Better predictions using big(ger) datasets

Thomas Debray, PhD

Assistant Professor

Risk prediction

- Risk prediction = foreseeing / foretelling
    … (probability) of something that is yet unknown

- Turn available information (predictors) into a statement about the probability:
    … of having a particular disease -> **diagnosis**
    … of developing a particular event -> **prognosis**
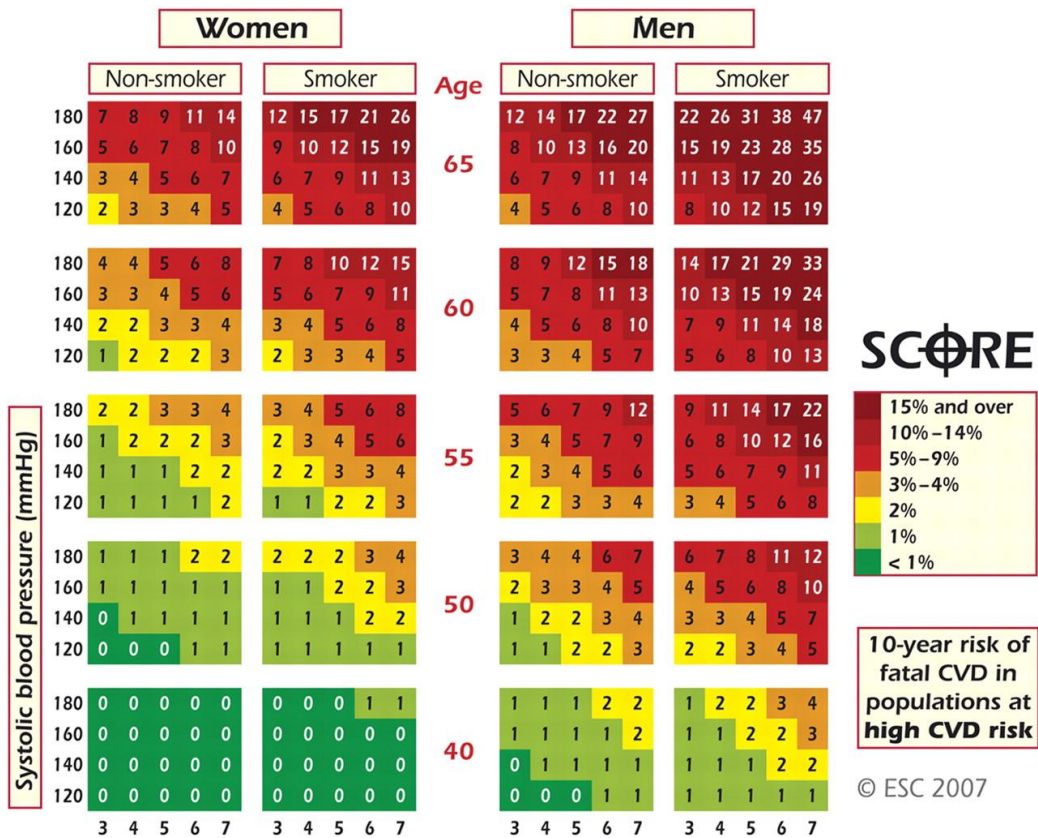
Why do we predict?

- Identification of high risk individuals
    - To inform patients and their families
    - To guide treatment decisions ("**precision medicine**")
    - To design randomized trials

- Data analysis
    - To deal with missing values
    - To match/subclassifiy patients
    - …

How do we predict?

- Combine information from multiple predictors
  - Subject characteristics (e.g. age, gender)
  - History and physical examination results (e.g. blood pressure)
  - Imaging results
  - (Bio)markers (e.g. coronary plaque)

- Develop a multivariable statistical model
  - Need for individual participant data (e.g. from cohort studies)
  - Many strategies available (e.g. logistic regression)

SCORE — 10-year risk of fatal CVD in populations at high CVD risk

© ESC 2007

@TPA_Debray

Dr. Watson

Watson for Oncology

"Bring personalized, evidence-supported cancer care plans to your patients"

- Interpret cancer patients' clinical information
- Digest doctor's notes, medical studies, and clinical guidelines
- Provide individualized treatment recommendations
- Adopted by more than 150 hospitals and healthcare organizations across 11 countries, including China

Hype meets reality

- Focus on US clinical practice and demographics
- Reliance on varies among hospitals
- Multiple examples of unsafe and incorrect treatment recommendations
- Lack of validation by independent scientists
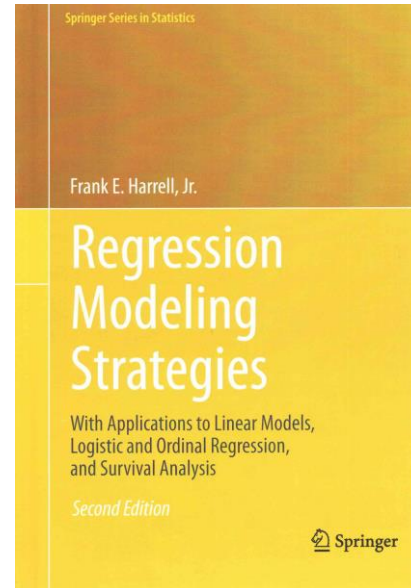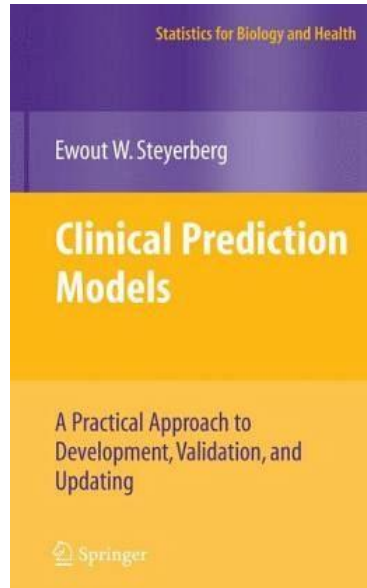- Lack of clinical trials to assess effectiveness

https://www.statnews.com/2017/09/05/watson-ibm-cancer/
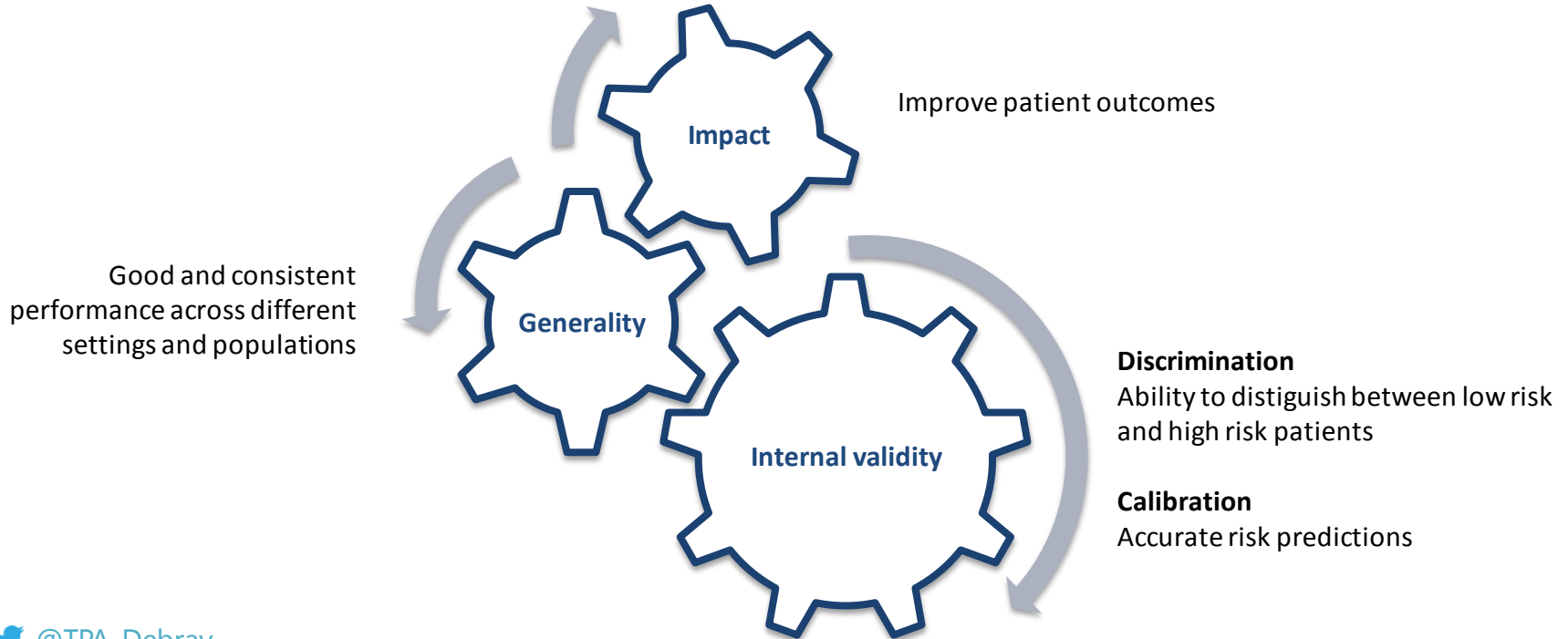
Most models are not as good as we think

"All models are wrong, but some are useful"

George Box

What is a "good" prediction model?

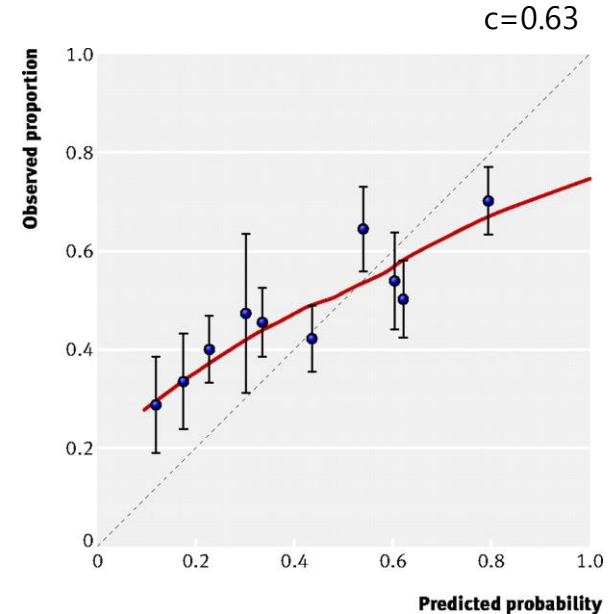# What is a "good" prediction model?

Impact — Improve patient outcomes

Generality — Good and consistent performance across different settings and populations

Internal validity

Discrimination
Ability to distiguish between low risk and high risk patients

Calibration
Accurate risk predictions

@TPA_Debray

What is a "good" prediction model?

Common measures of prediction model performance

- Discrimination
  - Concordance (c-) statistic
  - Area under the ROC curve

- Calibration
  - Calibration intercept (calibration-in-the-large)
  - Calibration slope
  - Ratio of expected versus observed events

c=0.63

10.1136/bmj.b605

Most models are not as good as we think

How to assess and improve the generalizability of prediction models?

The rise of "big" data sets

The rise of "big" data sets

Data increasingly available for thousands or even millions of patients from multiple practices, hospitals, or countries.

- Meta-analysis of individual participant data from multiple studies
  - Observational studies
  - Randomized controlled trials

- Analyses of databases and registry data containing e-health records

Examples of "big" data sets

International Prediction of Pre-eclampsia IPD Collaborative Network

- Target population
  - Pregnant women in the 1st or 2nd trimester of pregnancy

- So far, 81 datasets have been included
  - 15 UK studies
  - 66 international studies

Examples of "big" data sets

CALIBER

- EHR data encompassing more than 10 million adults with 400 million person-years of follow-up
- Primary care consultations and hospitalisations
- Clinical examination findings, blood laboratory results, prescriptions and vaccinations
- Diagnoses of diseases and mortality data

External validation of prediction models

- Studying (sources of) variation in model performance allows to assess the model's performance multiple times
  - In new patients from the same (target) population
  - In new patients from different (but related) populations

- Meta-analysis methods are needed to summarize prediction model performance and to investigate sources of heterogeneity.
  - Differences in case-mix
  - Invalid predictor effects
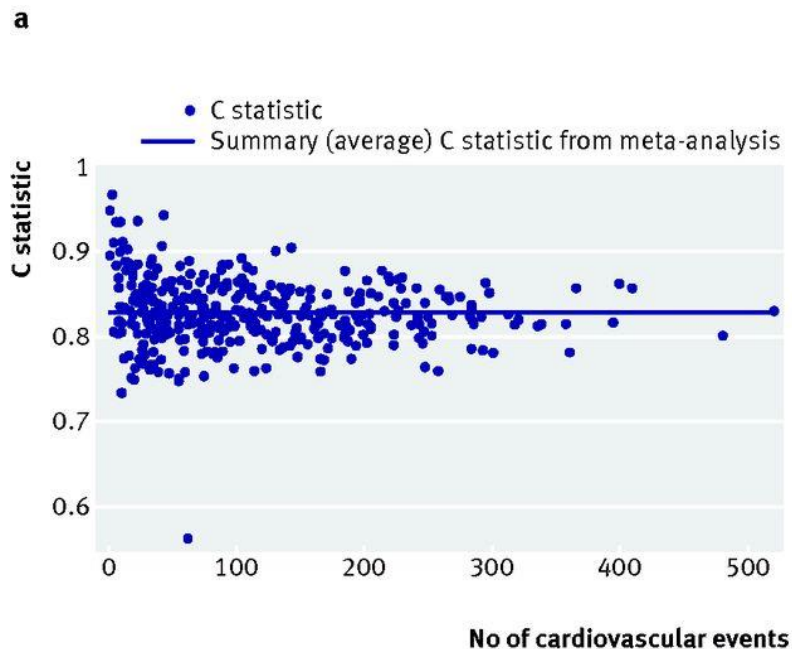
Validation of QRISK 2

Registry data with 1.58 million patients from 365 practices

Objective To evaluate the performance of the QRISK2 score for predicting 10-year cardiovascular disease in an independent UK cohort of patients from general practice records and to compare it with the NICE version of the Framingham equation and QRISK1.

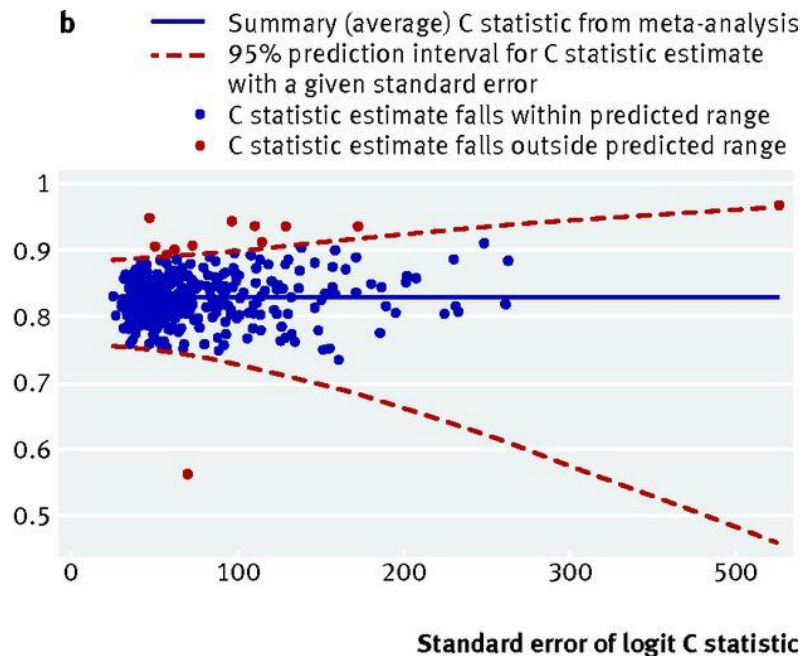**Design** Prospective cohort study to validate a cardiovascular risk score.

Setting 365 practices from United Kingdom contributing to The Health Improvement Network (THIN) database.

Participants 1.58 million patients registered with a general practice between 1 January 1993 and 20 June 2008, aged 35-74 years (9.4 million person years) with 71 465 cardiovascular events.
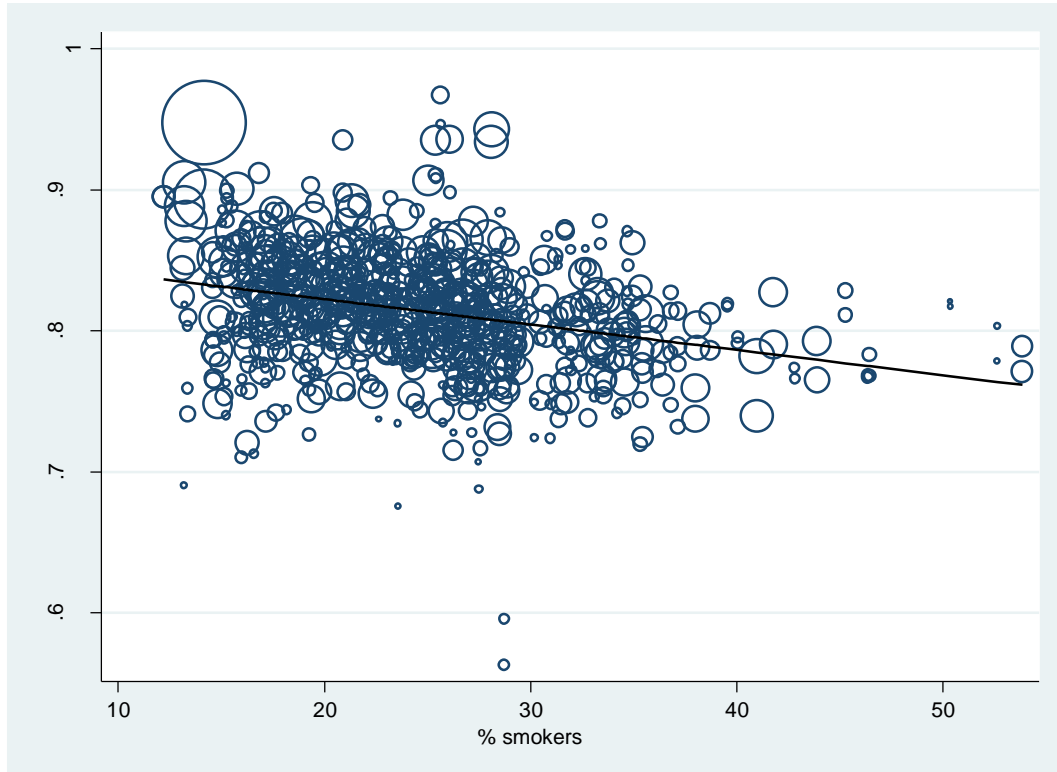
10.1136/bmj.c2442

a

- C statistic
— Summary (average) C statistic from meta-analysis

C statistic (y-axis): 1, 0.9, 0.8, 0.7, 0.6
No of cardiovascular events (x-axis): 0, 100, 200, 300, 400, 500

Summary C statistic = 0.83 (95% CI 0.826 to 0.833)
95% prediction interval for true C statistic in a
new practice = 0.76 to 0.88

b

— Summary (average) C statistic from meta-analysis
- - - 95% prediction interval for C statistic estimate
  with a given standard error
- C statistic estimate falls within predicted range
- C statistic estimate falls outside predicted range

(y-axis): 1, 0.9, 0.8, 0.7, 0.6, 0.5
Standard error of logit C statistic (x-axis): 0, 100, 200, 300, 500

Summary (average) C statistic = 0.83 (95% CI 0.826 to 0.833)
95% prediction interval for true C statistic in a
new practice = 0.76 to 0.88

External validation of prediction models

Investigating heterogeneity in discrimination performance

External validation of prediction models

Final remarks

- Heterogeneity in discrimination performance should be anticipated
- Good and consistent calibration is what matters most
- Multivariate meta-analysis methods can be used to determine probability of "adequate" performance in new settings, e.g.
  - 0.8 < ratio observed vs. expected events < 1.2, *and*
  - 0.8 < Calibration slope < 1.2, *and*
  - C-statistic > 0.7

Key references

- Debray et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. Stat Methods Med Res 2018
- Debray et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ 2017
- Riley et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ 2016
- Snell et al. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model. J Clin Epidemiol 2015
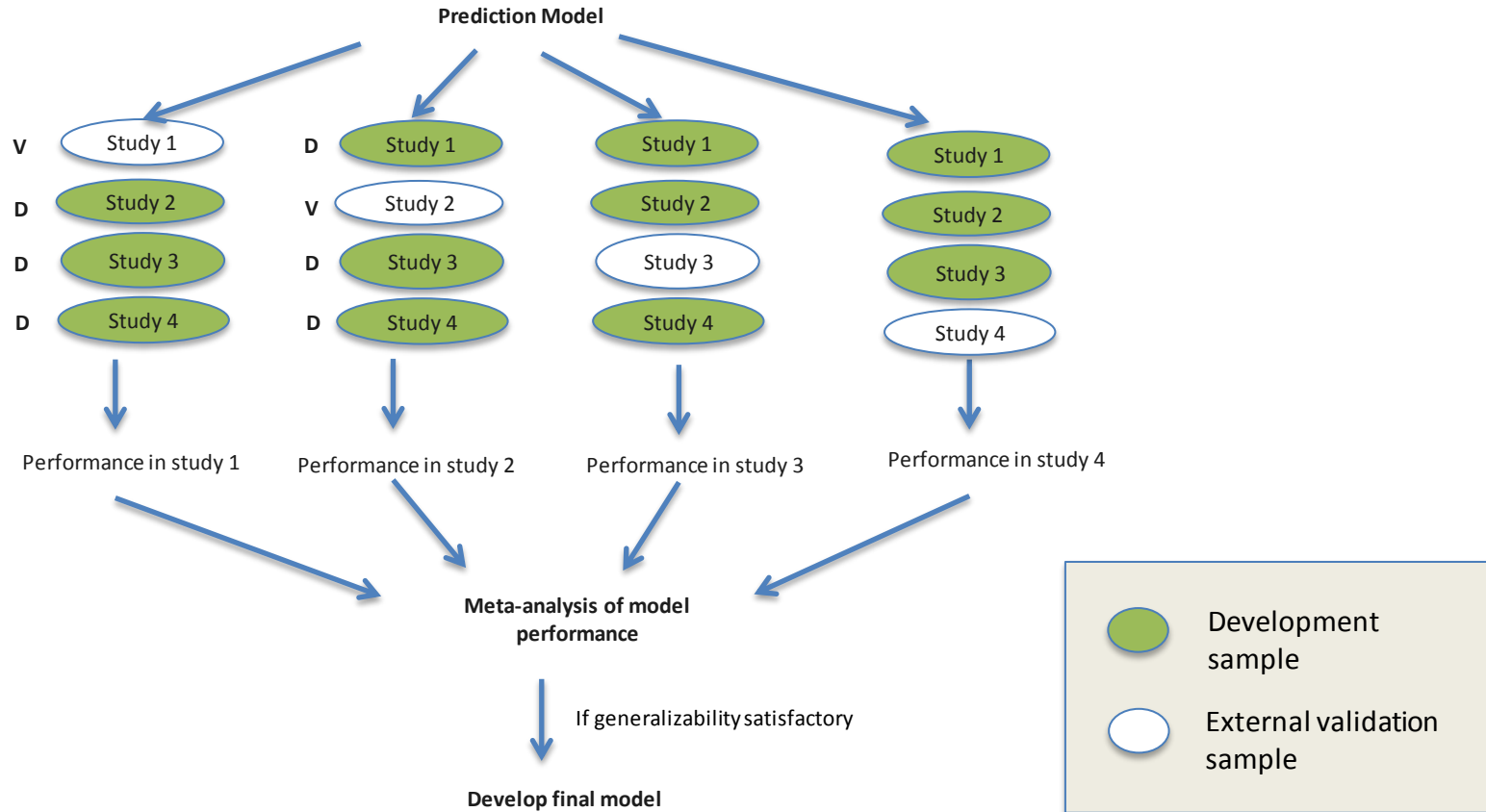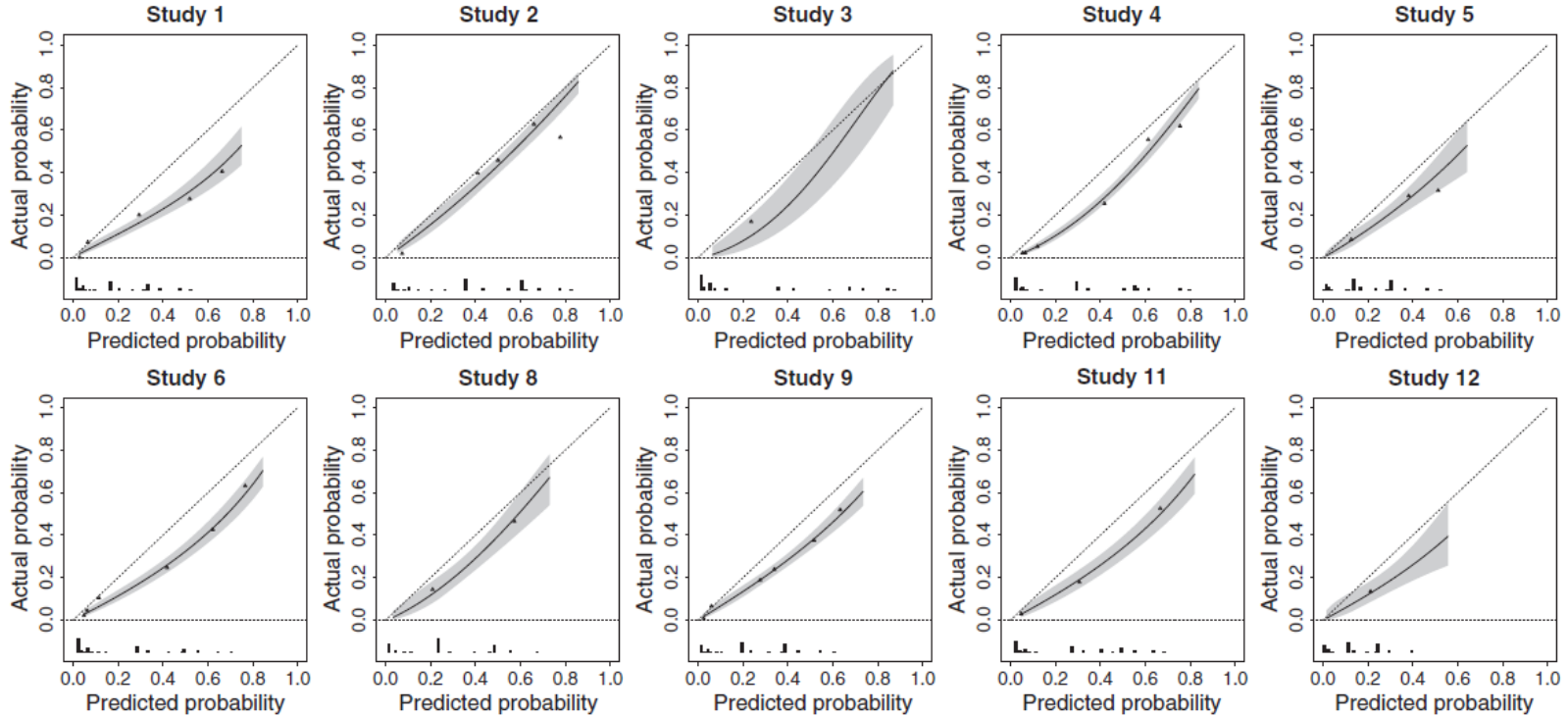
Development of prediction models

- Access to big(ger) datasets enables to assess model transportability (rather than merely reproduciblity) during its development

- Identify and account for heterogeneity in prediction model performance via <u>internal-external cross-validation</u>

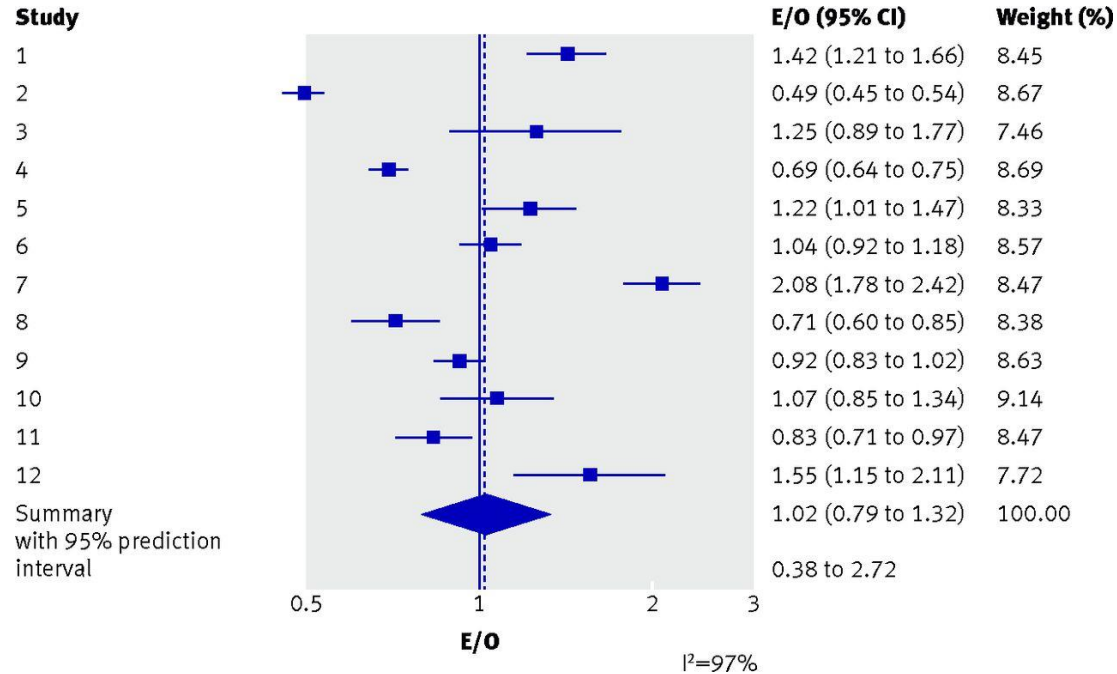- This allows to optimize prediction model generalizability

Internal-external cross-validation (IECV)

@TPA_Debray

IECV allows for many external validations

10.1002/sim.5732

IECV allows for meta-analysis

10.1136/bmj.i3140
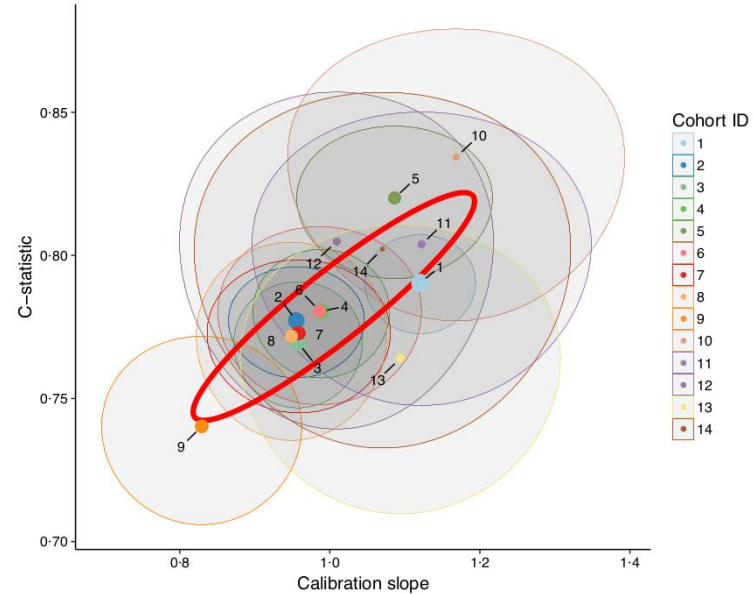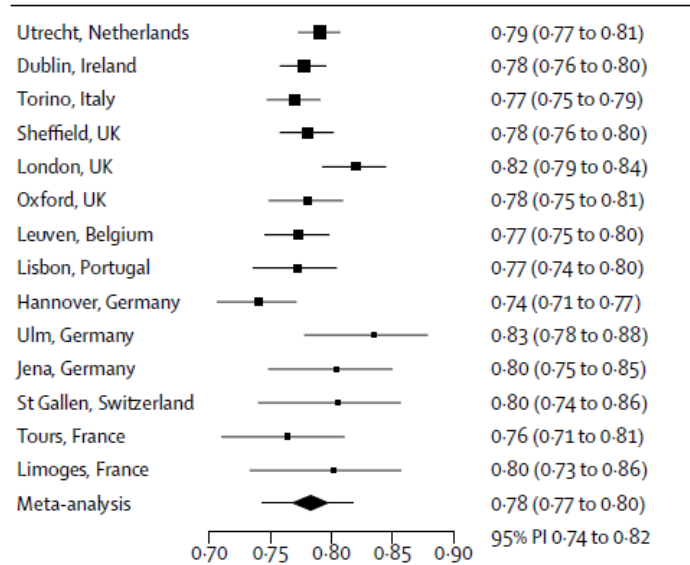
Development and validation of ENCALS

Prognosis for patients with amyotrophic lateral sclerosis (ALS)

- Cohort data from 11,475 patients from 14 European ALS centres
- Composite survival outcome (non-invasive ventilation for more than 23 h per day, tracheostomy, or death)
- Development of multivariable Royston-Parmar models
- Assessment of generalizability via IECV

THE LANCET
Neurology

Development and validation of ENCALS

10.1016/S1474-4422(18)30089-9

Development and validation of ENCALS

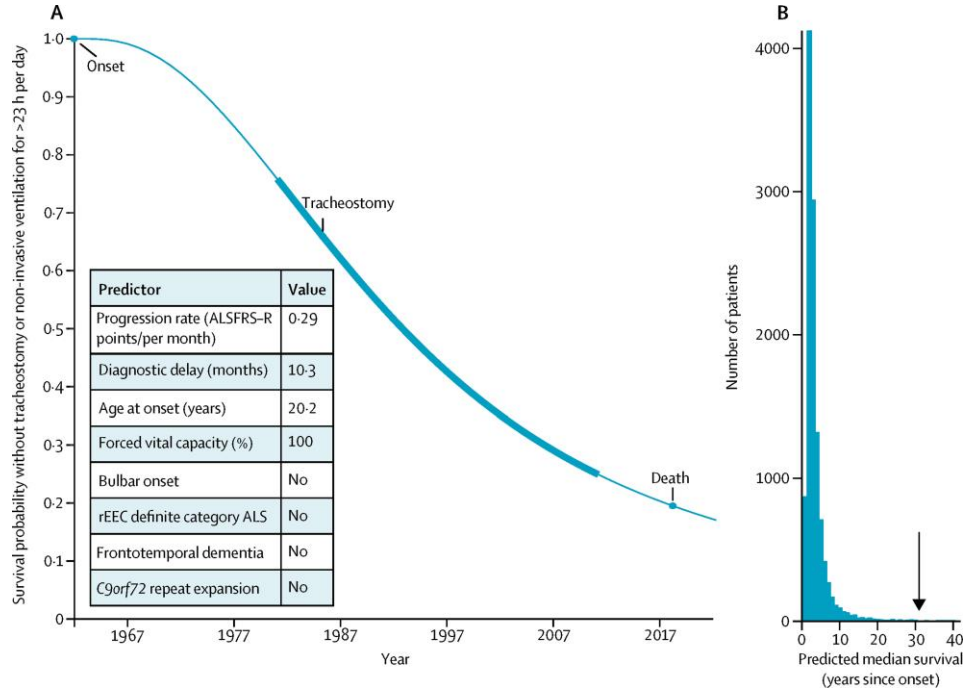| Measure	Criteria	Prob. of "good" performance	Joint probability
C-statistic	> 0.70	100%	98.3%
Calibration slope	0.80 to 1.20	97.1%	
Calibration-in-the-large	-0.587 to 0.587	85.5%	

The life expectancy of Stephen Hawking

"Using publicly available data, we examined whether Professor Hawking's survival was as rare as his intellectual performance, or could be predicted solely based on his disease characteristics at diagnosis in 1963."

- Predicted 10-year survival probability: 94%
- The IQR for his predicted survival lay between 1981 and 2011
- Young age of onset was the most important factor for his long survival

The life expectancy of Stephen Hawking

10.1016/S1474-4422(18)30241-2

Key references

- Ahmed et al. Developing and validating risk prediction models in an individual participant data meta-analysis. BMC Med Res Methodol 2014
- Debray et al. A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. Stat Med 2013
- Debray et al. Individual Participant Data (IPD) Meta- analyses of Diagnostic and Prognostic Modeling Studies: Guidance on Their Use. PLoS Med 2015

Future developments

Software

`metamisc: Diagnostic and Prognostic Meta-Analysis`

Meta-analysis of diagnostic and prognostic modeling studies. Summarize estimates of prognostic factors, diagnostic test accuracy and prediction model performance. Validate, update and combine published prediction models. Develop new prediction models with data from multiple studies.

Version:	0.1.9
Depends:	R ($\geq$ 3.2.0), stats, graphics
Imports:	metafor ($\geq$ 2.0.0), mvtnorm, ellipse, lme4, plyr, ggplot2
Suggests:	runjags, rjags, testthat ($\geq$ 1.0.2)
Published:	2018-05-13
Author:	Thomas Debray ⓘ [aut, cre], Valentijn de Jong [aut]
Maintainer:	Thomas Debray <thomas.debray at gmail.com>
License:	GPL-3
URL:	http://r-forge.r-project.org/projects/metamisc/
NeedsCompilation:	no
In views:	MetaAnalysis
CRAN checks:	metamisc results

`Downloads:`

Reference manual:	metamisc.pdf
Package source:	metamisc_0.1.9.tar.gz
Windows binaries:	r-devel: metamisc_0.1.9.zip, r-release: metamisc_0.1.9.zip, r-oldrel: metamisc_0.1.9.zip
OS X binaries:	r-release: metamisc_0.1.9.tgz, r-oldrel: metamisc_0.1.8.tgz
Old sources:	metamisc archive

`Linking:`

Please use the canonical form `https://CRAN.R-project.org/package=metamisc` to link to this page.

Guidance and methods

- **Prognostic Research in Health Care: concepts, methods and impact**
  editors: Richard Riley, Danielle Van der Windt, Peter Croft, Karel Moons

- **Evidence synthesis using individual participant data: Concepts, Methods and Guidance for Clinical Research**
  editors: Richard Riley, Jayne Tierney, Lesley Stewart

- **Handbook of Meta-analysis**
  editors: Christopher Schmid, Theo Stijnen, Ian White