UMC Utrecht
Julius Center

Collaborative research in risk prediction

Thomas Debray, PhD

Julius Center for Health Sciences and Primary Care,
University Medical Center Utrecht, The Netherlands

Cochrane Netherlands, Utrecht, The Netherlands

thomasdebray.be

Aims of my talk

- To describe the value of clinical prediction models
- To foster interdisciplinary collaboration
- To highlight the merits of data sharing
- To inspire you

Who am I?

Current position

- Assistant Professor, Julius Center for Health Sciences and Primary Care
- Affiliated Researcher, Cochrane Netherlands
- Honorary Senior Research Associate, University College London
- Honorary Departmental Senior Research Fellow, University of Oxford

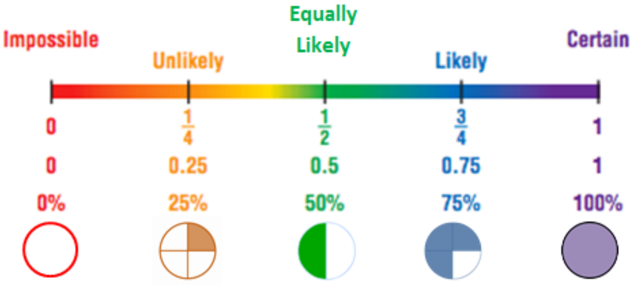
Background in

- Computer Science (BSc)
- Artificial Intelligence (MSc)
- Epidemiology (MSc, PhD)

What is risk prediction?

Estimating the probability of something that is yet unknown

- Presence of a certain disease → diagnosis
- Future occurrence of a particular event → prognosis



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 (e.g. computed tomography scan)
- (Bio)markers (e.g. coronary plaque)

Why do we predict?

- To inform patients and their families
- To decide upon further testing (e.g. magnetic resonance imaging)
- To decide upon patient referral (e.g. to secondary care)
- To decide on preemptive or therapeutic strategies
- To guide treatment decisions

Age	Systolic blood pressure (mmHg)	Women Non-smoker 3	4	5	6	7	Women Smoker 3	4	5	6	7	Men Non-smoker 3	4	5	6	7	Men Smoker 3	4	5	6	7
65	180	4	5	6	7	8	8	9	11	13	15	7	8	10	12	15	13	16	19	23	28
	160	3	3	4	5	6	6	6	8	9	11	5	6	7	9	11	9	11	13	16	20
	140	2	2	3	3	4	4	5	5	6	8	4	4	5	6	8	7	8	10	12	14
	120	1	2	2	2	3	3	3	4	5	6	2	3	4	4	6	5	6	7	8	10
60	180	2	3	3	4	5	4	5	6	7	9	5	5	7	8	10	9	10	12	15	19
	160	2	2	2	3	3	3	4	4	5	6	3	4	5	6	7	6	7	9	11	13
	140	1	1	2	2	2	2	2	3	4	4	2	3	3	4	5	4	5	6	8	10
	120	1	1	1	1	2	1	2	2	3	3	2	2	2	3	4	3	4	4	5	7
55	180	1	1	2	2	2	2	3	3	4	5	3	3	4	5	6	5	6	8	10	12
	160	1	1	1	1	2	2	2	2	3	3	2	2	3	4	5	4	5	6	7	9
	140	1	1	1	1	1	1	1	2	2	2	1	2	2	3	3	3	3	4	5	6
	120	0	0	1	1	1	1	1	1	1	2	1	1	1	2	2	2	2	3	3	4
50	180	1	1	1	1	1	1	1	1	2	2	2	2	2	3	4	3	4	5	6	7
	160	0	0	1	1	1	1	1	1	1	2	1	1	2	2	3	2	3	3	4	5
	140	0	0	0	0	1	0	1	1	1	1	1	1	1	2	2	2	2	2	3	4
	120	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	2	2	3
40	180	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	2	2
	160	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
	140	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1
	120	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

Total cholesterol: HDL Cholesterol ratio

SCORE

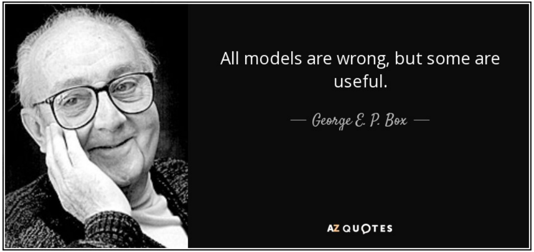
- 15% and over
- 10%–14%
- 5%–9%
- 3%–4%
- 2%
- 1%
- < 1%

10-year risk of fatal CVD in populations at low CVD risk

Risk prediction - the reality

Many prediction models perform more poorly than anticipated, do not affect clinical practice, or are implemented for the wrong reasons

- Small & poor quality studies
- Limited variation in studied patients, settings or populations
- Lack of validity and effectiveness assessments



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.

- Analyses of databases and registry data containing e-health records
- Meta-analysis of individual participant data from multiple studies

A common theme in these data is the presence of clustering, which allows to study generalizability of model predictions across different settings and populations

Exemplar projects

A selection of ongoing collaborative projects using “big” data sets

- Identification of patients in need of specialized trauma care

- Prognosis for patients with amyotrophic lateral sclerosis

- Personalized medicine for infectious diseases

Identification of patients in need of specialized trauma care

Prehospital trauma triage is essential to get the right patient to the right hospital

- Need for prehospital triage tool(s)
- Many existing tools have poor discrimination
- Most studies are based on small and local data sets
- Limited time and equipment to collect patient data on-scene

Identification of patients in need of specialized trauma care

The Trauma Triage App (TTApp)

- Mobile application
- Logistic regression model to estimate the need of specialized trauma care
- Based on data from a single Emergency Medical Service (EMS)

Improvements are underway

- Combine data from multiple EMS regions in the Netherlands
- Adopt machine learning methods (XGBoost)
- Investigate variation in prediction performance across NL regions

Prognosis for patients with amyotrophic lateral sclerosis (ALS)

ALS

- Neurodegenerative disease
- No cure or effective treatment
- Heterogeneous survival (several months to $>$ 10 years)

Prof. Stephen W Hawking

- Diagnosed with ALS in 1963
- Life expectancy upon diagnosis: 2 years
- Died on March 14, 2018 (aged 76)

Need for accurate tools that can predict survival in patients diagnosed with ALS

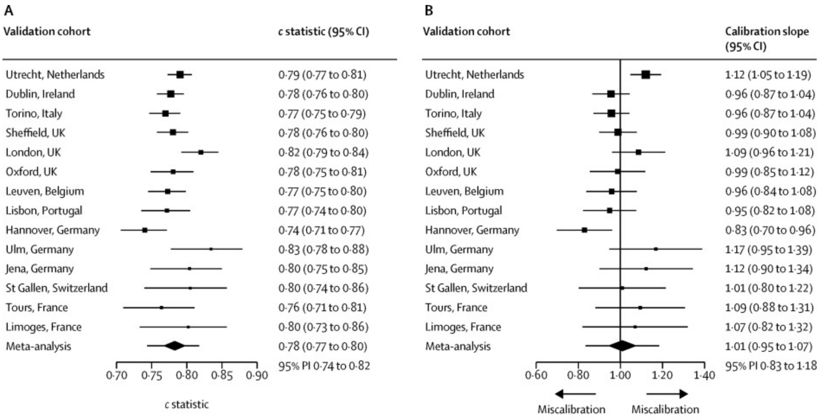
Prognosis for patients with amyotrophic lateral sclerosis (ALS)

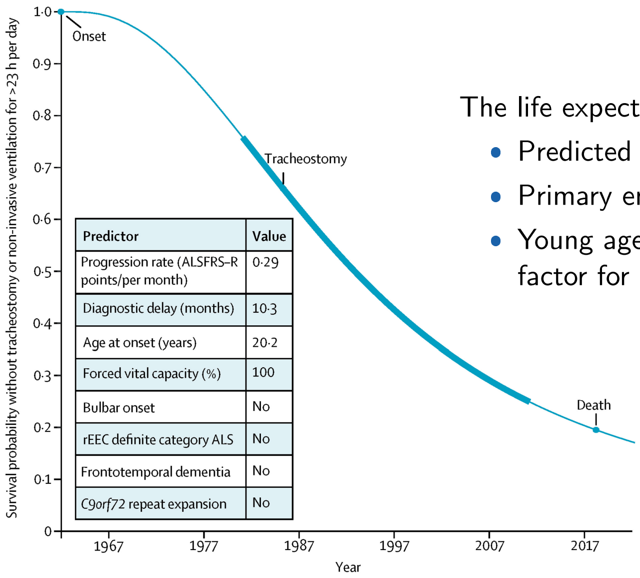
We set up an international collaboration between neurologists, epidemiologists and statisticians

- To combine cohort data
  - from 11,475 patients
  - from 14 ALS centres across Europe
- To implement recently developed statistical methodology
  - for dealing with missing values in large, clustered, data sets
  - for assessing prediction model performance across different ALS centres
- To publish a prediction model that is freely accessible by medical doctors
  - to provide estimates of prognosis in individual patients with ALS
  - to facilitate clinical practice and innovative trial design

Prognosis for patients with amyotrophic lateral sclerosis (ALS)

We externally validated the prediction model in each of the 14 ALS centres (probability of "good" performance in new centres: 98%)





Predictor	Value
Progression rate (ALSFRS–R points/per month)	0·29
Diagnostic delay (months)	10·3
Age at onset (years)	20·2
Forced vital capacity (%)	100
Bulbar onset	No
rEEC definite category ALS	No
Frontotemporal dementia	No
C9orf72 repeat expansion	No

The life expectancy of Stephen Hawking

- Predicted 10-year survival probability: 94%
- Primary endpoint reached after 22 years
- Young age of onset was the most important factor for his long survival

Prognosis for patients with amyotrophic lateral sclerosis (ALS)

Articles

Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model

Henk-Jan Westeneng, Thomas P A Debray, Anne E Visser, Ruben P A van Eijk, James P K Rooney, Andrea Calvo, Sarah Martin, Christopher J McDermott, Alexander G Thompson, Susana Pinto, Xenia Kobeleva, Angela Rosenbohm, Beatrice Stubendorff, Helma Sommer, Bas M Middelkoop, Annelot M Dekker, Joke J F A van Vugt, Wouter van Rheenen, Alice Vajda, Mark Heverin, Mbombe Kazoka, Hannah Hollinger, Marta Gromicho, Sonja Körner, Thomas M Ringer, Annekathrin Rödiger, Anne Gunkel, Christopher E Shaw, Annelien L Bredenoord, Michael A van Es, Philippe Corcia, Philippe Couratier, Markus Weber, Julian Grosskreutz, Albert C Ludolph, Susanne Petri, Mamede de Carvalho, Philip Van Damme, Kevin Talbot, Martin R Turner, Pamela J Shaw, Ammar Al-Chalabi, Adriano Chiò, Orla Hardiman, Karel G M Moons, Jan H Veldink, Leonard H van den Berg

Correspondence

The life expectancy of Stephen Hawking, according to the ENCALS model

Stephen W Hawking, one of the most famous physicists, died on March 14, 2018, at the age of 76 years. Although he was best known for his remarkable and survival can be exceptionally long in some cases.[3,4] We have examined Professor Hawking's clinical phenotype using our recently validated predictive model of survival (the ENCALS survival model), which is based on eight predictors.[3] The model was designed to generate survival probabilities on the basis of a composite endpoint, which we defined as death, tracheostomy, or dependency on non-invasive venti- 1985, which is within this interval, the endpoint was reached because he had a tracheostomy. Then, the model predicts that he had a 20% probability of surviving, to the time of his death some 33 years later. According to the ENCALS survival model, Professor Hawking's young age of onset was the most important factor for his long survival. However, more than half of his disease duration (ie, after

doi 10.1016/S1474-4422(18)30089-9
doi 10.1016/S1474-4422(18)30241-2

THE LANCET
Neurology

Personalized medicine for infectious diseases

Main challenges to manage, exchange and preserve research data efficiently

- Separate storage of research data hampers both within- and cross-study scientific advancement
- Density of generated data often exceeds the storage capacity of typical databases
- Lack of resources for integrated data analysis across cohorts
- Governance of sample- and data sharing is complicated
- Utility of OMICS data disproportionally affect countries in the global South.

Personalized medicine approaches hold tremendous promise for improving the detection of infectious diseases and for developing targeted treatment strategies.

Personalized medicine for infectious diseases

Horizon 2020 project of 6 mln EUR

- To develop an integrated sustainable platform for collating data within and across infectious disease cohorts
- To develop innovative solutions for shared ownership, linked data and biorepositories
- To develop and evaluate statistical methods for collaborative analysis (Utrecht)

European Commission

Collaboration with physicians, virologists, micro-biologists,epidemiologists, bio-informaticians, bio-statisticians, ethicists, and many more

Final thoughts

Big collaborative projects are increasingly important

- To improve the quality and reproducibility of research
- To address more complex questions
- To provide more generalizable answers
- To enhance impact of research findings

Funding organizations play an important role to identify and finance critical expertise. Academic institutions need to train scientists and foster their creativity.