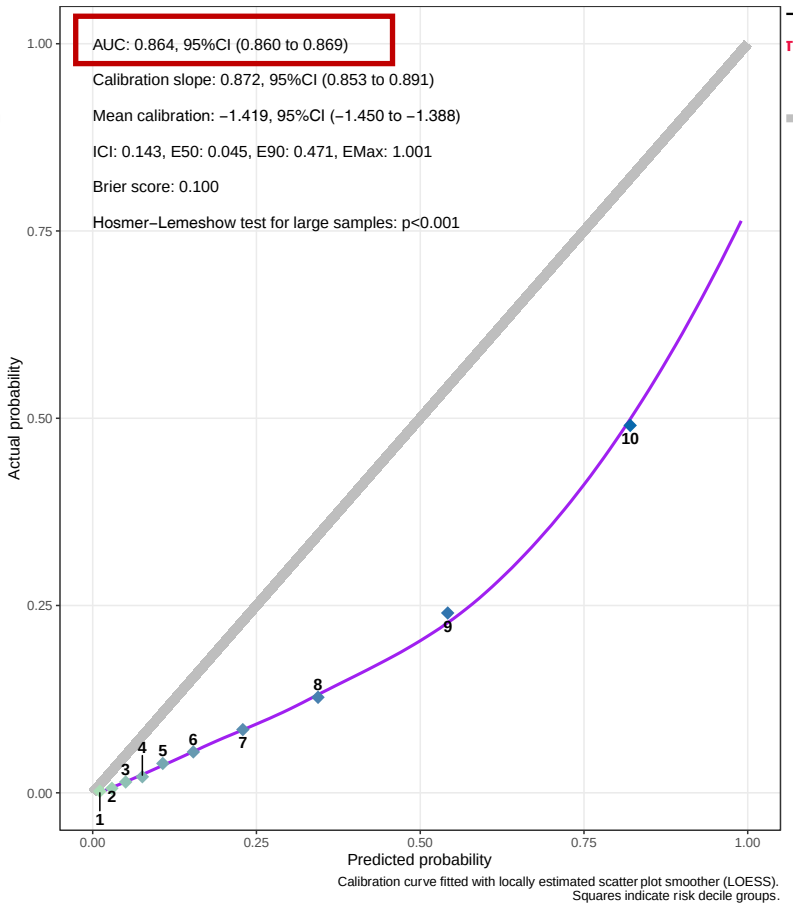
UNIVERSITÄT BERN

Level 1	Level 2	Level 3	Level 4
Mean	Weak	Moderate	Strong
Observed event rate equals **average** predicted risk?	Over- or underestimation of risk?	Predicted risks correspond to observed event rate?	Predicted risks correspond to observed event rate **among all covariate patterns**

https://doi.org/10.1016/j.jclinepi.2015.12.005

Discrimination and calibration





Source: 10.1016/j.jclinepi.2021.11.028

The myth about data splitting for 'validation'

ELSEVIER

Check for updates

Journal of Clinical Epidemiology 103 (2018) 131–133

Journal of Clinical Epidemiology

COMMENTARY

Validation in prediction research: the waste by data splitting

Ewout W. Steyerberg[a,b,*]

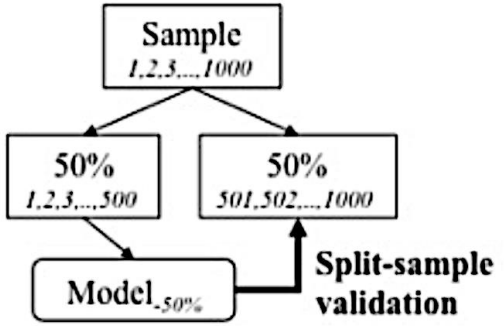
[a]*Professor of Clinical Biostatistics and Medical Decision Making, Chair, Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands*

[b]*Professor of Medical Decision Making, Department of Public Health, Erasmus MC, Rotterdam, The Netherlands*

Accepted 24 July 2018; Published online 29 July 2018

Abstract

Accurate prediction of medical outcomes is important for diagnosis and prognosis. The standard requirement in major medical journals is nowadays that validity outside the development sample needs to be shown. Is such data splitting an example of a waste of resources? In large samples, interest should shift to assessment of heterogeneity in model performance across settings. In small samples, cross-validation and bootstrapping are more efficient approaches. In conclusion, random data splitting should be abolished for validation of prediction models.



https://doi.org/10.1016/j.jclinepi.2015.04.005

The myth about data splitting for ‘validation’

UNIVERSITÄT BERN

- **Inefficient** use of data (poorer model is developed)

Better solution for ‘validation’:

- For large samples: **Internal-external** validation

- For small samples: **Bootstrapping** or **cross-validation**

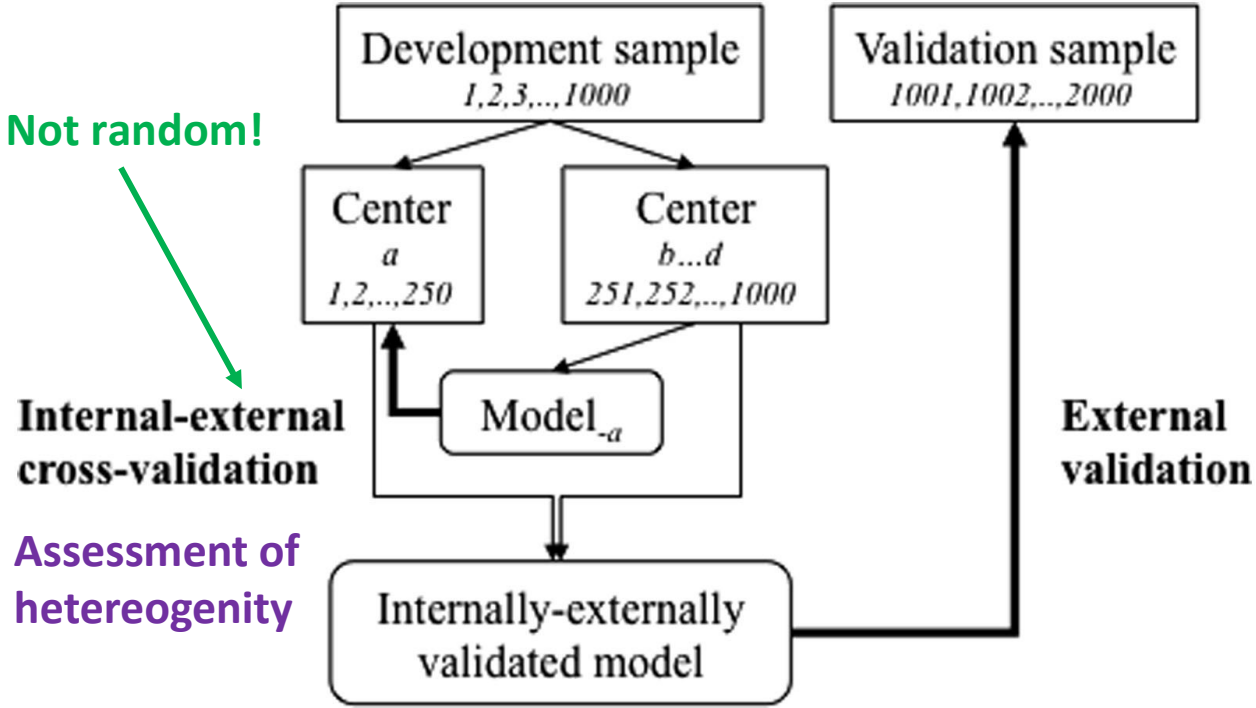
Important note: Machine learning approaches make us of data splitting methods (test and training) for internal optimization!

External validation

Good model ‘performance’ in one study population **does not imply** good model performance in **new patients**

New patients = Patients from different centres, countries or times (different, but similar!)

(Internal-)external validation

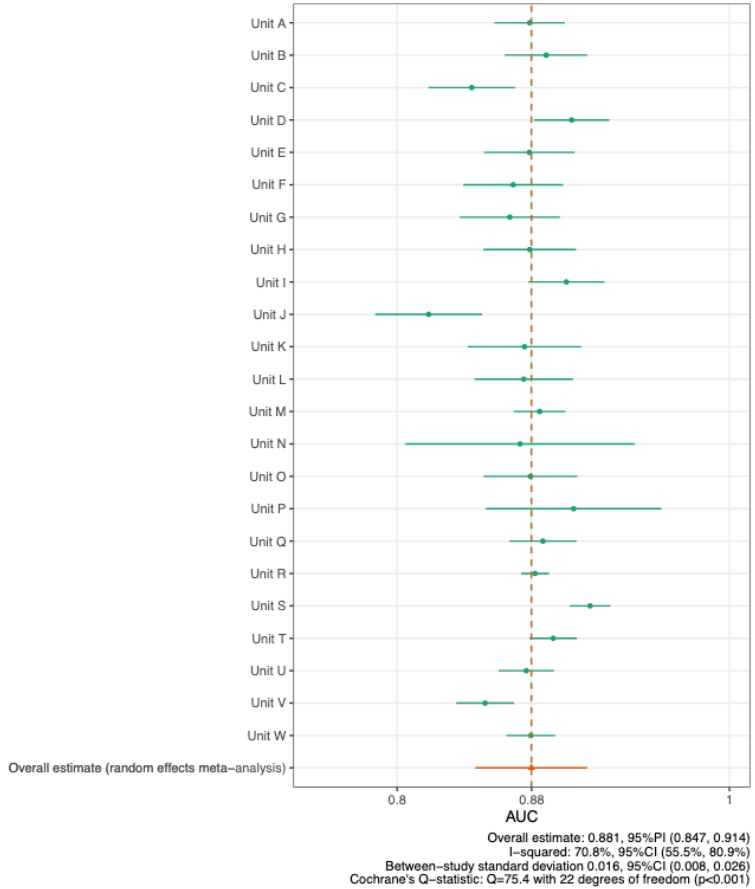


Test for generalizability

https://doi.org/10.1016/j.jclinepi.2015.04.005

(Internal-)external validation

Supplemental Figure 8: Geographic transportability: AUC random effects meta-analysis.



- We used a proposed framework for interpreting validation and transportability properties of prediction models and concluded that while our model showed good geographic discrimination transportability, other performance measures revealed a moderate to high heterogeneity between hospitals.

10.1016/j.jclinepi.2021.11.028

Reasons for poor validity

UNIVERSITÄT BERN

“There is no such that thing as a validated prediction model!”
https://doi.org/10.1186/s12916-023-02779-w

- **Modelling aspects**: Small samples, overfitting, optimism

- **Population** differences, changes of over time

- **Definition, coding and assessment** of predictors/outcomes differ across hospitals and time

Reasons for poor validity

"Rather, the current focus on developing new models **should shift to a focus on more extensive, well-conducted, and well-reported validation studies of promising models.**"

https://doi.org/10.1186/s12916-023-02779-w

Some references

Riley et al: **Calculating the sample size required for developing a clinical prediction model**. BMJ 2020;368:m441 doi: 10.1136/bmj.m441

Steyerberg and Harrell: **Prediction models need appropriate internal, internal–external, and external validation.** Journal of Clinical Epidemiology. 2016 Jan;69:245-247. doi: 10.1016/j.jclinepi.2015.04.005

Steyerberg: **Validation in prediction research: the waste by data splitting**. Journal of Clinical Epidemiology. 2018 Nov;103:131-133. doi: doi.org/10.1016/j.jclinepi.2018.07.010

Steyerberg et al.: **Assessing the performance of prediction models: a framework for traditional and novel measures**. Epidemiology. 2017 2010 Jan;21(1):128-38. doi: 10.1097/EDE.0b013e3181c30fb2

Steyerberg: **Clinical Prediction Models - A Practical Approach to Development, Validation, and Updating.** Springer Cham

Van Calster et al.: **A calibration hierarchy for risk models was defined: from utopia to empirical data.** J Clin Epidemiol. 2016 Jun;74:167-76. doi: 10.1016/j.jclinepi.2015.12.005

Van Calster et al.: **There is no such thing as a validated prediction model.** BMC Med 2023; Feb 24;21(1):70.

Van Smeden et al.: **Sample size for binary logistic prediction models: Beyond events per variable criteria.** Stat Methods Med Res. 2019 Aug;28(8):2455-2474. doi: 10.1177/0962280218784726