UMC Utrecht

Better predictions using big(ger) datasets

Thomas Debray, PhD
Assistant Professor

UMC Utrecht

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?

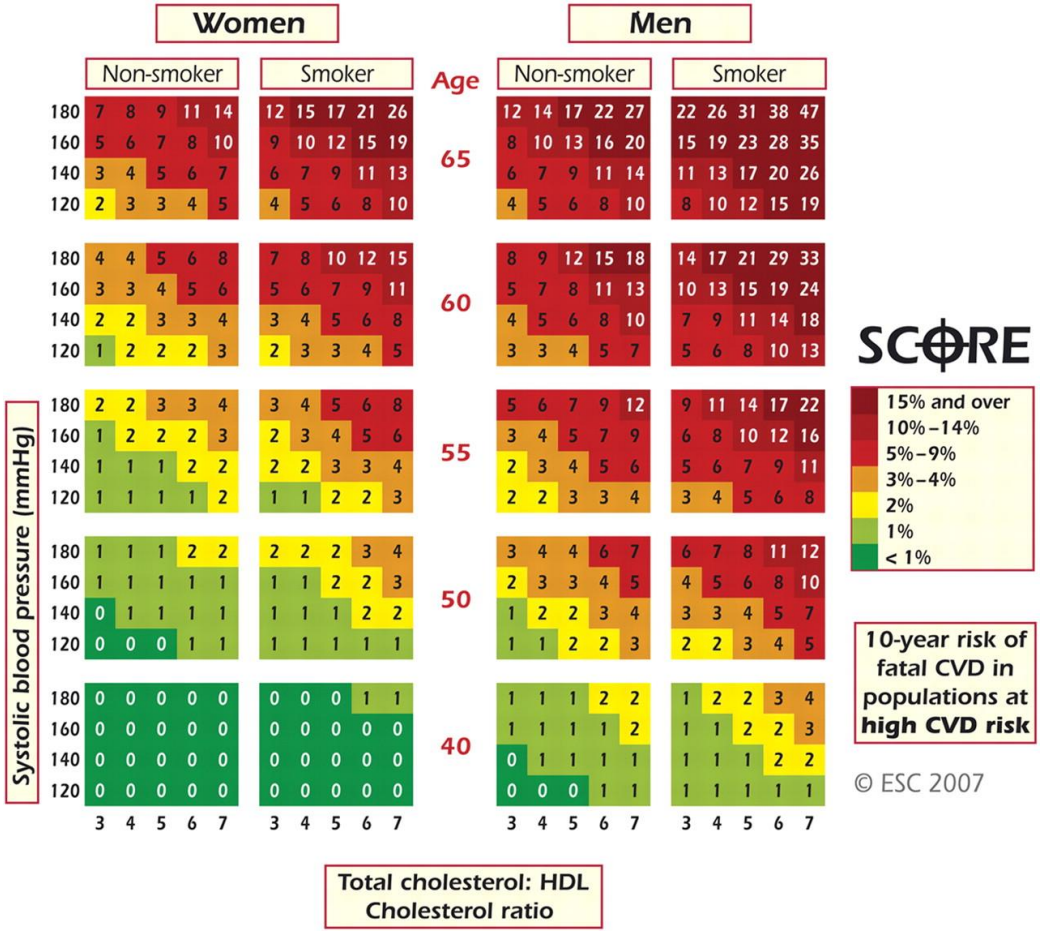
- Identification of high risk individuals
  - To inform patients and their families
  - To guide treatment decisions ("**precision medicine**")
  - To design randomized trials

- 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)


Women
Men
Non-smoker
Smoker
Non-smoker
Smoker
Age
Systolic blood pressure (mmHg)
180 7 8 9 11 14 12 15 17 21 26 12 14 17 22 27 22 26 31 38 47
160 5 6 7 8 10 9 10 12 15 19 8 10 13 16 20 15 19 23 28 35
140 3 4 5 6 7 6 7 9 11 13 6 7 9 11 14 11 13 17 20 26
120 2 3 3 4 5 4 5 6 8 10 4 5 6 8 10 8 10 12 15 19
65
180 4 4 5 6 8 7 8 10 12 15 8 9 12 15 18 14 17 21 29 33
160 3 3 4 5 6 5 6 7 9 11 5 7 8 11 13 10 13 15 19 24
140 2 2 3 3 4 3 4 5 6 8 4 5 6 8 10 7 9 11 14 18
120 1 2 2 2 3 2 3 3 4 5 3 3 4 5 7 5 6 8 10 13
60
180 2 2 3 3 4 3 4 5 6 8 5 6 7 9 12 9 11 14 17 22
160 1 2 2 2 3 2 3 4 5 6 3 4 5 7 9 6 8 10 12 16
140 1 1 1 2 2 2 2 3 3 4 2 3 4 5 6 5 6 7 9 11
120 1 1 1 1 2 1 1 2 2 3 2 2 3 3 4 3 4 5 6 8
55
180 1 1 1 2 2 2 2 2 3 4 3 4 4 6 7 6 7 8 11 12
160 1 1 1 1 1 1 1 2 2 3 2 3 3 4 5 4 5 6 8 10
140 0 1 1 1 1 1 1 1 2 2 1 2 2 3 4 3 3 4 5 7
120 0 0 0 1 1 1 1 1 1 1 1 1 2 2 3 2 2 3 4 5
50
180 0 0 0 0 0 0 0 0 1 1 1 1 2 2 1 2 2 3 4
160 0 0 0 0 0 0 0 0 0 0 1 1 1 1 2 1 1 2 2 3
140 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 2 2
120 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1
40
3 4 5 6 7 3 4 5 6 7 3 4 5 6 7 3 4 5 6 7
Total cholesterol: HDL Cholesterol ratio
SCORE
15% and over
10%–14%
5%–9%
3%–4%
2%
1%
< 1%
10-year risk of fatal CVD in populations at high CVD risk
© ESC 2007

Dr. Watson


PIENSE
THINK
SMAOMIŞ
सोचिए
ΣΚΕΨΟΥ
DENKE
PENSER
$300,000
Who is Stoker?
(I FOR ONE WELCOME OUR NEW COMPUTER OVERLORDS)
$1,000
$1,000,000
Who is Bram Stoker?
$ 17,973
$200,000
WHO IS BRAM STOKER?
$5600

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

https://www.ibm.com/watson/health/oncology-and-genomics/oncology/

Hype meets reality

- Reliance on Watson for Oncology varies among hospitals
- Focus on US clinical practice and demographics
- Examples of poor advise on how to treat patients' cancers
  - "multiple examples of unsafe and incorrect treatment recommendations"
  - "serious questions about the process for building content and the underlying technology."
- Lack of validation by independent scientists
- Lack of clinical trials to assess effectiveness

https://www.statnews.com/2017/09/05/watson-ibm-cancer/

What is a “good” prediction model?

Accurate predictions

Ability to distinguish between low and high risk patients

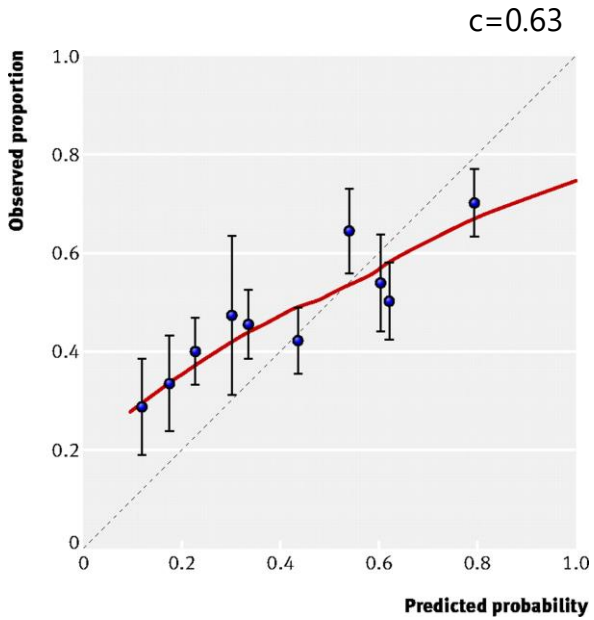
Good and consistent performance across different settings and populations

Influence decision making

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

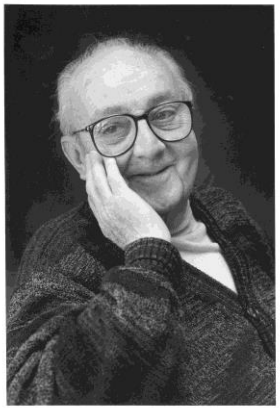


10.1136/bmj.b605

Key problems

- Poor conduct and reporting
  - Poor adherence to TRIPOD reporting guidelines (Heus et al. BMC Med 2018)
- Lack of reproducibility
  - Invalid model predictions in new patients from the same population
  - Overfitting
- Lack of transportability
  - Invalid model predictions in new patients from different but related populations

Most models are not as good as we think



“All models are wrong, but some are useful”

George Box

Most models are not as good as we think

Prediction model performance often varies across settings, populations and time periods

- Invalid predictor effects
  (e.g. missed predictors, non-linear terms or interactions)
- Discrepancies in outcome and predictor assessement
  (e.g. differences in measurement error)
- Case-mix variation
  (e.g. differences is patient characteristics)

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

UMC Utrecht

External validation of prediction models

- Studying (sources of) variation in model performance allows to assess the model's potential generalizability and clinical usefulness

- Meta-analysis methods are needed to combine model performance estimates across studies/clusters, and to investigate sources of heterogeneity.

External validation of prediction models

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

Validation of QRISK 2

Registry data with 1.58 million patients from 365 practices

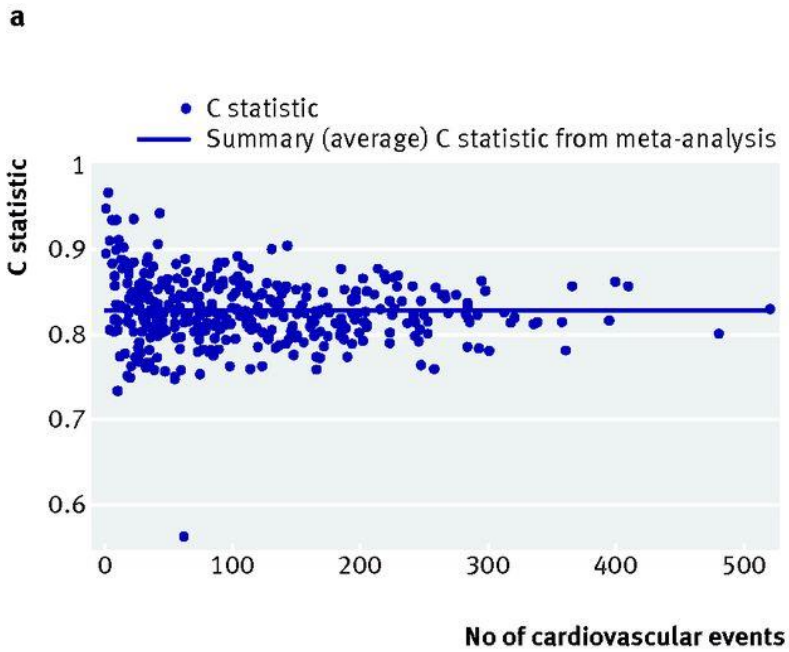
Objective To evaluate the performance of the QRISK2 score for predicting 10-year cardiovascular disease in an independent UK cohort of patients from general practice records and to compare it with the NICE version of the Framingham equation and QRISK1.

Design Prospective cohort study to validate a cardiovascular risk score.

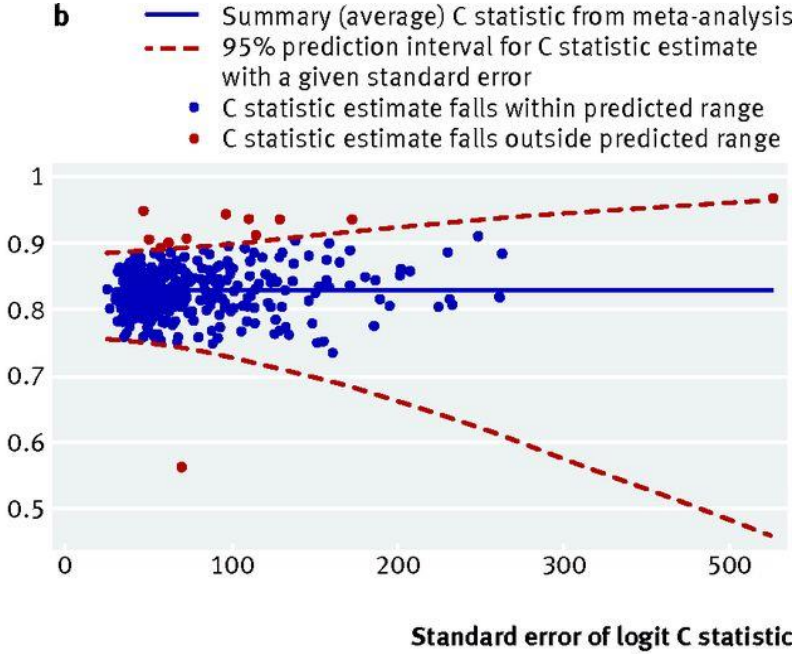
Setting 365 practices from United Kingdom contributing to The Health Improvement Network (THIN) database.

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



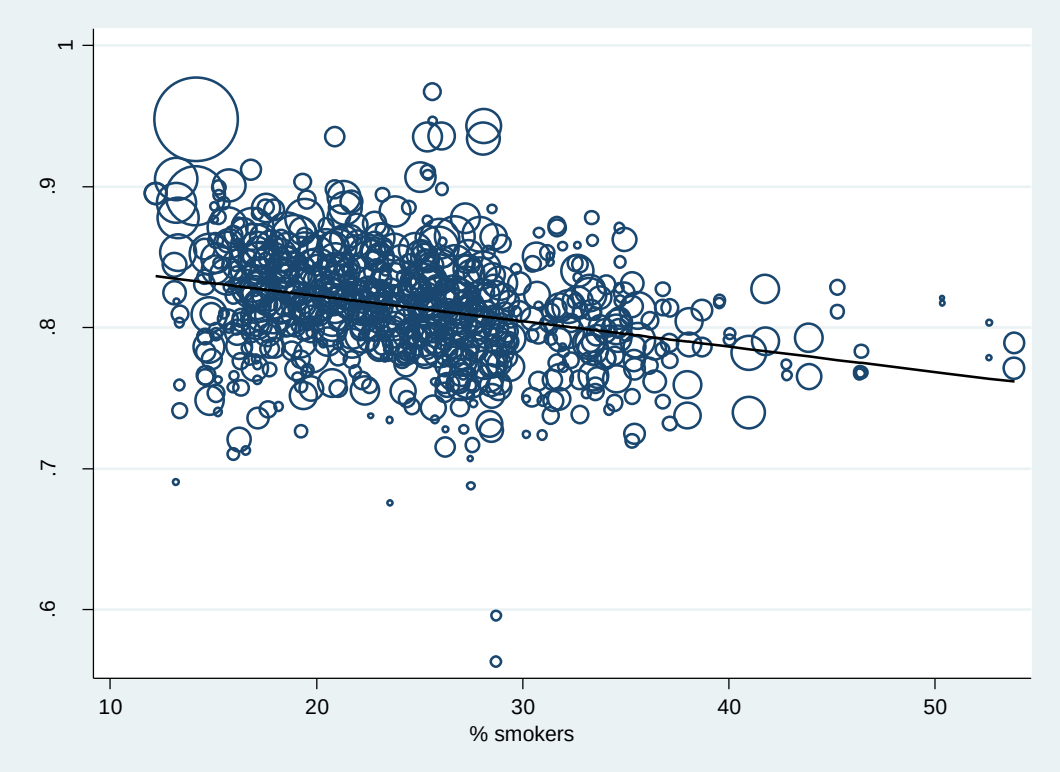
Summary C statistic = 0.83 (95% CI 0.826 to 0.833)
95% prediction interval for true C statistic in a new practice = 0.76 to 0.88



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

10.1136/bmj.c2442

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

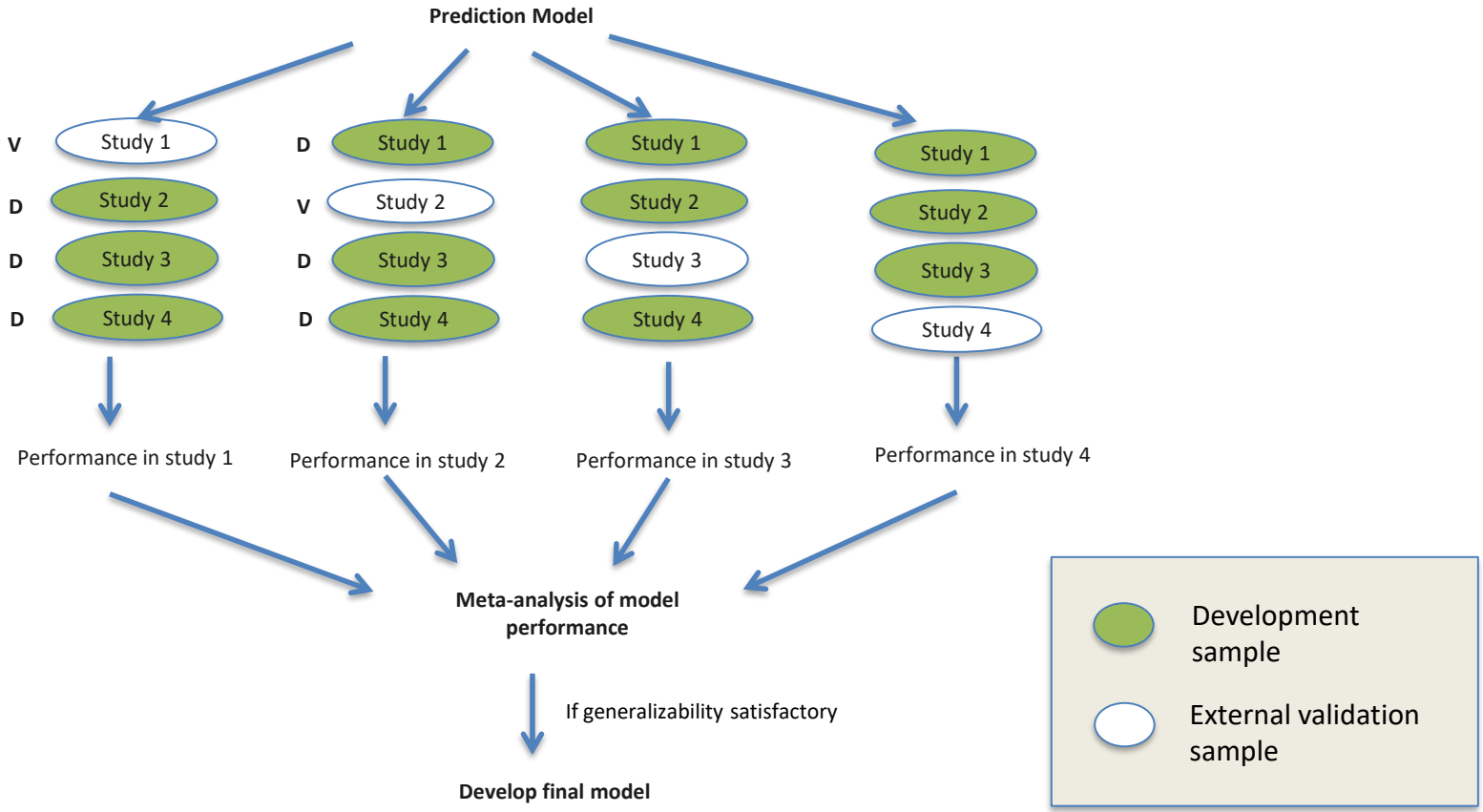
UMC Utrecht

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 internal-external cross-validation
- This allows to optimize prediction model generalizability

10.1002/sim.5732

Internal-external cross-validation



Development of prediction models

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

Development and validation of ENCALS

Prognosis for patients with amyotrophic lateral sclerosis (ALS)

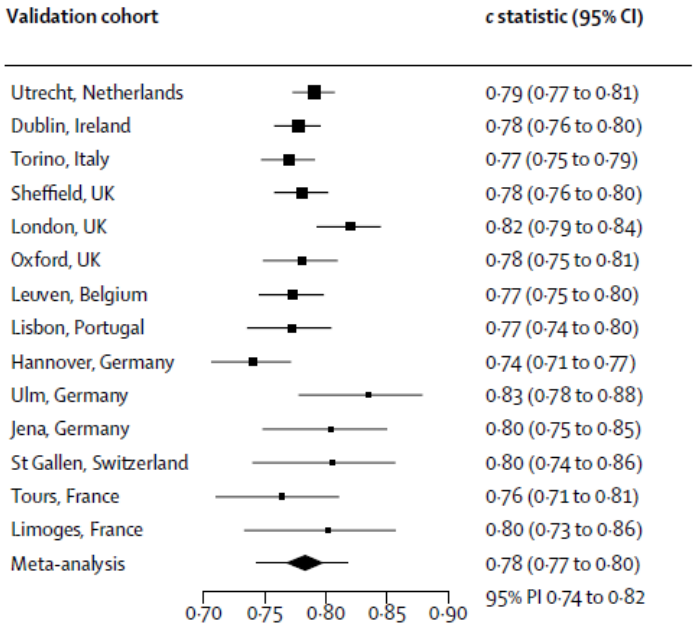
- Cohort data from 11,475 patients from 14 European ALS centres
- Composite survival outcome (non-invasive ventilation for more than 23 h per day, tracheostomy, or death)
- Development of multivariable Royston-Parmar models
- Assessment of generalizability via IECV

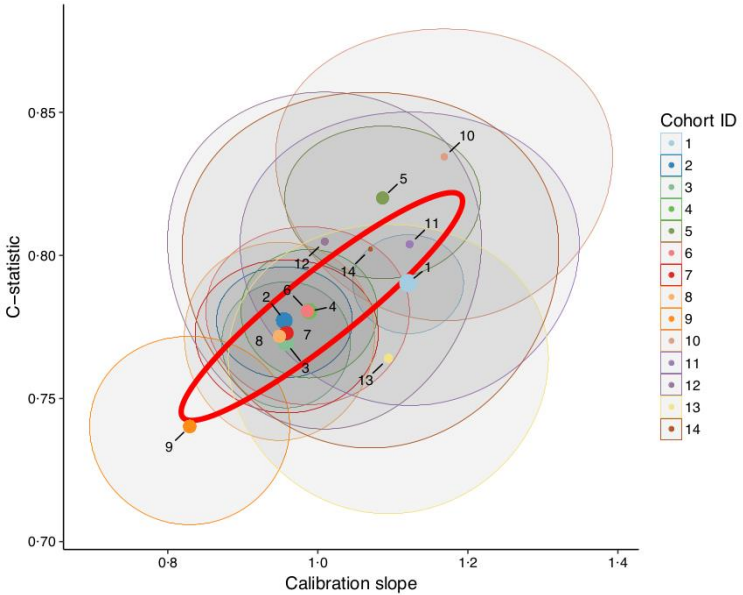
THE LANCET Neurology

@TPA_Debray

10.1016/S1474-4422(18)30089-9

Development and validation of ENCALS





@TPA_Debray

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%	

10.1016/S1474-4422(18)30089-9

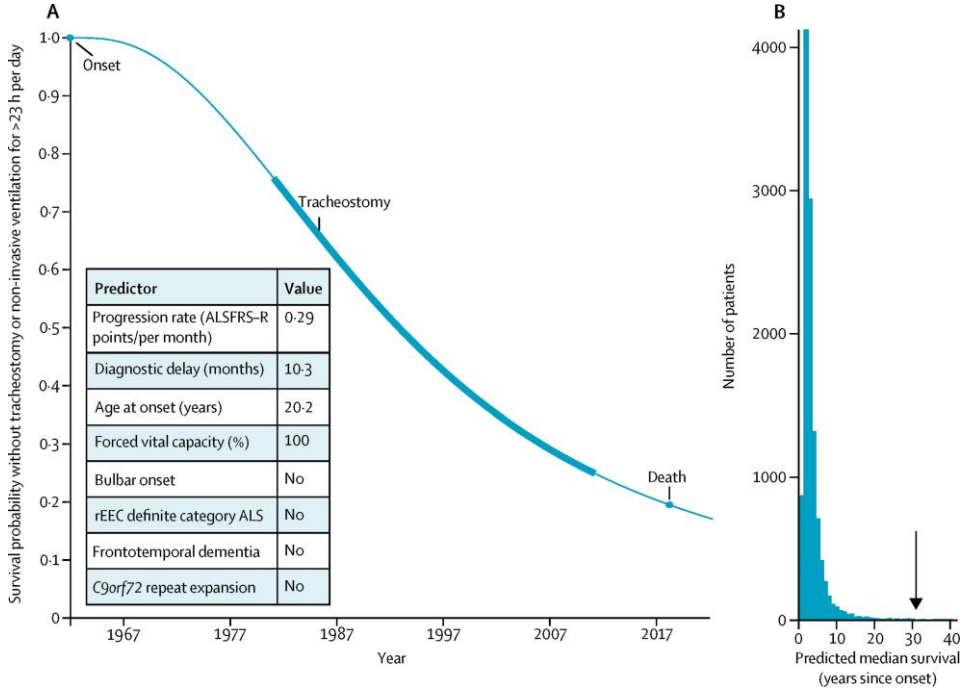
The life expectancy of Stephen Hawking

"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

10.1016/S1474-4422(18)30241-2

The life expectancy of Stephen Hawking



10.1016/S1474-4422(18)30241-2

Future developments

UMC Utrecht

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 (≥ 3.2.0), stats, graphics
Imports:	metafor (≥ 2.0.0), mvtnorm, ellipse, lme4, plyr, ggplot2
Suggests:	runjags, rjags, testthat (≥ 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

- **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