UMC Utrecht
Julius Center

Predicting the unknown in health care
challenges and opportunities

Thomas Debray, PhD

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

Cochrane

December 1, 2016

University Medical Center Utrecht

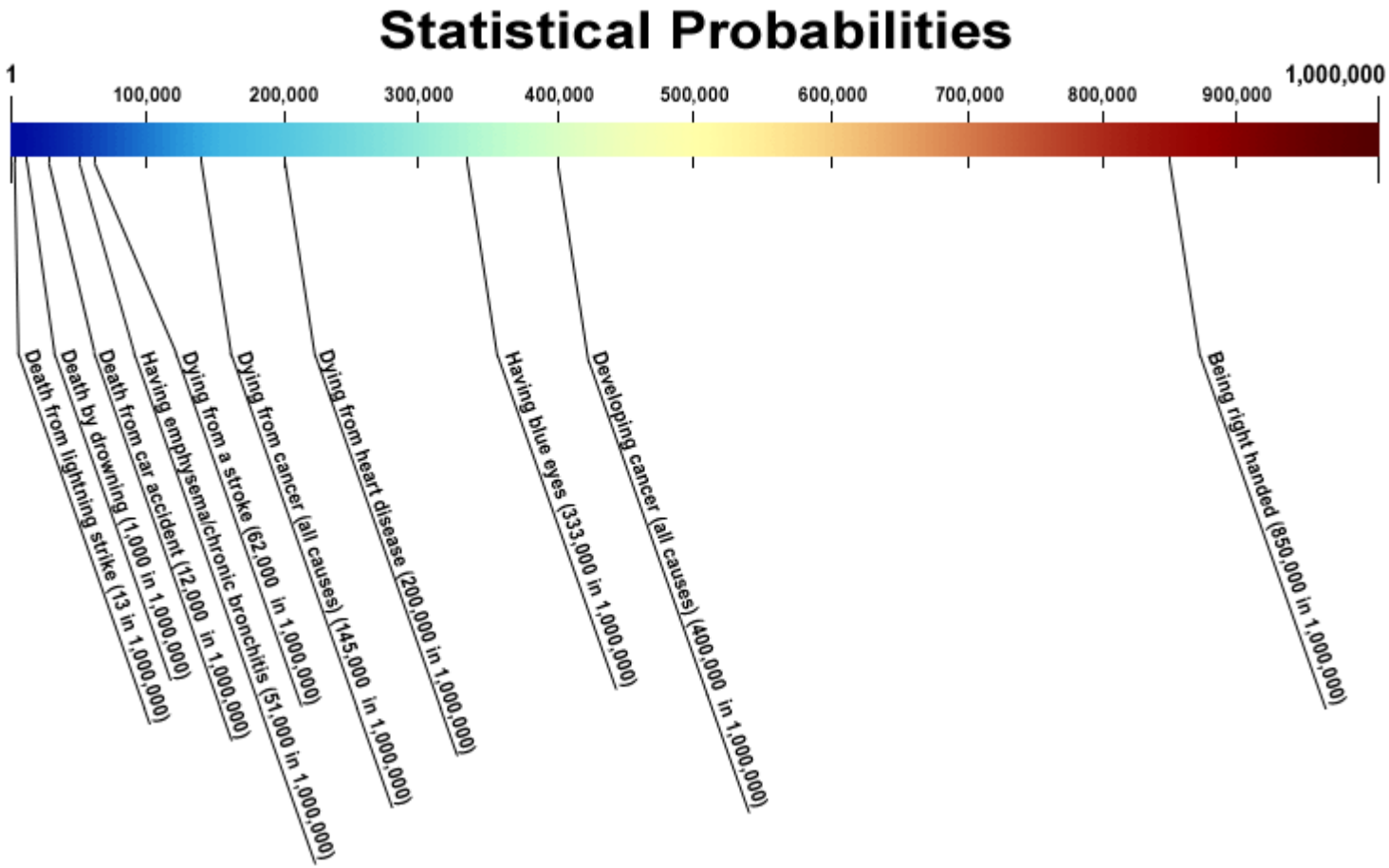

Stratenum

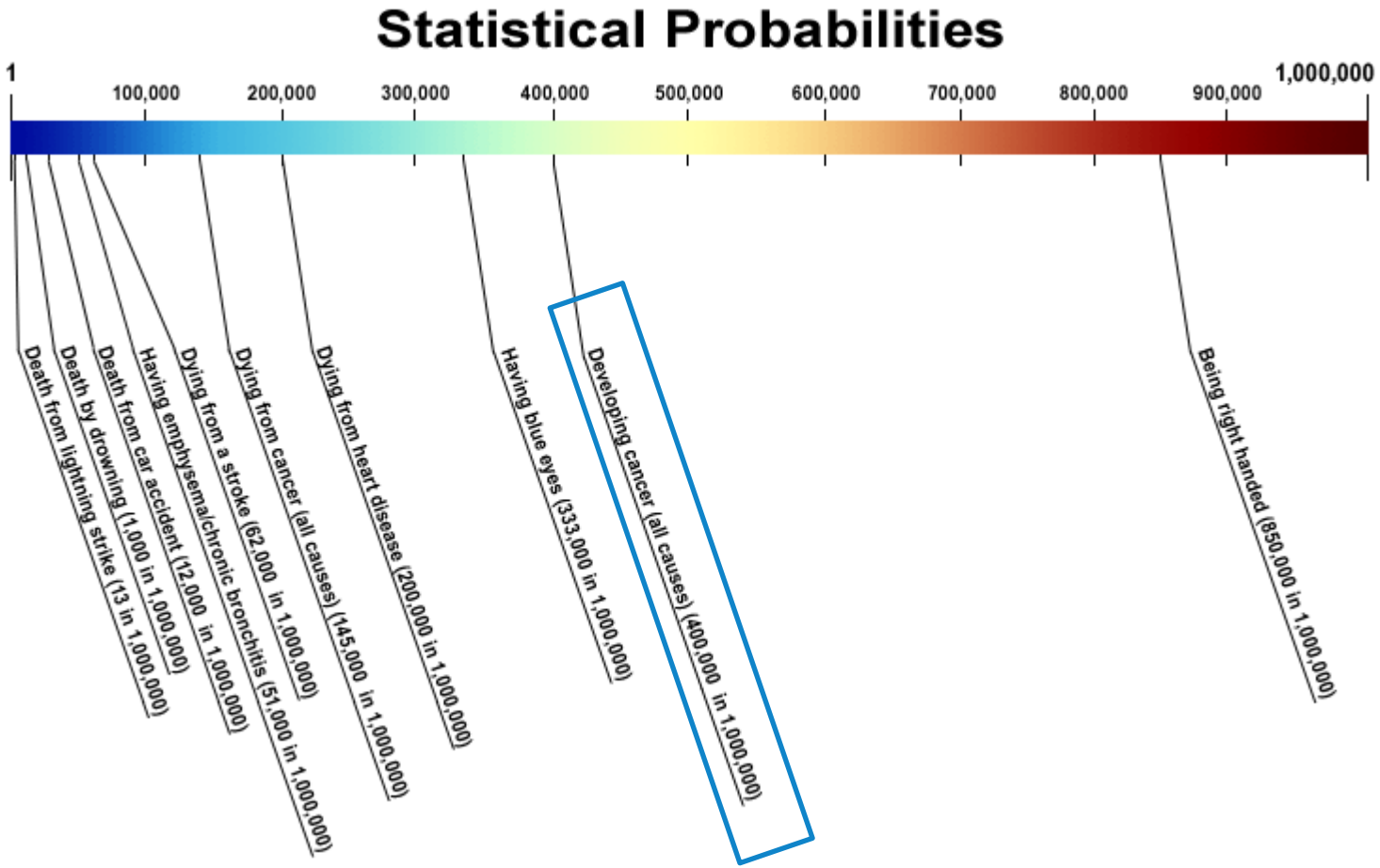
Prediction

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

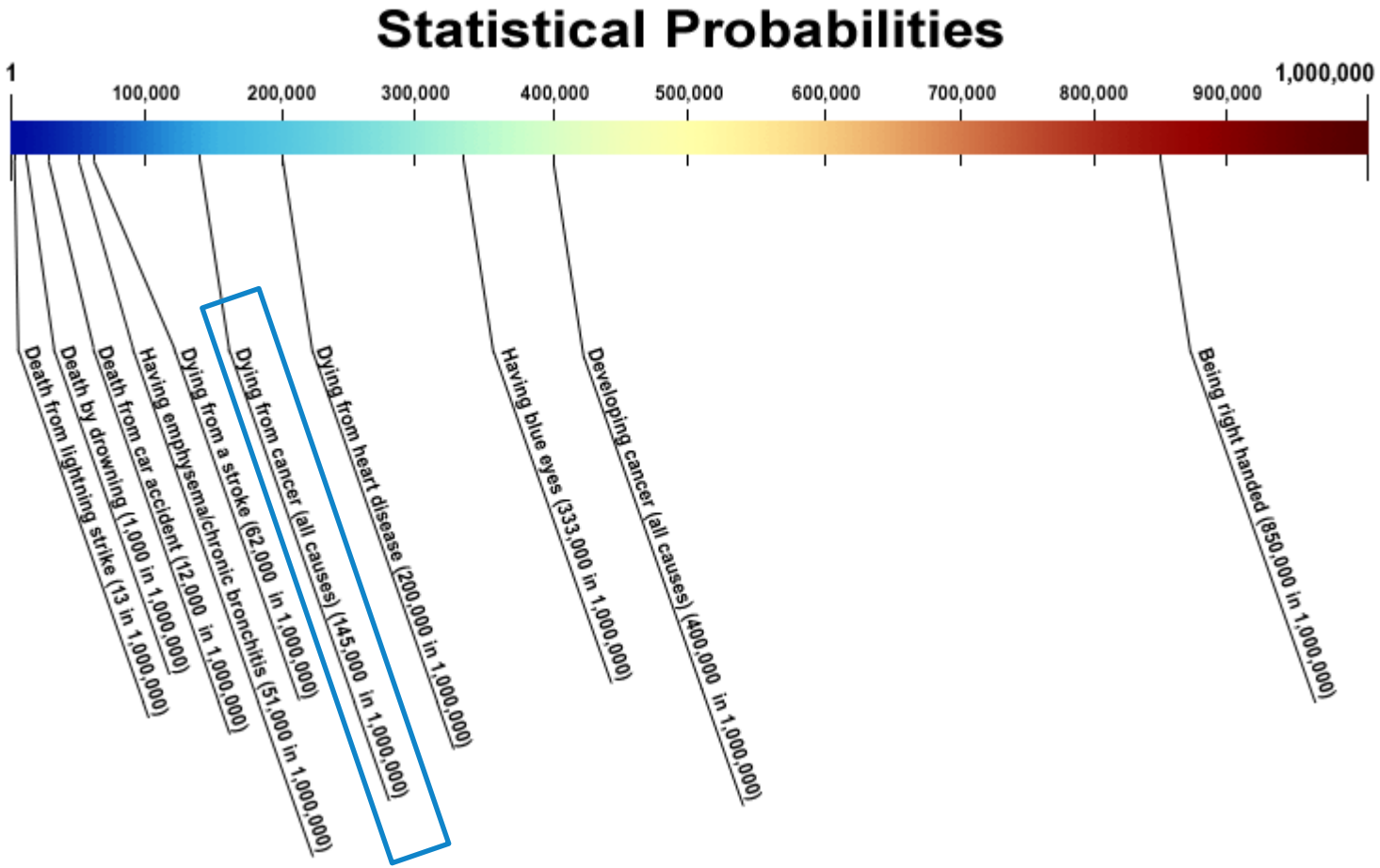
- Turn available information (predictors) into a statement about the probability:
  … of having a particular disease -> **diagnosis**
  … of developing a particular event -> **prognosis**

- Use of prognostic information:
  - to inform patients and their families
  - to guide treatment and other clinical decisions
  - to create risk groups
  - …


Statistical Probabilities
1
100,000
200,000
300,000
400,000
500,000
600,000
700,000
800,000
900,000
1,000,000
Death from lightning strike (13 in 1,000,000)
Death by drowning (1,000 in 1,000,000)
Death from car accident (12,000 in 1,000,000)
Having emphysema/chronic bronchitis (51,000 in 1,000,000)
Dying from a stroke (62,000 in 1,000,000)
Dying from cancer (all causes) (145,000 in 1,000,000)
Dying from heart disease (200,000 in 1,000,000)
Having blue eyes (333,000 in 1,000,000)
Developing cancer (all causes) (400,000 in 1,000,000)
Being right handed (850,000 in 1,000,000)



Risk of developing cancer

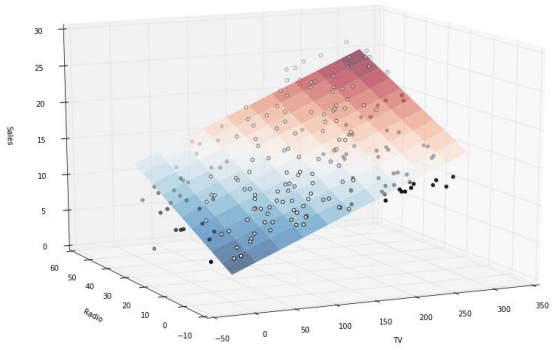


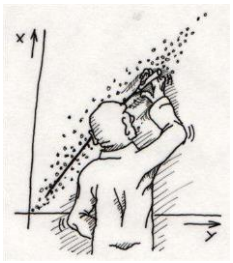
Risk of dying from cancer

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 patient data from large cohort studies
  - Many strategies available (Regression, decision trees, neural networks, …)





Breast Cancer Risk Assessment Tool

An interactive tool to help estimate a woman's risk of developing breast cancer

Last modified date: 05/16/2011

- Get Started with the Risk Tool
- About the Tool
- Breast Cancer Risk Factors
- Download Source Code

Page Options

Print Page

Quick Links

- Breast Cancer Home Page
- Breast Cancer: Prevention, Genetics, Causes
- Current Clinical Trials: Breast Cancer *In Situ*: Treatment
- Current Clinical Trials: Breast Cancer Prevention
- Current Clinical Trials: Breast Cancer Screening
- Breast Cancer Risk in American Women

Need Help?

Contact us by phone, Web, and e-mail

1-800-4-CANCER

The Breast Cancer Risk Assessment Tool is an interactive tool designed by scientists at the National Cancer Institute (NCI) and the National Surgical Adjuvant Breast and Bowel Project (NSABP) to estimate a woman's risk of developing invasive breast cancer. See About the Tool for more information.

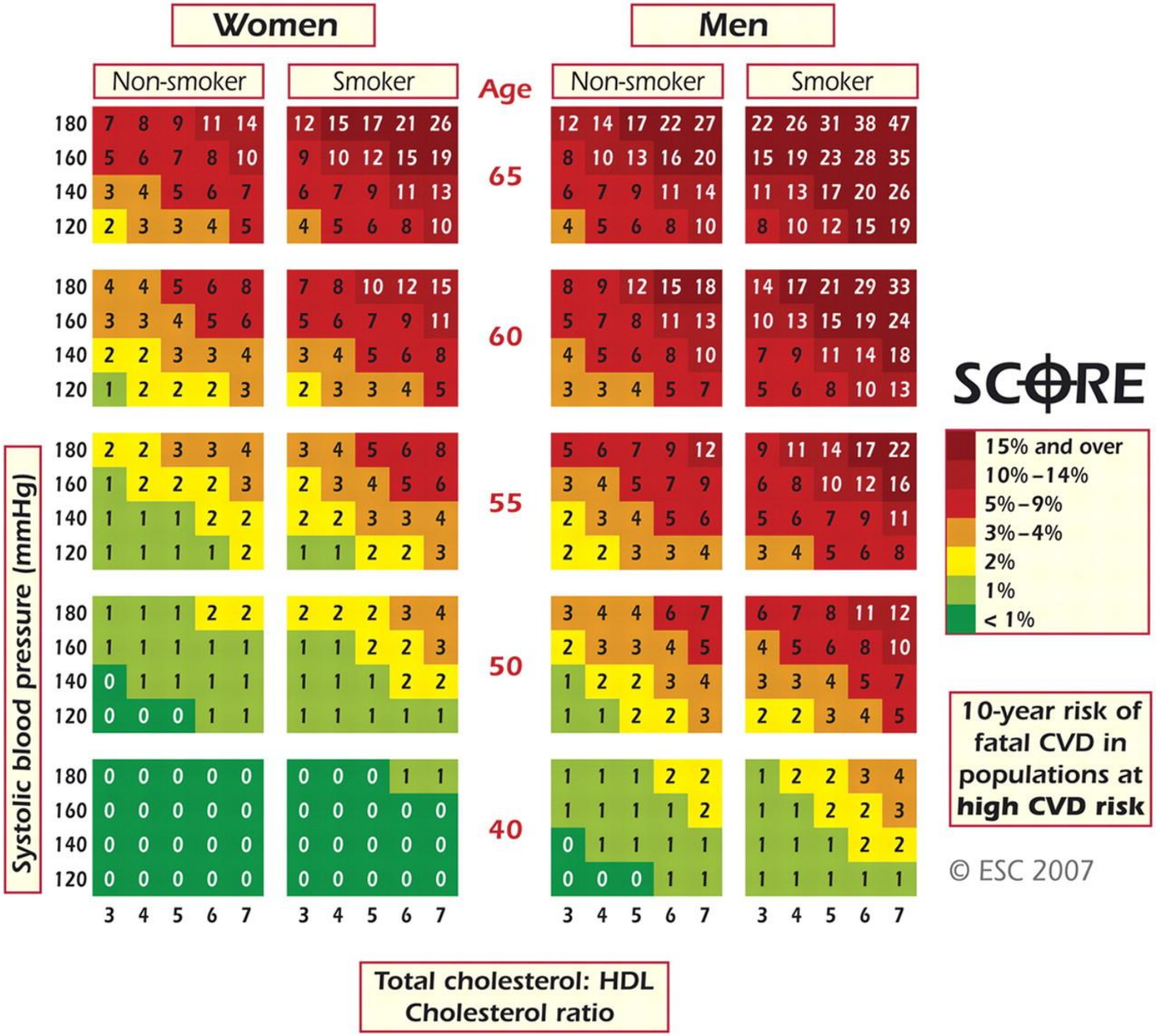
The Breast Cancer Risk Assessment Tool may be updated periodically as new data or research becomes available.

Risk Tool

(Click a question number for a brief explanation, or read all explanations.)

1. Does the woman have a medical history of any breast cancer or of ductal carcinoma in situ (DCIS) or lobular carcinoma in situ (LCIS) or has she received previous radiation therapy to the chest for treatment of Hodgkin lymphoma? — Select
2. Does the woman have a mutation in either the *BRCA1* or *BRCA2* gene, or a diagnosis of a genetic syndrome that may be associated with elevated risk of breast cancer? — Select
3. What is the woman's age? *This tool only calculates risk for women 35 years of age or older.* — Select
4. What was the woman's age at the time of her first menstrual period? — Select
5. What was the woman's age at the time of her first live birth of a child? — Select
6. How many of the woman's first-degree relatives - mother, sisters, daughters - have had breast cancer? — Select
7. Has the woman ever had a breast biopsy? — Select
   - 7a. How many breast biopsies (positive or negative) has the woman had? — Select
   - 7b. Has the woman had at least one breast biopsy with atypical hyperplasia? — Select
8. What is the woman's race/ethnicity? — Select
   - 8a. What is the sub race/ethnicity? — Select

Calculate Risk >


Women
Men
Non-smoker
Smoker
Non-smoker
Smoker
Age
65
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
60
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
55
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
50
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
40
180 0 0 0 0 0 | 0 0 0 1 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
3 4 5 6 7 | 3 4 5 6 7 | 3 4 5 6 7 | 3 4 5 6 7
Systolic blood pressure (mmHg)
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

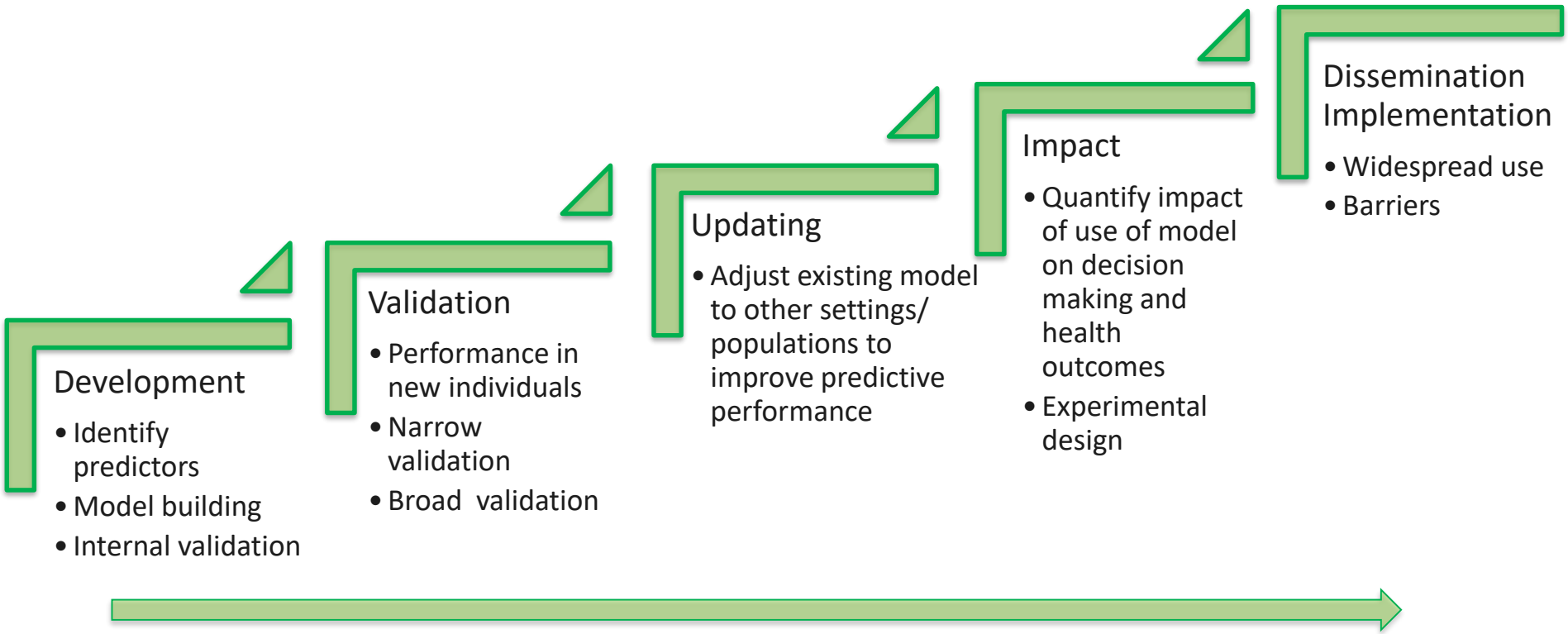
Prediction

What is a good model?

- Generates accurate predictions in individuals from potential population(s) for clinical use
- Ability to discriminate between different risk groups
- Improves patient outcomes by informing treatment decisions

Phases of prediction model evaluation

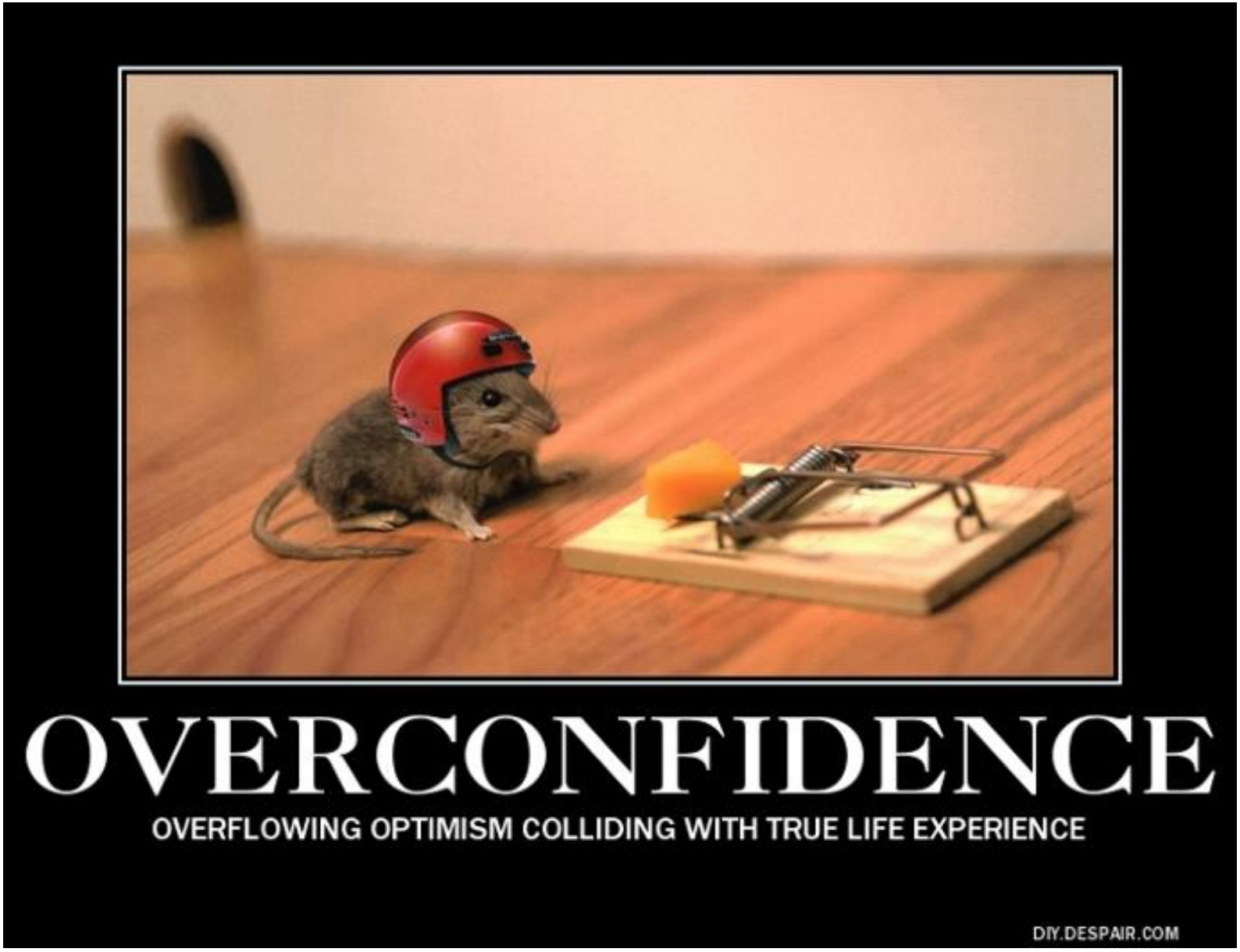
Series in BMJ 2009 and in Heart 2012, Moons et al.



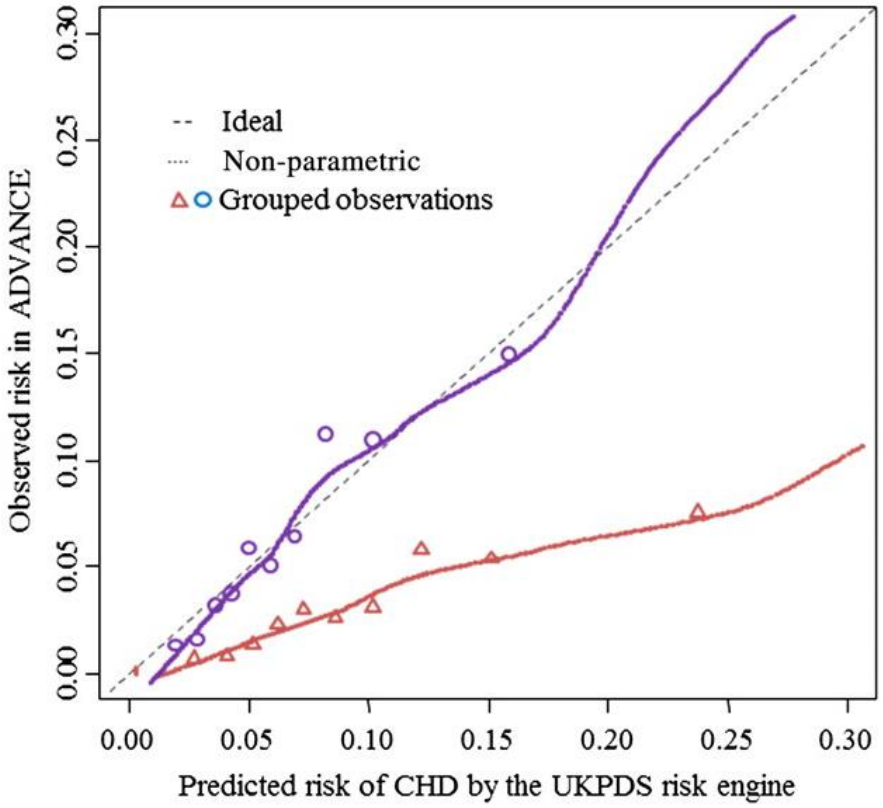
Increasing level of evidence for use of model in practice

Common pitfalls

UMC Utrecht


OVERCONFIDENCE
OVERFLOWING OPTIMISM COLLIDING WITH TRUE LIFE EXPERIENCE
DIY.DESPAIR.COM

Poor predictive accuracy

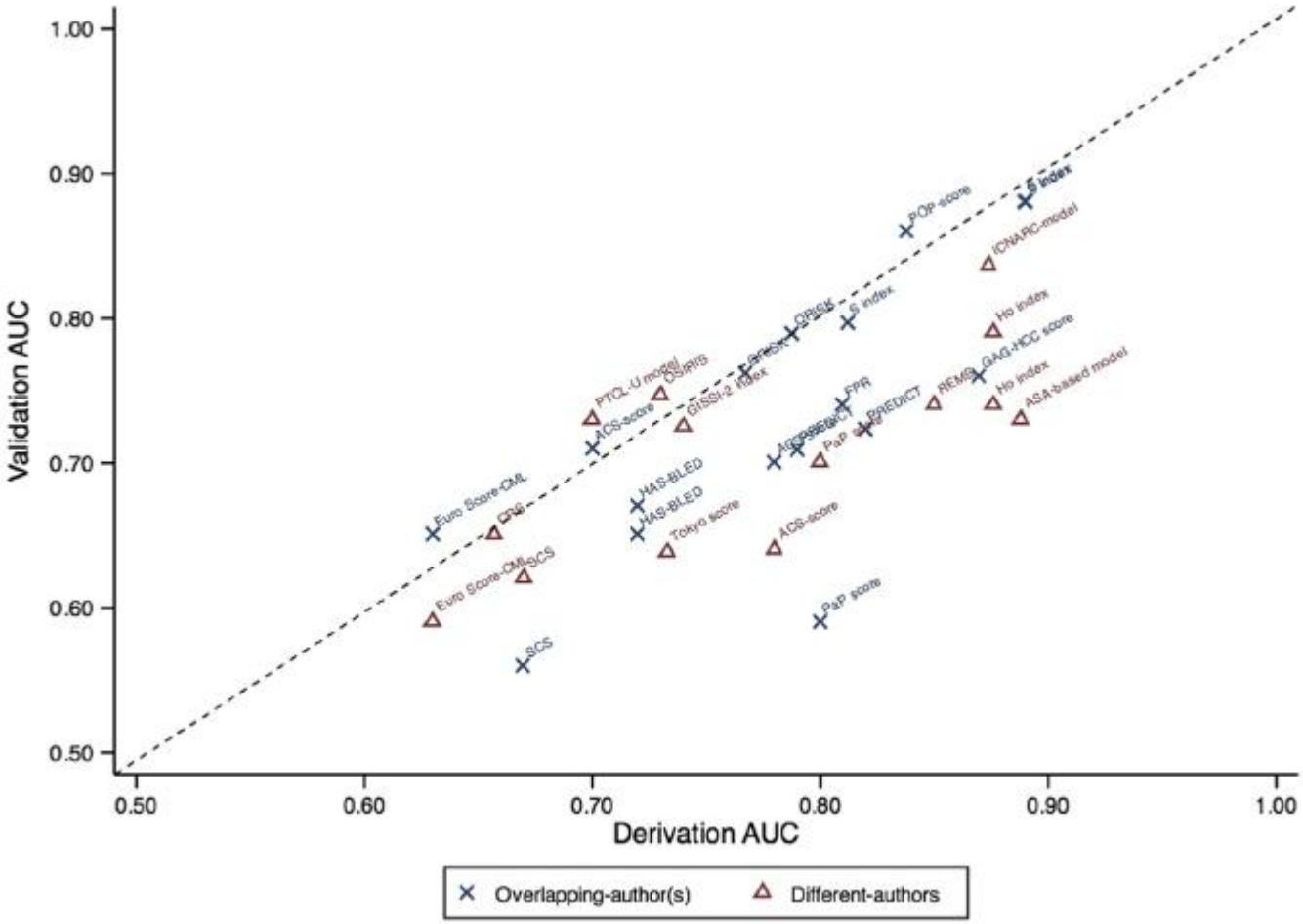


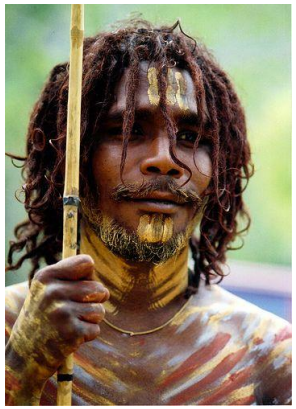
Calibration plot for the 4-year predicted risk of major coronary heart disease (CHD)

Results for the original risk equation and the updated risk equation .

Ref: Moons KGM, *et al*. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012 Mar 7;98(9):691–8.

Lack of transportability





Ref: Siontis *et al.* External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. Journal of Clinical Epidemiology. 2014.

Reasons for performance changes

- Over-fitting
- Missed interactions and non-linear trends of predictors
- Biomarkers: different measurement method, recording time point or cut-off across settings
- Case-mix variation
- Different standards of care and treatment strategies
- Different startpoints (e.g. due to screening)

Lack of (independent) validation

Summarized

Most models are not as good as we think

- Poor quality of prognostic modelling studies
  - Limited sample size
  - Incomplete registrations & reporting
  - Absent study protocols
- Poor transportability
  - Case-mix variation across populations
  - Differences in measurement methods
  - Time-varying predictor effects
  - Changes in standards of care and treatment strategies
- Lack of external validation

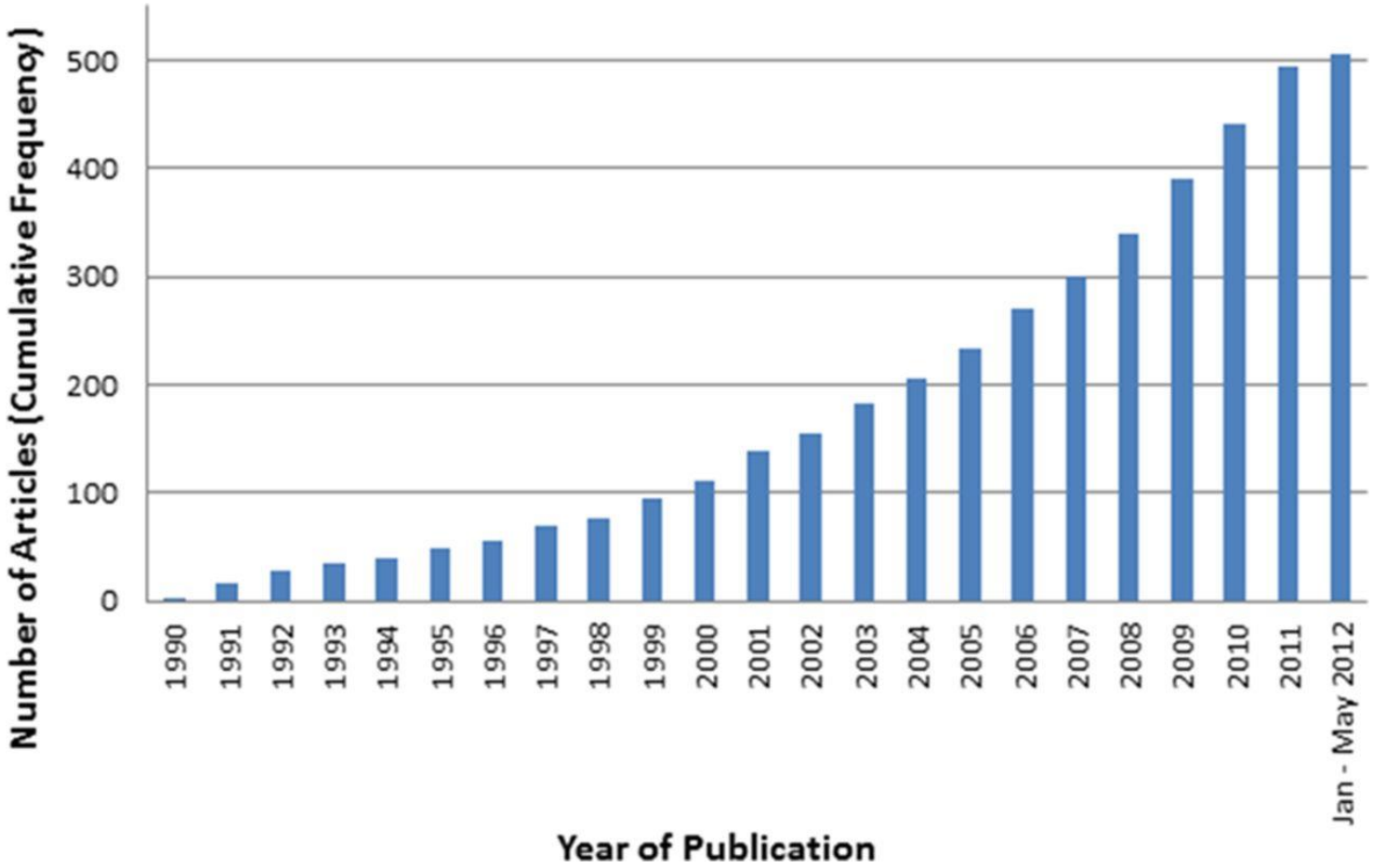
But wait… this is not the end

There are numerous models for same target population and outcomes

- >150 models alike Framingham, SCORE, Qrisk
- >100 models for brain trauma patients
- > 100 diabetes type 2 models
- > 60 models for breast cancer prognosis

Why focus on prediction models?

Cumulative growth in published CPM articles over time



American Heart Association.

Ref: Benjamin S. Wessler et al. Circ Cardiovasc Qual Outcomes. 2015;8:368-375

Numerous models for same target population + outcomes

"Comparing risk prediction models should be routine when deriving a new model for the same purpose" (Collins 2012)

"Substantial work is needed to understand how competing prediction models compare and how they can best be applied to individualize care." (Wessler 2015)

"There is an excess of models predicting incident CVD in the general population. The usefulness of most of the models remains unclear." (Damen 2016)

thebmj

American Heart Association.

Opportunities

Evidence Synthesis

Big Data

Machine Learning

UMC Utrecht

Evidence synthesis

Why?

- Improve estimation of prediction models
- Evaluate sources of variability in predictive performance
- Evaluate need for tailoring

How?

- Synthesis of prognostic factors
- Synthesis of prediction models
- Synthesis of prediction model performance

Evidence synthesis

Combining information on prognostic factors

Concept: Use previously published risk factor associations to update multivariable coefficients in “own” data set

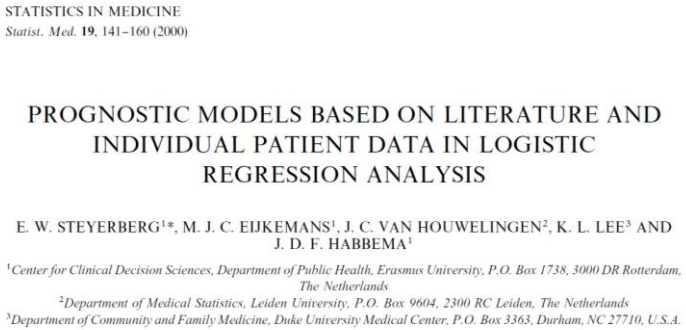
Debray *et al. BMC Medical Research Methodology* 2012, **12**:121
http://www.biomedcentral.com/1471-2288/12/121

BMC Medical Research Methodology

TECHNICAL ADVANCE Open Access

Incorporating published univariable associations in diagnostic and prognostic modeling

Thomas P A Debray[1*], Hendrik Koffijberg[1], Difei Lu[2], Yvonne Vergouwe[1,2], Ewout W Steyerberg[2†] and Karel G M Moons[1†]



STATISTICS IN MEDICINE
Statist. Med. **19**, 141–160 (2000)

PROGNOSTIC MODELS BASED ON LITERATURE AND INDIVIDUAL PATIENT DATA IN LOGISTIC REGRESSION ANALYSIS

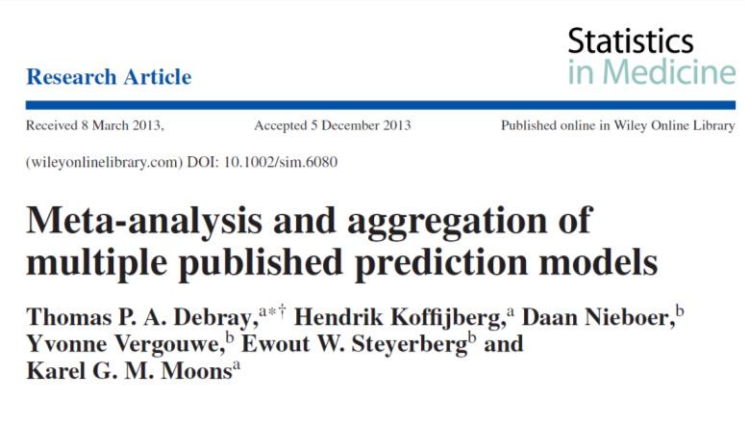
E. W. STEYERBERG[1]*, M. J. C. EIJKEMANS[1], J. C. VAN HOUWELINGEN[2], K. L. LEE[3] AND J. D. F. HABBEMA[1]

[1]*Center for Clinical Decision Sciences, Department of Public Health, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands*
[2]*Department of Medical Statistics, Leiden University, P.O. Box 9604, 2300 RC Leiden, The Netherlands*
[3]*Department of Community and Family Medicine, Duke University Medical Center, P.O. Box 3363, Durham, NC 27710, U.S.A.*

Evidence synthesis

Combining previously published prediction models

Concept: Use limited patient-level data at hand to combine and tailor previously published models



Research Article

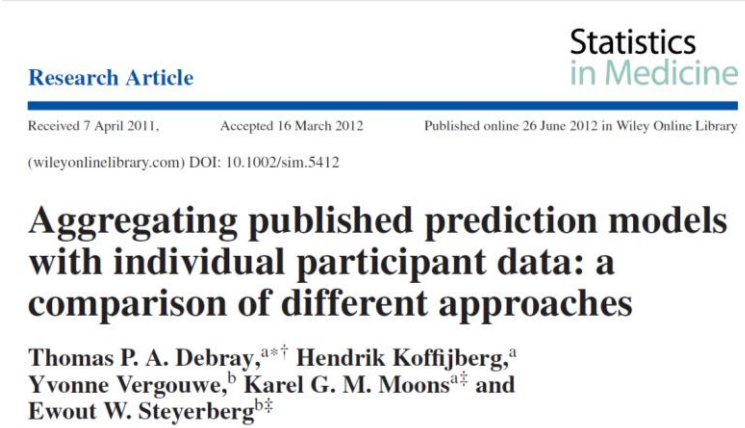
Statistics in Medicine

Received 8 March 2013, Accepted 5 December 2013 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6080

Meta-analysis and aggregation of multiple published prediction models

Thomas P. A. Debray,[a*†] Hendrik Koffijberg,[a] Daan Nieboer,[b] Yvonne Vergouwe,[b] Ewout W. Steyerberg[b] and Karel G. M. Moons[a]



Research Article

Statistics in Medicine

Received 7 April 2011, Accepted 16 March 2012 Published online 26 June 2012 in Wiley Online Library

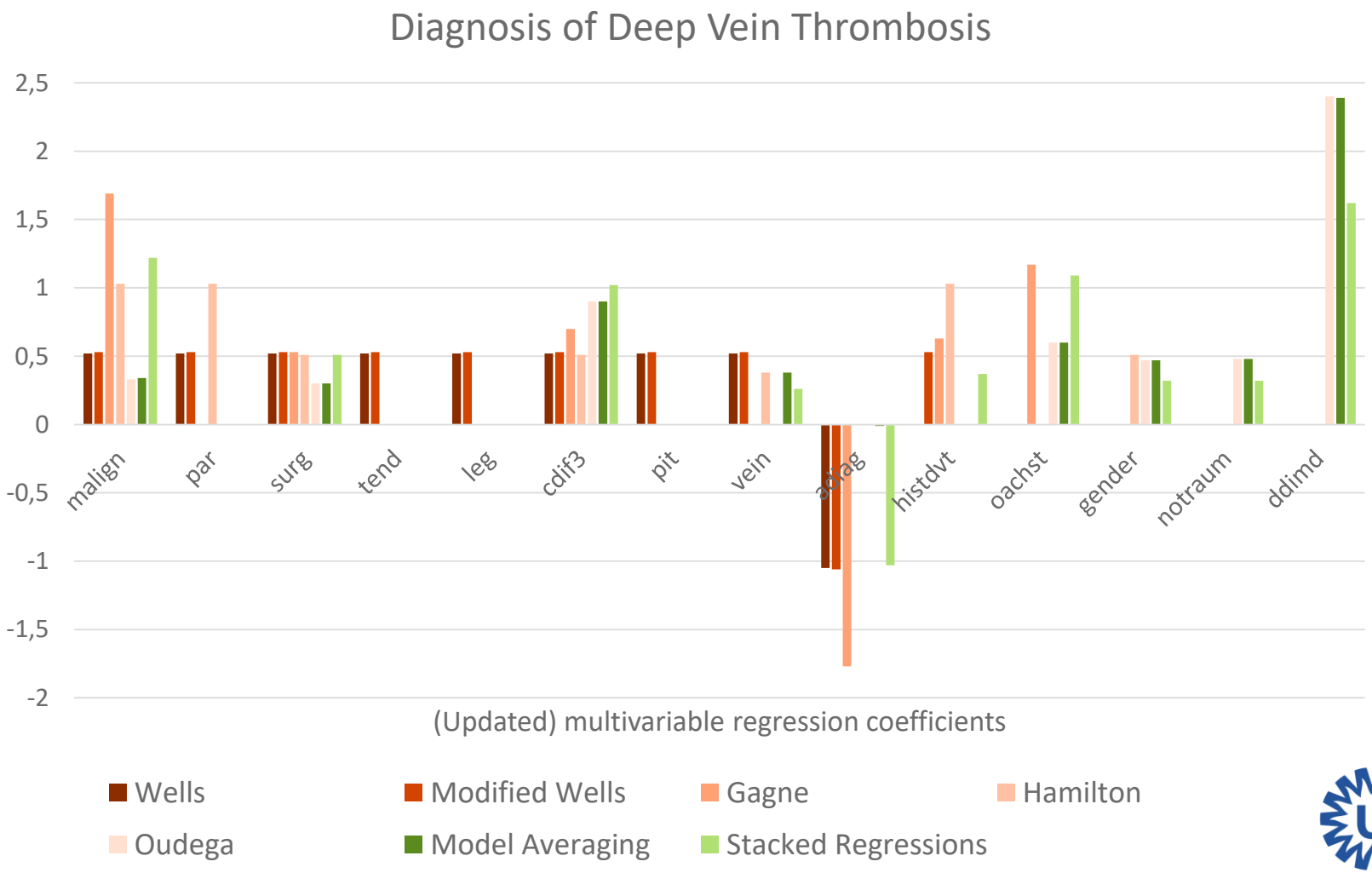
(wileyonlinelibrary.com) DOI: 10.1002/sim.5412

Aggregating published prediction models with individual participant data: a comparison of different approaches

Thomas P. A. Debray,[a*†] Hendrik Koffijberg,[a] Yvonne Vergouwe,[b] Karel G. M. Moons[a‡] and Ewout W. Steyerberg[b‡]

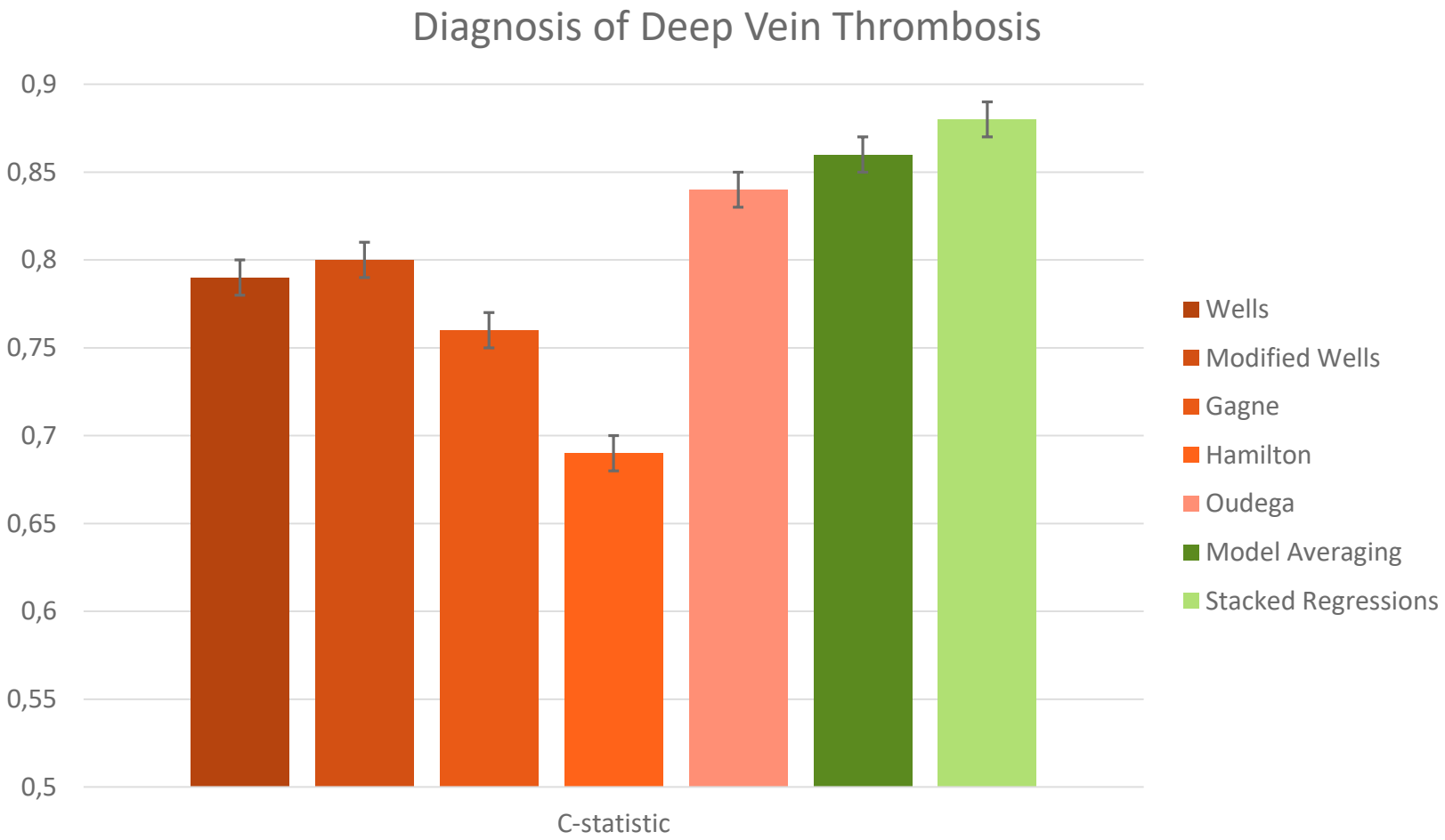
Evidence synthesis

Combining previously published prediction models



Evidence synthesis

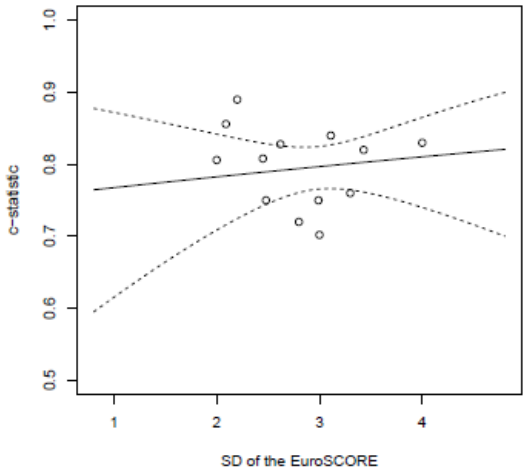
Combining previously published prediction models

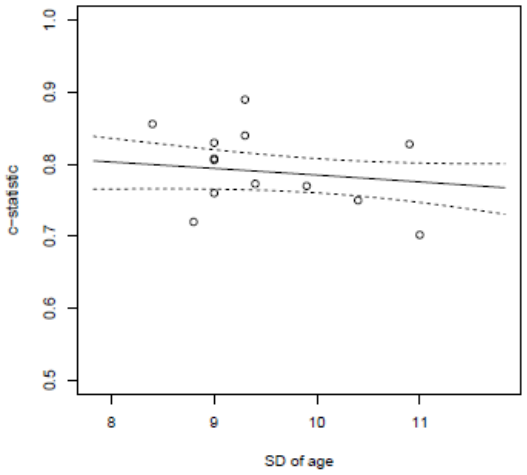


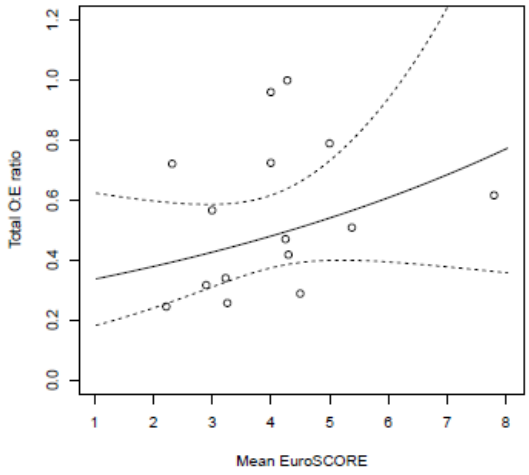
Evidence synthesis

Summarizing external validation study results

Concept: Systematically review external validation studies of a certain prediction model and summarize their results







Ref: Debray TPA, *et al*. A guide to systematic review and meta-analysis of prediction model performance. BMJ 2016 (Accepted for publication)

Opportunities

Evidence Synthesis

Big Data

Machine Learning

UMC Utrecht

The rise of big data

What is 'big data'?

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

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

Example: QRISK2 was developed using e-health data from the QRESEARCH database using over 1.5 million patients (with over 95000 new cardiovascular events) from 355 randomly selected general practices

Prediction research using big data

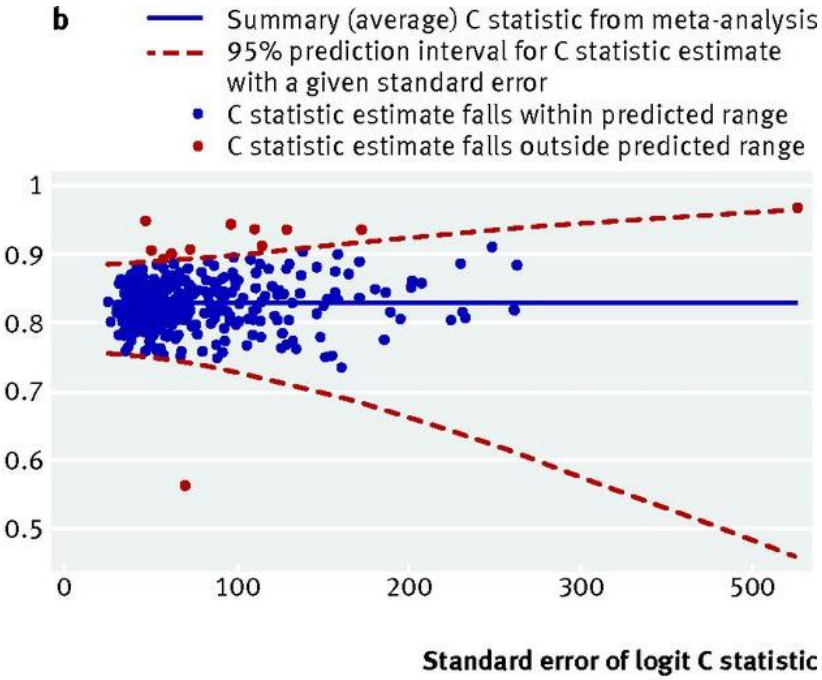
Why do we need 'big data'?

- Development of better prediction models
  - Reduced risk of overfitting
  - Ability to address wider spectrum of patients
  - Ability to investigate more complex associations

- More extensive testing of model performance
  - Ability to externally validate across multiple settings (also upon model development)
  - Ability to investigate sources of poor or inconsistent model performance
  - Ability to assess usability of prediction models across different situations

Prediction research using big data

Evaluate model generalizability



Discrimination performance of QRISK2, across 364 general practice surgeries

Ref: Riley RD, *et al*. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ. 2016;353:i3140.

Prediction research using big data

Identify most promising development strategy

- To model complex associations
- To account for differences between study populations

Table 2. Joint predicted probability of "good" discrimination and calibration performance of the DVT model for each of the three implementation strategies, derived using the multivariate meta-analysis results for the C statistic and calibration slope shown in Table 1

		Joint predicted probability of meeting criteria in new population		
Calibration slope required	Minimum C statistic required	Strategy (1): Develop using logistic regression and implement with intercept estimated in external validation study	Strategy (2): Develop using logistic regression and implement with average study intercept taken from developed model	Strategy (3): Develop using logistic regression and implement with intercept taken from a study used in development data with a similar prevalence
0.9–1.1	0.70	0.027	0.037	0.037
0.8–1.2	0.70	0.146	0.158	0.156
0.9–1.1	0.65	0.427	0.413	0.409
0.8–1.2	0.65	0.728	0.712	0.707

Abbreviation: DVT, deep vein thrombosis.

Refs: Debray TPA, *et al*. A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. Stat Med. 2013 Aug 15;32(18):3158–80.
Snell KIE, *et al*. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model. J Clin Epidemiol. 2015 May;69:40–50.

Prediction research using big data

PLOS | MEDICINE

GUIDELINES AND GUIDANCE

Individual Participant Data (IPD) Meta-analyses of Diagnostic and Prognostic Modeling Studies: Guidance on Their Use

Thomas P. A. Debray[1,2]*, Richard D. Riley[3], Maroeska M. Rovers[4], Johannes B. Reitsma[1,2], Karel G. M. Moons[1,2], Cochrane IPD Meta-analysis Methods group[¶]

1 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands, **2** The Dutch Cochrane Centre, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands, **3** Research Institute for Primary Care and Health Sciences, Keele University, Staffordshire, The United Kingdom, **4** Radboud Institute for Health Sciences, Radboudumc Nijmegen, The Netherlands

¶ Membership of the Cochrane IPD Meta-analysis Methods group is listed in the Acknowledgments.
* T.Debray@umcutrecht.nl

CrossMark
click for updates

Opportunities

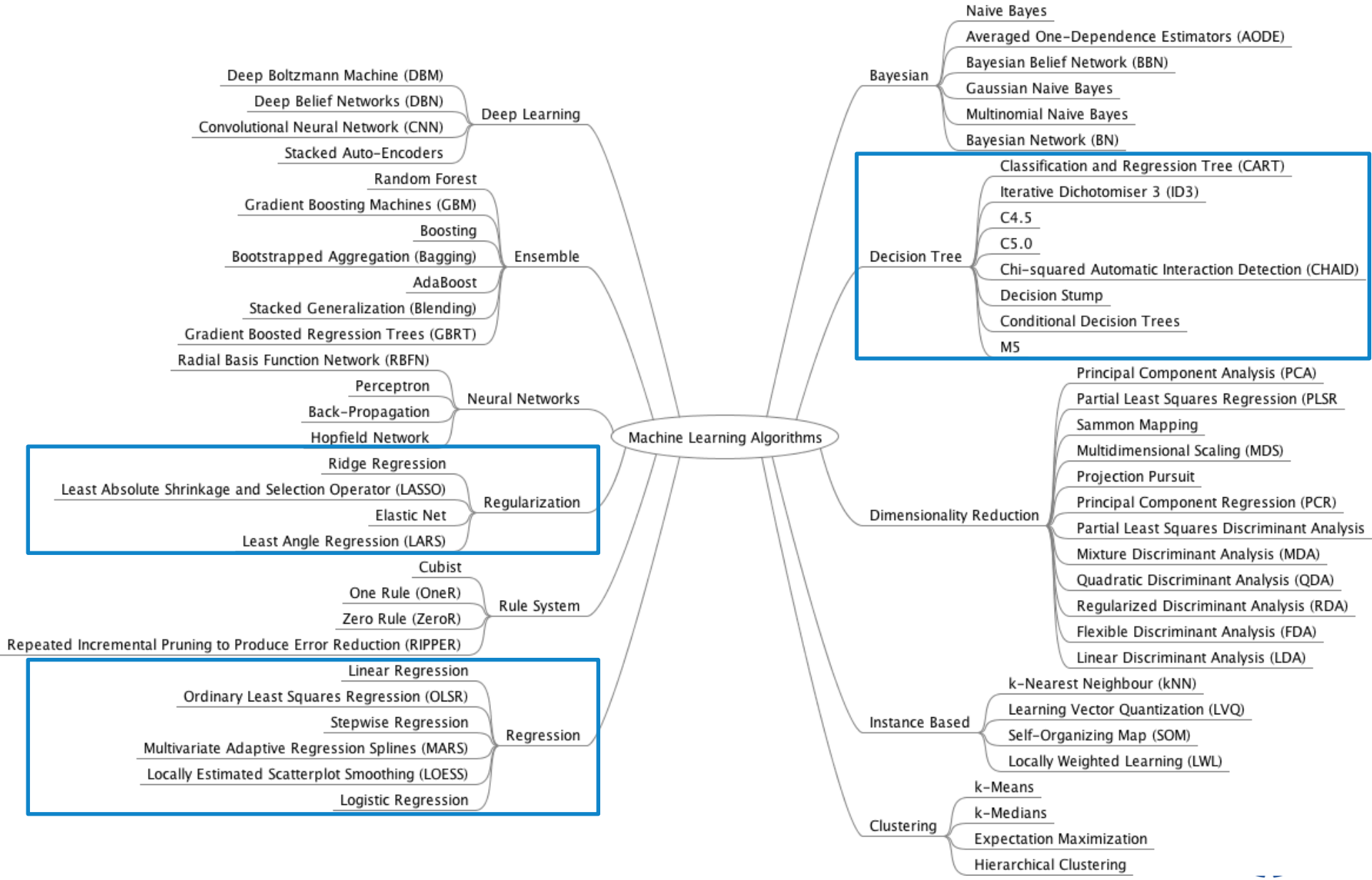
Evidence Synthesis

Big Data

Machine Learning

UMC Utrecht


Machine Learning Algorithms
Deep Learning
Deep Boltzmann Machine (DBM)
Deep Belief Networks (DBN)
Convolutional Neural Network (CNN)
Stacked Auto-Encoders
Ensemble
Random Forest
Gradient Boosting Machines (GBM)
Boosting
Bootstrapped Aggregation (Bagging)
AdaBoost
Stacked Generalization (Blending)
Gradient Boosted Regression Trees (GBRT)
Neural Networks
Radial Basis Function Network (RBFN)
Perceptron
Back-Propagation
Hopfield Network
Regularization
Ridge Regression
Least Absolute Shrinkage and Selection Operator (LASSO)
Elastic Net
Least Angle Regression (LARS)
Rule System
Cubist
One Rule (OneR)
Zero Rule (ZeroR)
Repeated Incremental Pruning to Produce Error Reduction (RIPPER)
Regression
Linear Regression
Ordinary Least Squares Regression (OLSR)
Stepwise Regression
Multivariate Adaptive Regression Splines (MARS)
Locally Estimated Scatterplot Smoothing (LOESS)
Logistic Regression
Bayesian
Naive Bayes
Averaged One-Dependence Estimators (AODE)
Bayesian Belief Network (BBN)
Gaussian Naive Bayes
Multinomial Naive Bayes
Bayesian Network (BN)
Decision Tree
Classification and Regression Tree (CART)
Iterative Dichotomiser 3 (ID3)
C4.5
C5.0
Chi-squared Automatic Interaction Detection (CHAID)
Decision Stump
Conditional Decision Trees
M5
Dimensionality Reduction
Principal Component Analysis (PCA)
Partial Least Squares Regression (PLSR)
Sammon Mapping
Multidimensional Scaling (MDS)
Projection Pursuit
Principal Component Regression (PCR)
Partial Least Squares Discriminant Analysis
Mixture Discriminant Analysis (MDA)
Quadratic Discriminant Analysis (QDA)
Regularized Discriminant Analysis (RDA)
Flexible Discriminant Analysis (FDA)
Linear Discriminant Analysis (LDA)
Instance Based
k-Nearest Neighbour (kNN)
Learning Vector Quantization (LVQ)
Self-Organizing Map (SOM)
Locally Weighted Learning (LWL)
Clustering
k-Means
k-Medians
Expectation Maximization
Hierarchical Clustering


Machine Learning Algorithms
Bayesian
Naive Bayes
Averaged One-Dependence Estimators (AODE)
Bayesian Belief Network (BBN)
Gaussian Naive Bayes
Multinomial Naive Bayes
Bayesian Network (BN)
Decision Tree
Classification and Regression Tree (CART)
Iterative Dichotomiser 3 (ID3)
C4.5
C5.0
Chi-squared Automatic Interaction Detection (CHAID)
Decision Stump
Conditional Decision Trees
M5
Dimensionality Reduction
Principal Component Analysis (PCA)
Partial Least Squares Regression (PLSR)
Sammon Mapping
Multidimensional Scaling (MDS)
Projection Pursuit
Principal Component Regression (PCR)
Partial Least Squares Discriminant Analysis
Mixture Discriminant Analysis (MDA)
Quadratic Discriminant Analysis (QDA)
Regularized Discriminant Analysis (RDA)
Flexible Discriminant Analysis (FDA)
Linear Discriminant Analysis (LDA)
Instance Based
k-Nearest Neighbour (kNN)
Learning Vector Quantization (LVQ)
Self-Organizing Map (SOM)
Locally Weighted Learning (LWL)
Clustering
k-Means
k-Medians
Expectation Maximization
Hierarchical Clustering
Deep Learning
Deep Boltzmann Machine (DBM)
Deep Belief Networks (DBN)
Convolutional Neural Network (CNN)
Stacked Auto-Encoders
Ensemble
Random Forest
Gradient Boosting Machines (GBM)
Boosting
Bootstrapped Aggregation (Bagging)
AdaBoost
Stacked Generalization (Blending)
Gradient Boosted Regression Trees (GBRT)
Neural Networks
Radial Basis Function Network (RBFN)
Perceptron
Back-Propagation
Hopfield Network
Regularization
Ridge Regression
Least Absolute Shrinkage and Selection Operator (LASSO)
Elastic Net
Least Angle Regression (LARS)
Rule System
Cubist
One Rule (OneR)
Zero Rule (ZeroR)
Repeated Incremental Pruning to Produce Error Reduction (RIPPER)
Regression
Linear Regression
Ordinary Least Squares Regression (OLSR)
Stepwise Regression
Multivariate Adaptive Regression Splines (MARS)
Locally Estimated Scatterplot Smoothing (LOESS)
Logistic Regression

Potential of Machine Learning

Machine Learning not widely implemented yet…

- Loss of transparency
- Performance gain often disappointing

ELSEVIER

Journal of Clinical Epidemiology 65 (2012) 404–412

Journal of Clinical Epidemiology

Development and validation of clinical prediction models: Marginal differences between logistic regression, penalized maximum likelihood estimation, and genetic programming

Kristel J.M. Janssen[a,*], Ivar Siccama[b], Yvonne Vergouwe[a], Hendrik Koffijberg[a], T.P.A. Debray[a], Maarten Keijzer[c], Diederick E. Grobbee[a], Karel G.M. Moons[a]

[a]*Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, P.O. Box 85500, 3508 AB Utrecht, The Netherlands*
[b]*Department of Neurology, Erasmus Medical Center, Rotterdam, The Netherlands*
[c]*Pegasystems Benelux, Amsterdam, The Netherlands*

Accepted 9 August 2011; Published online 02 January 2012

Potential of Machine Learning

With the rise of big data, the appeal of machine learning is increasing.

Key strenghts

- Handling enormous numbers of predictors
- Modeling highly interactive and nonlinear effects

Potential of Machine Learning

Promising areas of application

- Analysis of unstructured data
  - Text (e.g. medical records)
  - Images (e.g. CT, MRI, …)
- Analysis of high velocity data
  - Brain signals (e.g. restoration of motor control)
  - Wearable devices
  - Social media
- Diagnosis
  - Generation of differential diagnoses
  - Suggestion of high-value tests

Ref: Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med. 2016 Sep 29;375(13):1216–9.

Reasons to be optimistic?