UMC Utrecht
Julius Center

Systematic reviews of prognostic studies: a meta-analytical approach

Thomas PA Debray, Karel GM Moons

for the Cochrane Prognosis Review Methods Group
(Co-convenors: Doug Altman, Katrina Williams, Jill Hayden, Sue Woolfenden, Richard Riley, Karel Moons)

UNIVERSITY OF BIRMINGHAM

The Cochrane Collaboration
Working together to provide the best evidence for health care

University Medical Center Utrecht

Conflict of interest

We have no actual or potential conflict of interest in relation to this presentation

Overview Cochrane Prognostic Methods Group (PMG) Workshops

PMG Workshop	Facilitators	When?
W72. Systematic reviews of prediction modelling studies	Karel Moons, Lotty Hooft and Hans Reitsma	**Day 1 - 23 September, Tuesday: 13.30 to 15.00**
W36. Individual Participant Data (IPD) Meta-analysis of prediction modelling studies	Thomas Debray, Hans Reitsma and Karel Moons	**Day 1 - 23 September, Tuesday: 15.30 to 17.00**
W60. PROBAST: Introduction to a new risk of bias tool for prediction modelling studies	Robert Wolff, Penny Whiting and Karel Moons	**Day 3 - 25 September, Thursday: 13.30 to 15.00**
W73. Systematic reviews of prognostic studies: a meta-analytical approach	Thomas Debray and Karel Moons	**Day 4 - 26 September, Friday: 13.30 to 15.00**

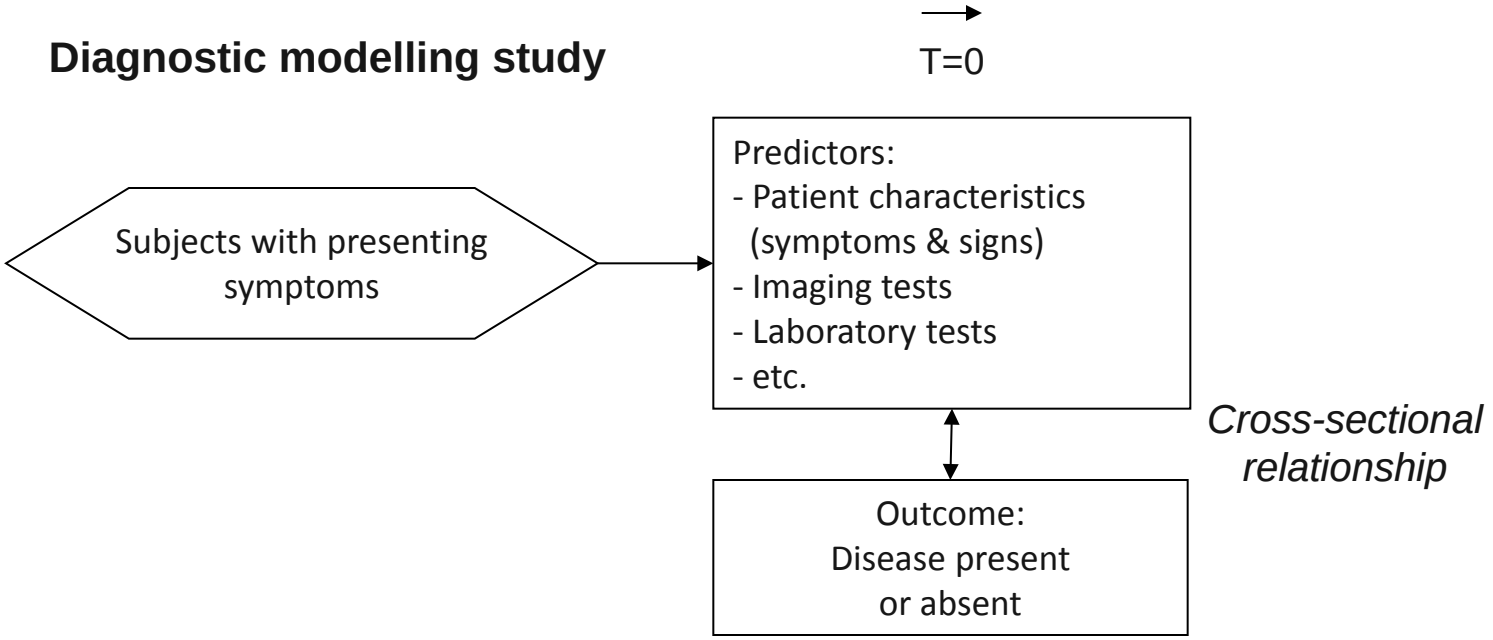
Prediction

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

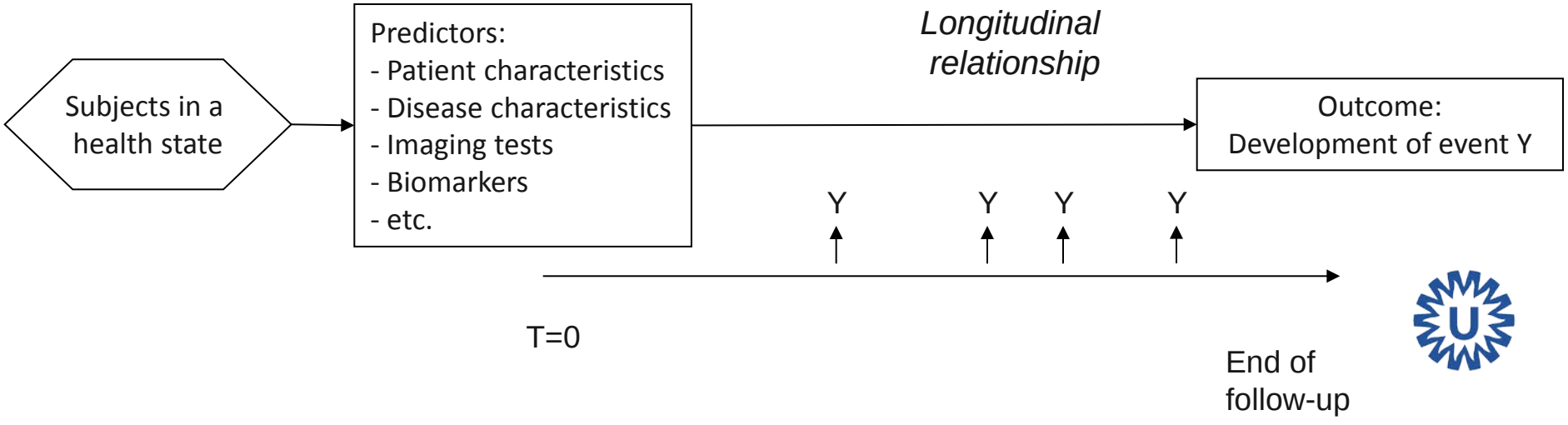
- Turn available information (predictors) into a statement about the probability:
  … diagnosis
  … prognosis

What is the big difference between diagnostic and prognostic ‘prediction’?

Diagnostic modelling study



Prognostic modelling study



Prognosis BMJ series 2009 (Altman, Moons, Royston, Vergouwe)

- **Prognosis**: Probable course or prediction of specific outcome of people with certain health condition
  - Not necessarily sick people

- Prognosis studies: Aim to understand the course and determinants of outcome in people with certain health condition

- Use of prognostic information:
  - To inform people/patients
  - Identify target groups for intervention/treatment
  - To select individuals for RCTs

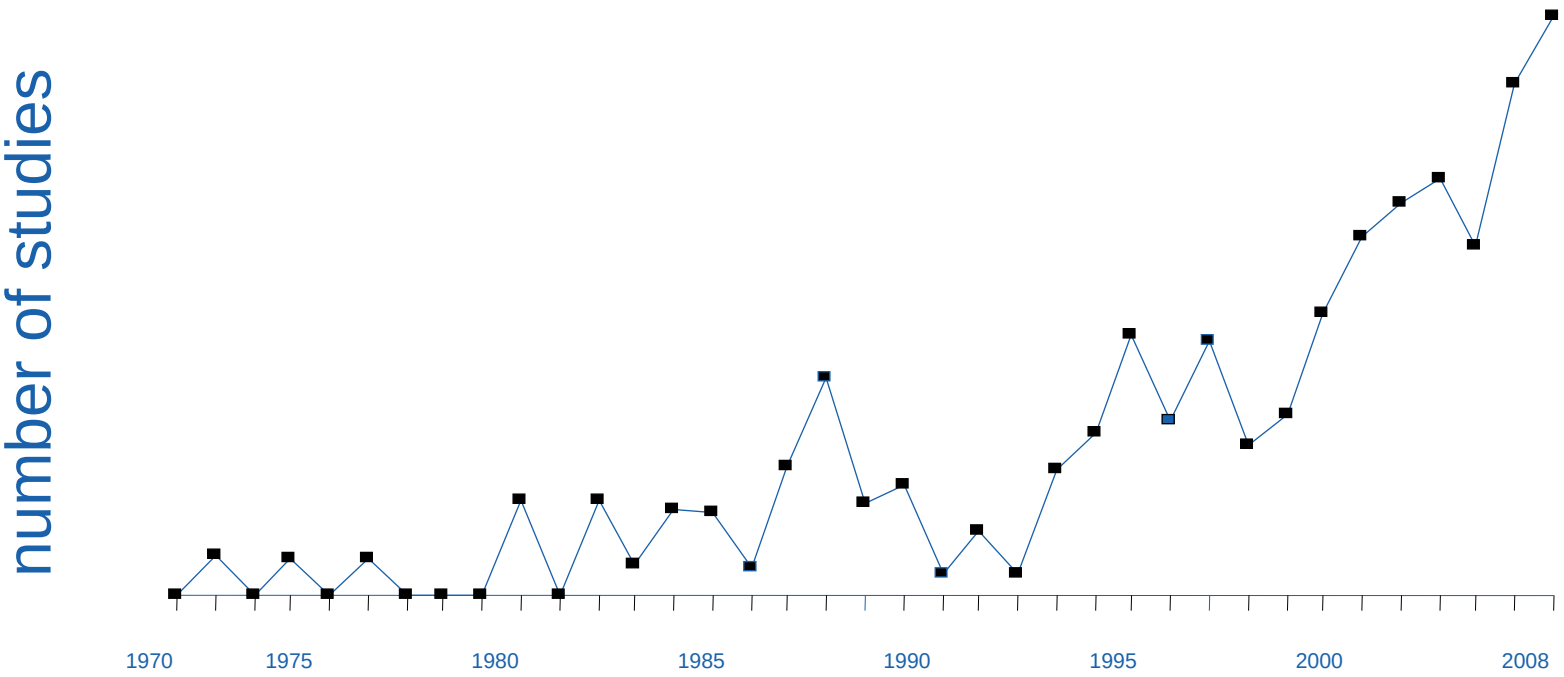
Three main types of prognosis studies
PROGRESS series 2013: BMJ and Plos Med

- Average/overall prognosis: 'What is the most likely course (outcome) of people with this health condition?'
- Prognostic factors: 'What factors are associated with that outcome?
- Prognostic (prediction) models: 'Are there risk groups who are likely to have different outcomes?'

Focus this workshop: MA of prediction model studies

BOTH: PROGNOSTIC AND DIAGNOSTIC

Why focus on prediction models? (Steyerberg 2009)

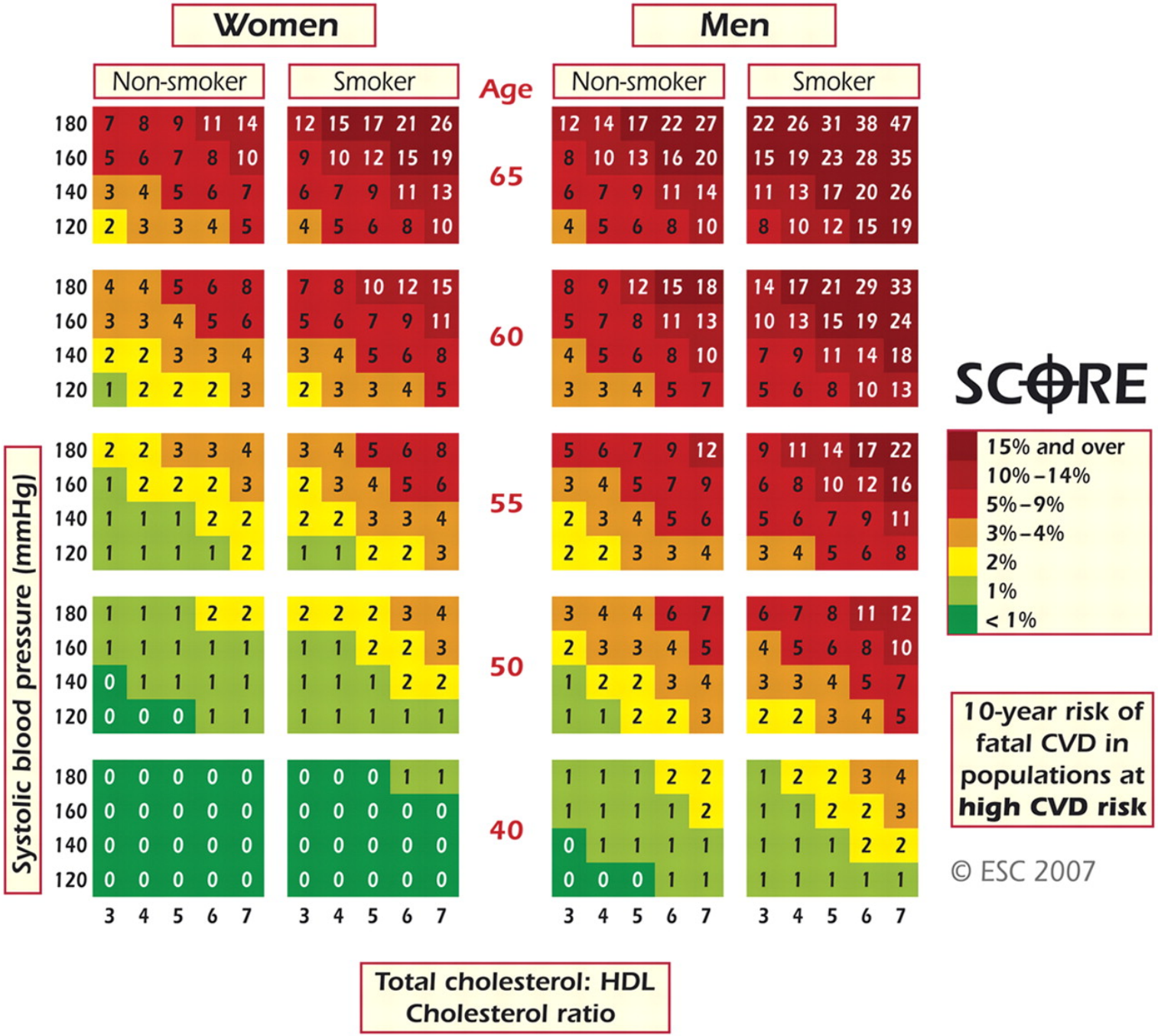


Apgar score in neonates (JAMA 1958)

Table 9–1. Apgar scoring.

Signs	0	1	2
Heartbeat per minute	Absent	Slow (<100)	Over 100
Respiratory effort	Absent	Slow, irregular	Good, crying
Muscle tone	Limp	Some flexion of *extremities*	Active *motion*
Reflex irritability	No response	Grimace	Cry or cough
Color	Blue or pale	Body pink, extremities blue	Completely pink

Σ = Apgar score (0-10)


Women
Men
Non-smoker
Smoker
Non-smoker
Smoker
Age
65
60
55
50
40
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
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
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
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
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
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

Your Disease Risk

THE SOURCE ON PREVENTION

my results: No Results Yet

Tome el cuestionario en Español

Cancer
Diabetes
Heart disease
Osteoporosis
Stroke

8 ways to prevent disease

What is...?
Prevention
Risk
A Screening Test

How to...
Estimate Risk

Community Action

Disclaimer
Privacy Policy
About This Site

Welcome to *Your Disease Risk*, the source on prevention. Here, you can find out your risk of developing five of the most important diseases in the United States and get personalized tips for preventing them.

Developed over the past ten years by world-renowned experts, *Your Disease Risk* collects the latest scientific evidence on disease risk factors into one easy-to-use tool.

To get started, choose one of the diseases below.

What is your risk?

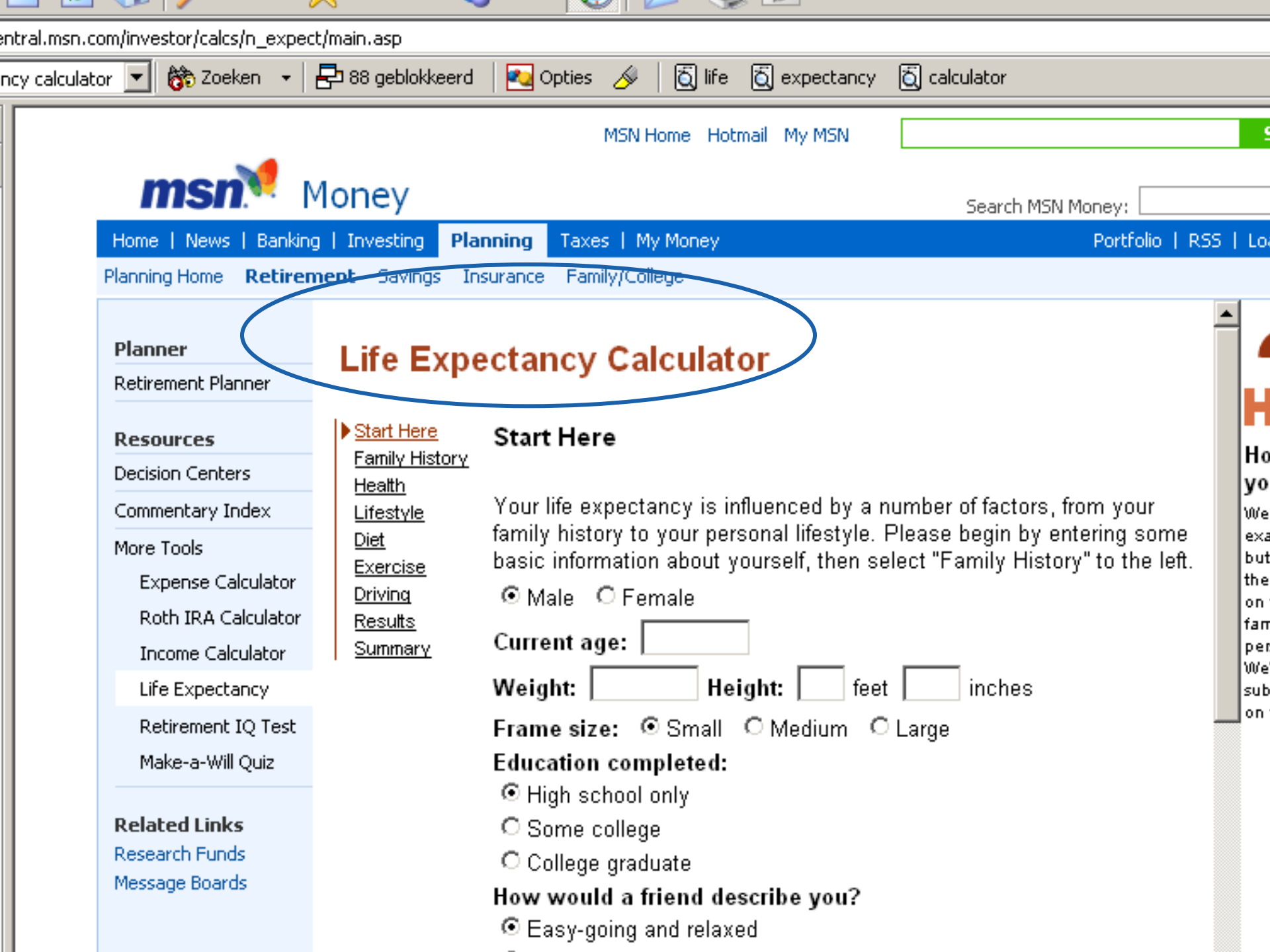
Cancer: There's much more to it than just smoking and lung cancer. — What's your cancer risk?

Diabetes: Over 18 million in the U.S. suffer from it. Take steps now to lower your risk. — What's your diabetes risk?

Heart disease: The #1 killer in the U.S. is also one of the most preventable. — What's your heart disease risk?

Osteoporosis: Calcium isn't the only way (or even the best way) to protect yourself. — What's your osteoporosis risk?

Stroke: Most cases of this feared disease can be avoided by lifestyle changes. — What's your stroke risk?


entral.msn.com/investor/calcs/n_expect/main.asp
ncy calculator
Zoeken
88 geblokkeerd
Opties
life
expectancy
calculator
MSN Home
Hotmail
My MSN
msn Money
Search MSN Money:
Home | News | Banking | Investing | Planning | Taxes | My Money
Portfolio | RSS | Lo
Planning Home
Retirement
Savings
Insurance
Family/College
Planner
Retirement Planner
Resources
Decision Centers
Commentary Index
More Tools
Expense Calculator
Roth IRA Calculator
Income Calculator
Life Expectancy
Retirement IQ Test
Make-a-Will Quiz
Related Links
Research Funds
Message Boards
Life Expectancy Calculator
Start Here
Family History
Health
Lifestyle
Diet
Exercise
Driving
Results
Summary
Start Here
Your life expectancy is influenced by a number of factors, from your family history to your personal lifestyle. Please begin by entering some basic information about yourself, then select "Family History" to the left.
Male
Female
Current age:
Weight:
Height:
feet
inches
Frame size:
Small
Medium
Large
Education completed:
High school only
Some college
College graduate
How would a friend describe you?
Easy-going and relaxed

Four phases of Prediction Modelling

BMJ series 2009 (Altman, Moons, Royston, Vergouwe)

1. Developing a prediction model
2. Validate the model in other subjects
3. Update existing model to local situation
4. Quantify model's impact on doctor's decision making and patient outcome (cost-effectiveness)

What is big difference between 4 versus 1-3?

Focus on 1-3

Overview Cochrane Prognostic Methods Group (PMG) Workshops

PMG Workshop	Facilitators	When?
W72. Systematic reviews of prediction modelling studies	Karel Moons, Lotty Hooft and Hans Reitsma	**Day 1 - 23 September, Tuesday: 13.30 to 15.00**
W36. Individual Participant Data (IPD) Meta-analysis of prediction modelling studies	Thomas Debray, Hans Reitsma and Karel Moons	**Day 1 - 23 September, Tuesday: 15.30 to 17.00**
W60. PROBAST: Introduction to a new risk of bias tool for prediction modelling studies	Robert Wolff, Penny Whiting and Karel Moons	**Day 3 - 25 September, Thursday: 13.30 to 15.00**
W73. Systematic reviews of prognostic studies: a meta-analytical approach	Thomas Debray and Karel Moons	**Day 4 - 26 September, Friday: 13.30 to 15.00**

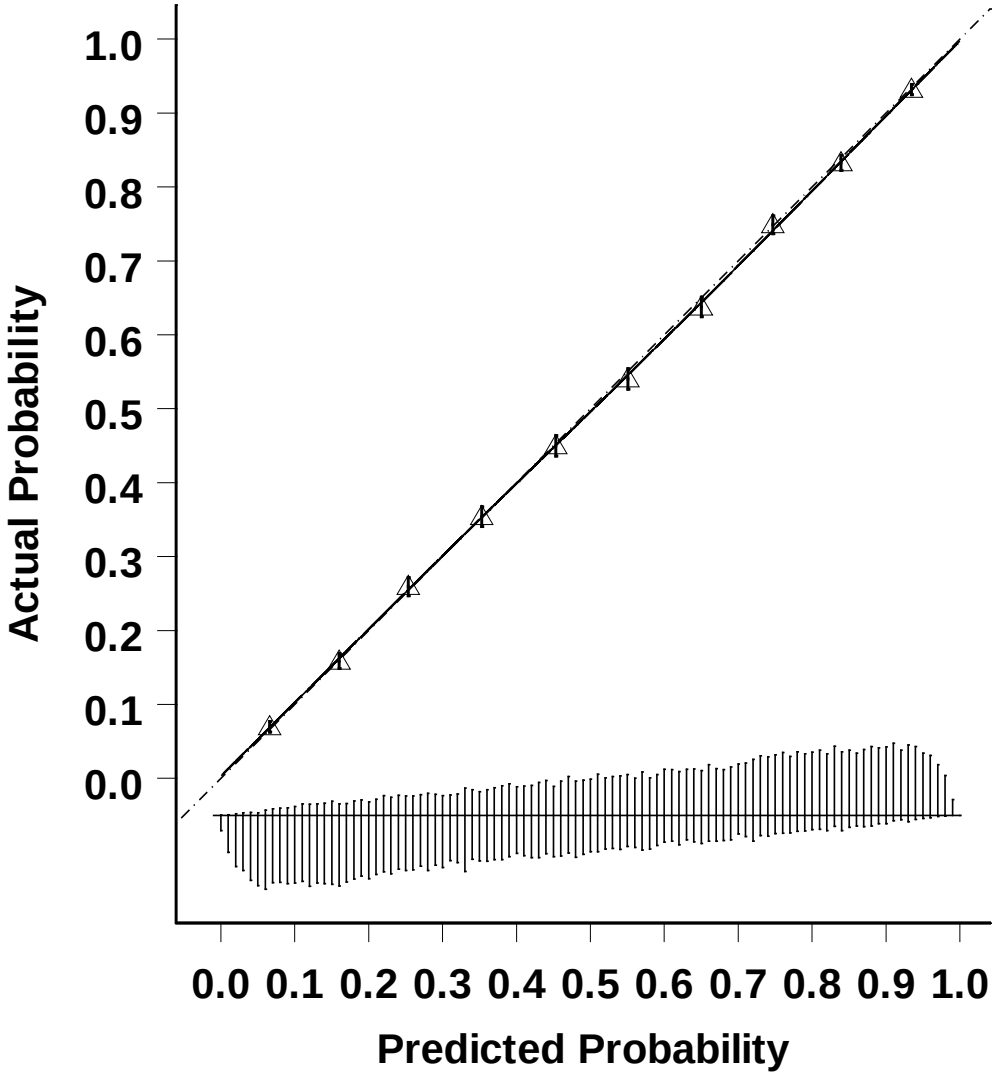
Prediction model performance measures

- Calibration plot
  (for specific time point in case of survival models)

- Discrimination
  - C-statistic (ROC area for logistic regression)

- (Re)classification → requires probability thresholds
  - Two by to tables → diagnostic test accuracy MA procedures
  - NRI → in case of model comparison / addition of new predictor → requires thresholds → beyond this workshop

Calibration plot – good model?

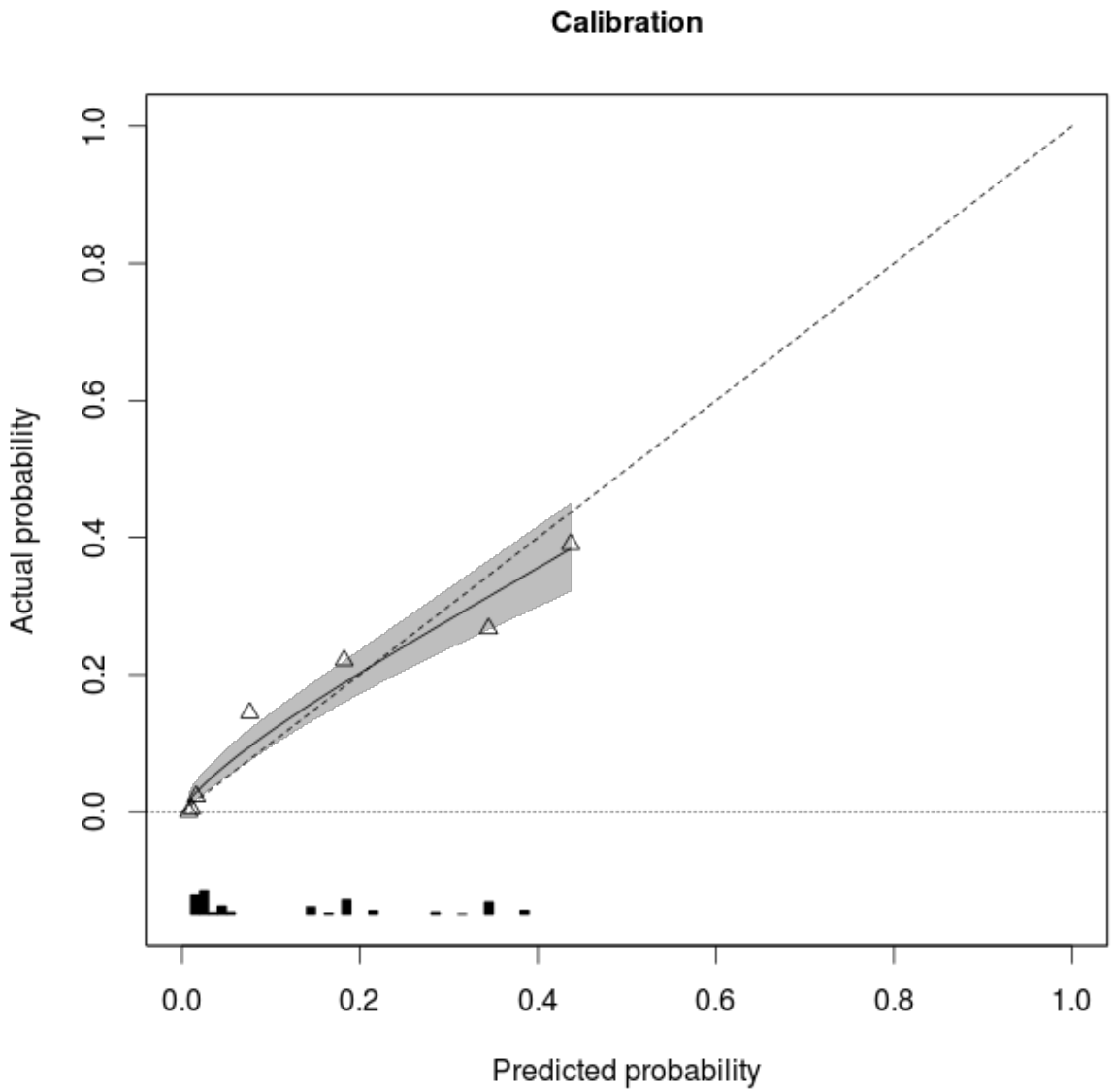


Ideal calibration

O:E = 1

Slope = 1

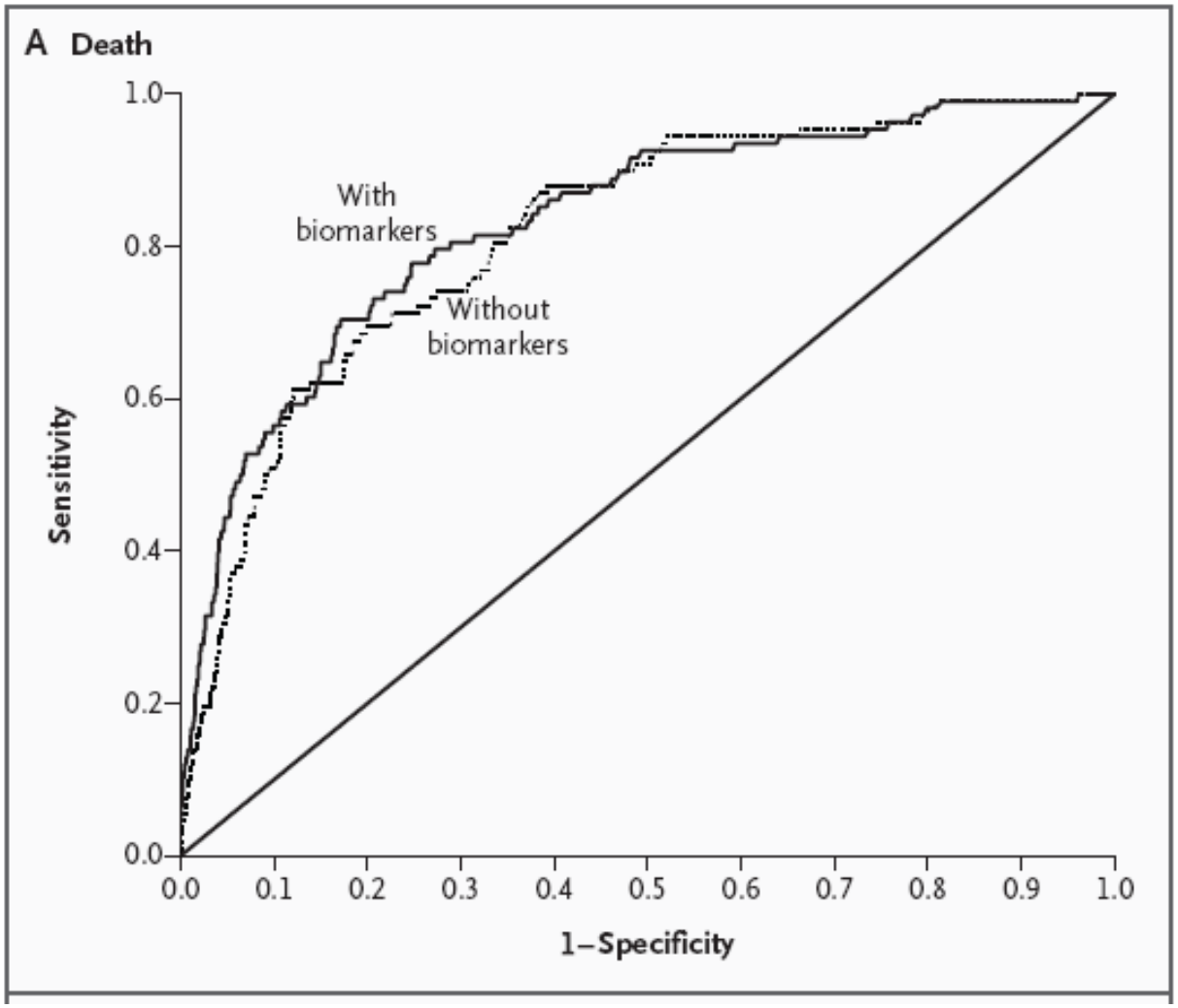
Calibration plot – good model?



O:E = 1
Slope = 0.79

Sub-obtimal slope because curve does not follow reference line

Model to predict cardiovascular outcomes – added value biomarkers?



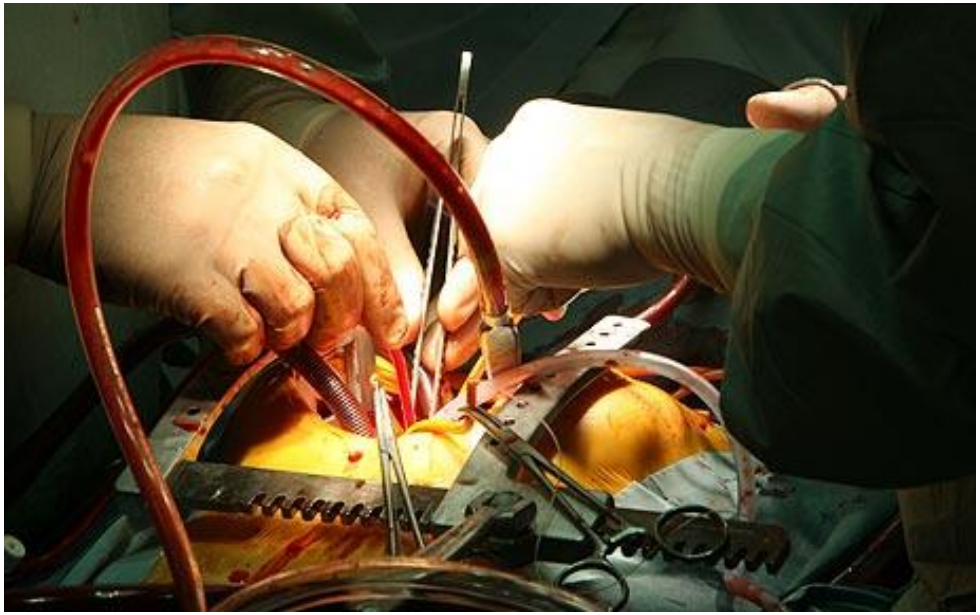
AUC 0.76

AUC 0.77

Wang TJ, et al. NEJM

Workshop example: predicting mortality after cardiac surgery

- Cardiac surgery in high-risk population
- Need for risk stratification
- Need for quality of care assessment (benchmarking)
- Establish risk profile of cardiac surgical patients using **multivariable prediction models**



Predicting mortality after cardiac surgery

ELSEVIER

European Journal of Cardio-thoracic Surgery 15 (1999) 816–823

EUROPEAN JOURNAL OF CARDIO-THORACIC SURGERY

Risk factors and outcome in European cardiac surgery: analysis of the EuroSCORE multinational database of 19030 patients☆

F. Roques*, S.A.M. Nashef, P. Michel, E. Gauducheau, C. de Vincentiis, E. Baudet, J. Cortina, M. David, A. Faichney, F. Gabrielle, E. Gams, A. Harjula, M.T. Jones, P. Pinna Pintor, R. Salamon, L. Thulin

Service de chirurgie cardiovasculaire, CHU de Fort de France, 97200 Martinique, France

Received 22 September 1998; received in revised form 8 March 1999; accepted 11 March 1999

External validation of EuroSCORE

- External validation in patients undergoing off-pump coronary artery bypass grafting
- Over-estimation of in-hospital mortality
  - Predicted mortality rate: 5.0%
  - Observed mortality rate: 1.3%
- Poor calibration
  - Hosmer-Lemeshow test (p=0.04)
- Adequate discrimination
  - AUC=0.71

Ref: Youn et al. Can the EuroSCORE Predict the Early and Mid-Term Mortality After Off-Pump Coronary Artery Bypass Grafting? Ann Thorac Surg 2007

External validation

- Is the model reliable?
- Does the model generalize well across populations?
- Does the model require improvements/changes?
- Or, should we rather develop a new model from scratch?

RESEARCH ARTICLE Open Access

External validation of multivariable prediction models: a systematic review of methodological conduct and reporting

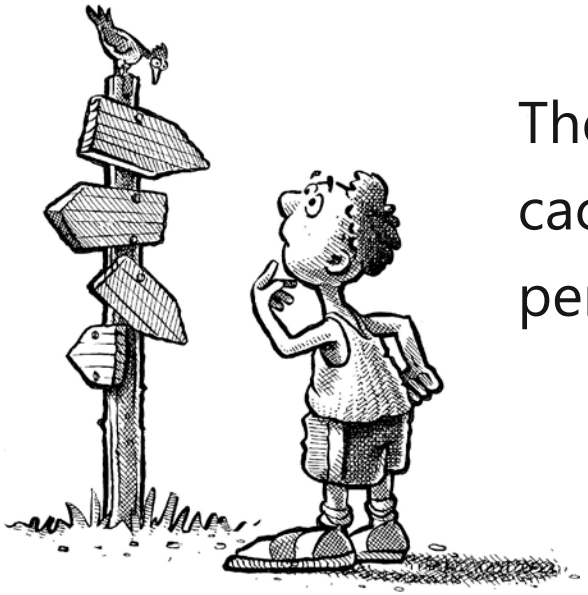
Gary S Collins[1*], Joris A de Groot[2], Susan Dutton[1], Omar Omar[1], Milensu Shanyinde[1], Abdelouahid Tajar[1], Merryn Voysey[1], Rose Wharton[1], Ly-Mee Yu[1], Karel G Moons[2] and Douglas G Altman[1]

External validation

- Assess model performance in a new sample
- Compare predicted probabilities to observed outcomes
- Discrimination and calibration

Caveats in prediction modeling research

- Most models are never validated
- Model redevelopment versus model updating
- Prior knowledge not optimally used
- How to choose between competing models?
- Incompatibility and confusion



The user must typically choose between a cacophony of existing models for which performance may be obscure

Numerous models for same target population + outcomes

- Reflex: develop 'own new' model from their study data
  → certainly if poor validation of existing model
  - >150 models alike Framingham, SCOPE, Qrisk
  - >100 models for brain trauma patients
  - >60 models for breast cancer prognosis
  - > 100 diabetes type 2 models

- Understandable:
  - We finally learned the 'tricks' to develop models (in standard software)
  - 'Own' model makes you famous
    (Apgar; Goldman; Gail; Wells)
  - Validation is only to support (citation index of) others

Numerous models for same target population + outcomes

Ref: Reilly Ann Int Med 2009; Moons BMJ 2009 + Heart 2012;Steyerberg+Moons 2013

- We need more SRs + MA of prediction models
- Every model development or validation study should be preceded by SR of existing models

BMJ

BMJ 2012;344:e3186 doi: 10.1136/bmj.e3186 (Published 24 May 2012)

EDITORIALS

Comparing risk prediction models

Should be routine when deriving a new model for the same purpose

Gary S Collins *senior medical statistician*[1], Karel G M Moons *professor of clinical epidemiology*[2]

Meta-analysis of prediction models

Two types

1. **In case no own (validation) IPD set – aggregate data only: 2 cases**
    1. MA of a specific prediction model across multiple 'model-validation-studies'
    2. MA of a specific predictor when added to a specific model across multiple 'added-value-studies'

2. **In case own (validation) IPD set – combination of aggregate and IPD**

Ad. Meta-analysis of prediction models
In case no own (validation) IPD set

1. MA of a specific prediction model across multiple 'model-validation-studies'
2. MA of a specific predictor/marker/test when added to a specific model across multiple 'added-value-studies'

Type 1. SR and MA of specific model across multiple model-validation-studies

- Systematic review of model performance
- Pool measures of discrimination and calibration
- Investigate heterogeneity in model performance

Example aggregate meta-analysis of a specific prediction model → the EuroSCORE model

44 validation studies with information on:

- Model discrimination (AUC)
- Model calibration (O:E ratio)

European Journal of Cardio-Thoracic Surgery 41 (2012) 746–754
doi:10.1093/ejcts/ezr285 Advance Access publication 26 January 2012

REVIEW

Performance of the original EuroSCORE

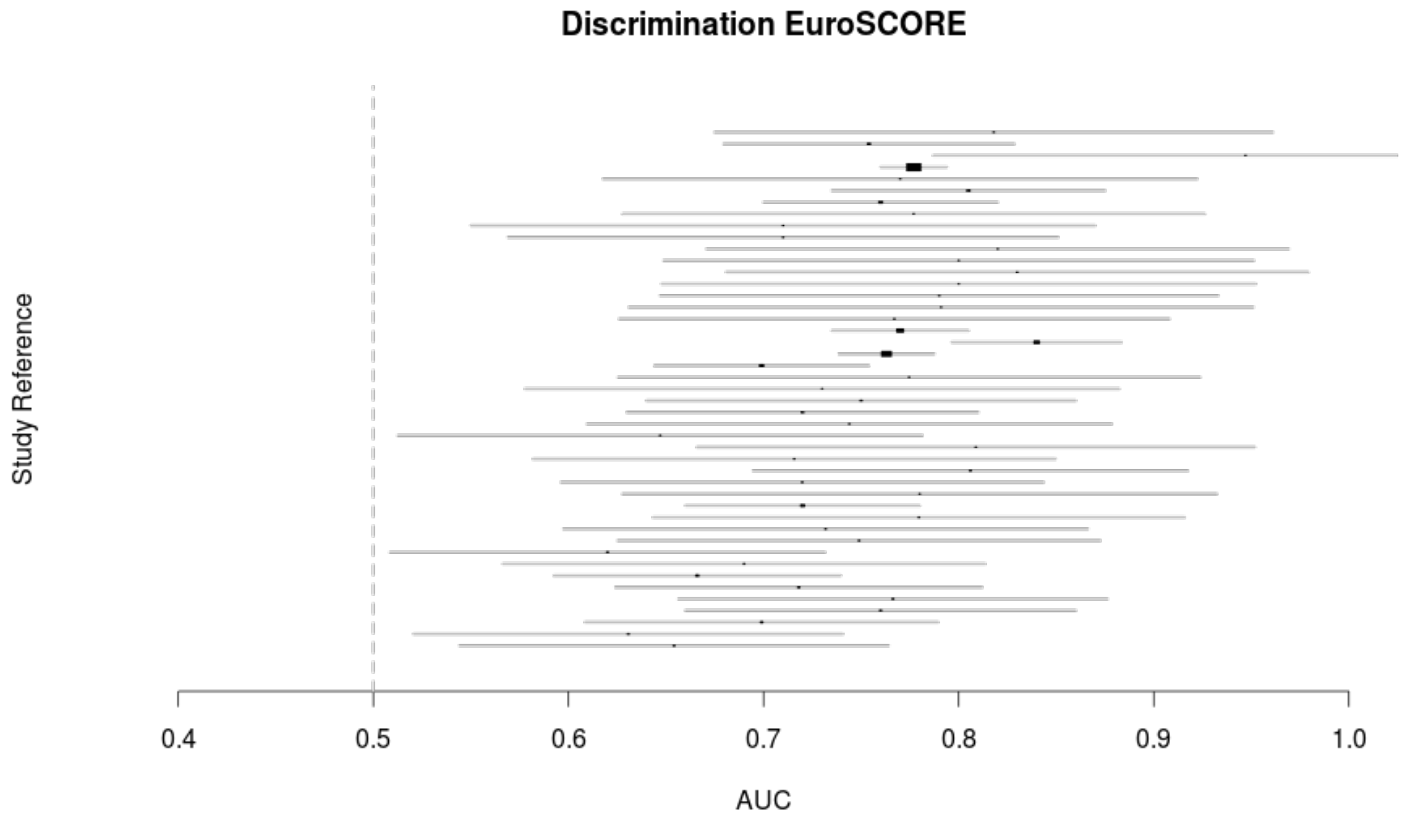
Sabrina Siregar[a,b,*], Rolf H.H. Groenwold[b], Frederiek de Heer[a], Michiel L. Bots[b], Yolanda van der Graaf[b] and Lex A. van Herwerden[a]

[a] Department of Cardio-Thoracic Surgery, Heart and Lungs Division, University Medical Center Utrecht, Utrecht, The Netherlands
[b] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

* Corresponding author. Department of Cardio-Thoracic Surgery, Division Heart and Lungs, University Medical Center Utrecht, Heidelberglaan 100, E03.511, PO Box 85500, 3508 GA Utrecht, The Netherlands. Tel: +31-88-7556179; fax: +31-88-7555058; e-mail: s.siregar@umcutrecht.nl (S. Siregar).

Received 24 June 2011; received in revised form 22 November 2011; accepted 24 November 2011

Example 1: Meta-analysis of the EuroSCORE model on aggregate level

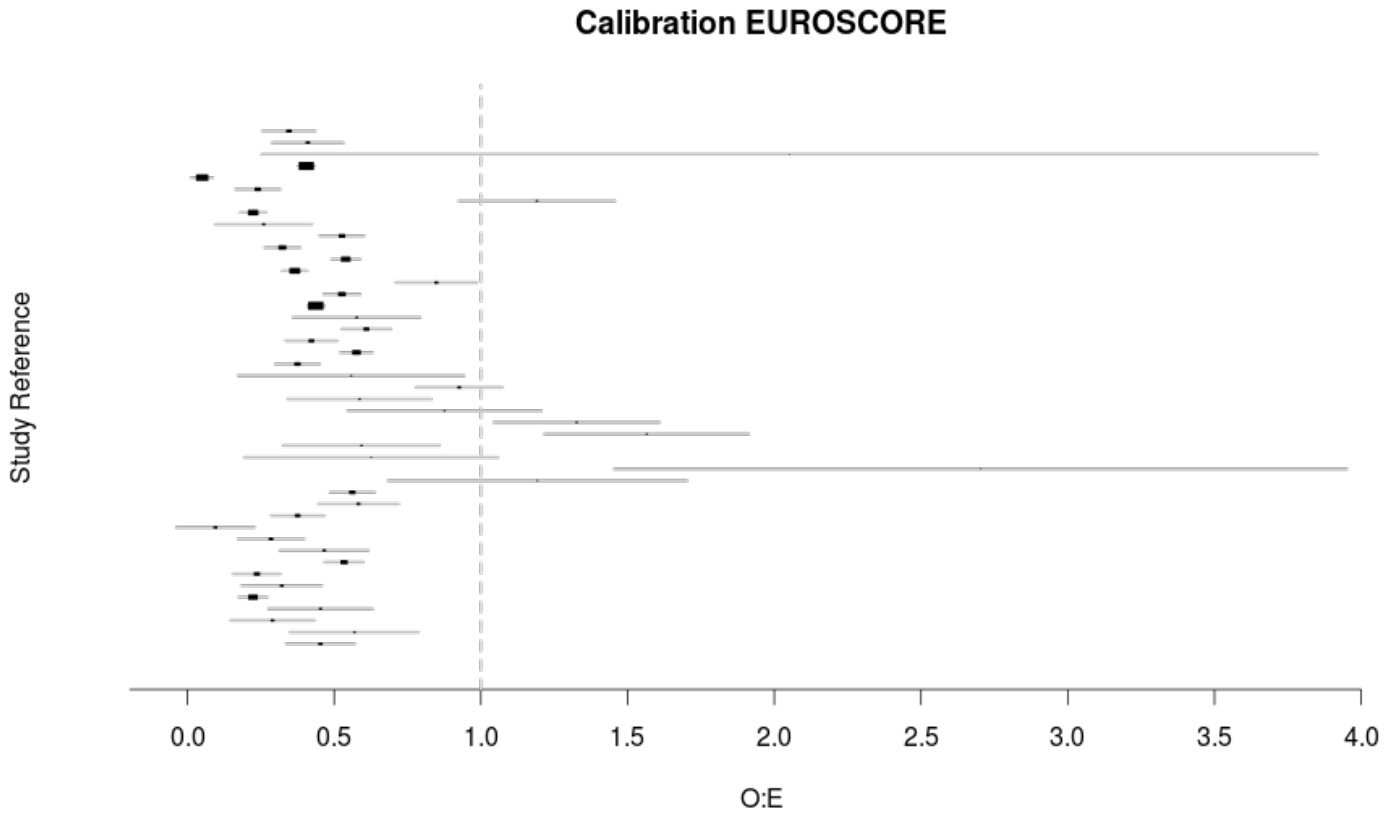


Meta-analysis of EuroSCORE performance

Pooled estimates of discrimination EuroSCORE

- Pooled estimate: **0.7516**
- Standard error: 0.0089
- Std. dev. between studies: 0.0318
- 95% confidence interval: 0.73 – 0.77
- 95% prediction interval: 0.69 – 0.82
- I2 statistic: 32.3%
- Cochran Q-test for heterogeneity: p-value = 0.0216

Meta-analysis of EuroSCORE performance



Meta-analysis of EuroSCORE performance

Pooled estimates of calibration EuroSCORE

- Pooled estimate: **0.5205**
- Standard error: 0.0438
- Std. dev. between studies: 0.2748
- 95% confidence interval: 0.43 – 0.61
- 95% prediction interval: 0.00 – 1.07
- I2 statistic: 95.3%
- Cochran Q-test for heterogeneity: p-value = 0.0000

Meta-analysis of EuroSCORE performance

Heterogeneity across validation studies

- Type of study: prospective vs. retrospective
- Surgical categories
  - Cardiac surgery
  - Isolated coronary artery bypass grafting (CABG)
  - Isolated valve and mixed CABG
  - Valve
- Mortality
  - 30-day mortality
  - In-hospital mortality
  - Operative mortality

Meta-analysis of EuroSCORE performance

Pooled estimates of discrimination EuroSCORE

- Surgical categories:
  - CABG and valve: **0.70** (95% PI: 0.64 – 0.75)
  - Cardiac surgery: **0.78** (95% PI: 0.73 – 0.82)
  - Isolated CABG: **0.78** (95% PI: 0.73 – 0.83)
  - Isolated valve: **0.74** (95% PI: 0.69 – 0.79)
- I2 statistic: 1%
- Cochran Q-test for heterogeneity: p-value = 0.5299

Meta-analysis of EuroSCORE performance

Pooled estimates of calibration EuroSCORE

- Surgical categories:
  - CABG and valve: **0.35** (95% PI: 0.00 – 0.80)
  - Cardiac surgery: **0.53** (95% PI: 0.08 – 0.97)
  - Isolated CABG: **0.39** (95% PI: 0.00 – 0.84)
  - Isolated valve: **0.81** (95% PI: 0.36 – 1.27)
- I2 statistic: 93.4%
- Cochran Q-test for heterogeneity: p-value = 0.0000

Recall Meta-analysis of prediction models In case no own (validation) IPD set

1. MA of a specific prediction model across multiple 'model-validation-studies'
2. MA of a specific predictor/marker/test when added to a specific model across multiple 'added-value-studies'

Type 2. SR and MA of specific predictor when added to a specific model across multiple 'added-value-studies'

- Systematic review of added value in discrimination of the predictor

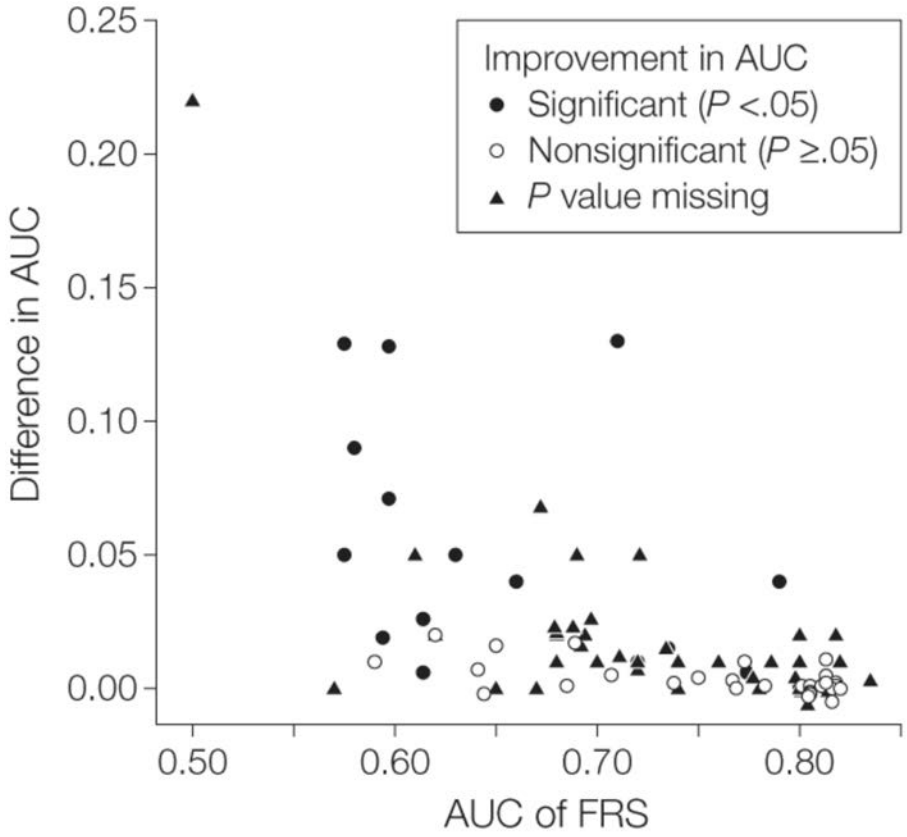
- Investigate heterogeneity in this

Example: Added value of new (bio)markers in Framingham Risk Score

- Systematic review of studies that …
  - … evaluated various candidate prognostic factors in their ability to improve prediction of coronary hearth disease or other outcomes
  - … beyond what the Framingham risk score (FRS) can achieve
- Reported test statistics:
  - AUC of FRS alone
  - AUC of FRS with additional predictor(s)
  - Δ AUC

Example: Added value of new (bio)markers in Framingham Risk Score

Possible extension: pooling of Δ AUC statistic using same methods as for pooling AUC of a specific model (see above) example 1 !



Meta-analysis of prediction models

Two types

1. In case no own (validation) IPD set – aggregate data only

2. **In case own (validation) IPD set – combination of aggregate and IPD**
   - **a. Models with similar predictors**
   - **b. Models with different predictors**

Meta-analysis of prediction models in case of own IPD set

Models with similar predictors

- Meta-analysis (**therapeutic research**)
  - Synthesize evidence from multiple trials
  - Obtain a summary estimate of treatment effect
  - Facilitate detailed analyses of effect modification

- Meta-analysis (**prediction research**)
  - Synthesize evidence on prognostic factors
  - Summarize model performance
  - Aggregate literature models into a meta-model

Meta-analysis of prediction models in case of own IPD set

Models with similar predictors

Statistics
in Medicine

Research Article

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

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

Article first published online: 26 JUN 2012

DOI: 10.1002/sim.5412

CrossMark click for updates

Issue

Statistics in Medicine
Volume 31, Issue 23, pages 2697–2712, 15 October 2012

SEARCH

In this issue

Advanced > Saved Searches >

ARTICLE TOOLS

Get PDF (179K)
Save to My Profile
E-mail Link to this Article
Export Citation for this Article
Get Citation Alerts
Request Permissions

Share

Additional Information (Show All)

How to Cite | Author Information | Publication History | Funding Information

Meta-analysis of prediction models in case of own IPD set

Models with similar predictors

- Identify common predictors
  - Restore missing coefficients and standard errors where necessary (imputation)
- Pooling of predictor effects
  - Calculate weighted average of regression coefficients
  - Account for differences in precision
  - Account for heterogeneity across studies
- Meta-model for average or specific study population
  - Relevance of literature versus validation sample
  - Adjust intercept term to local circumstances

Meta-analysis of prediction models in case of own IPD set

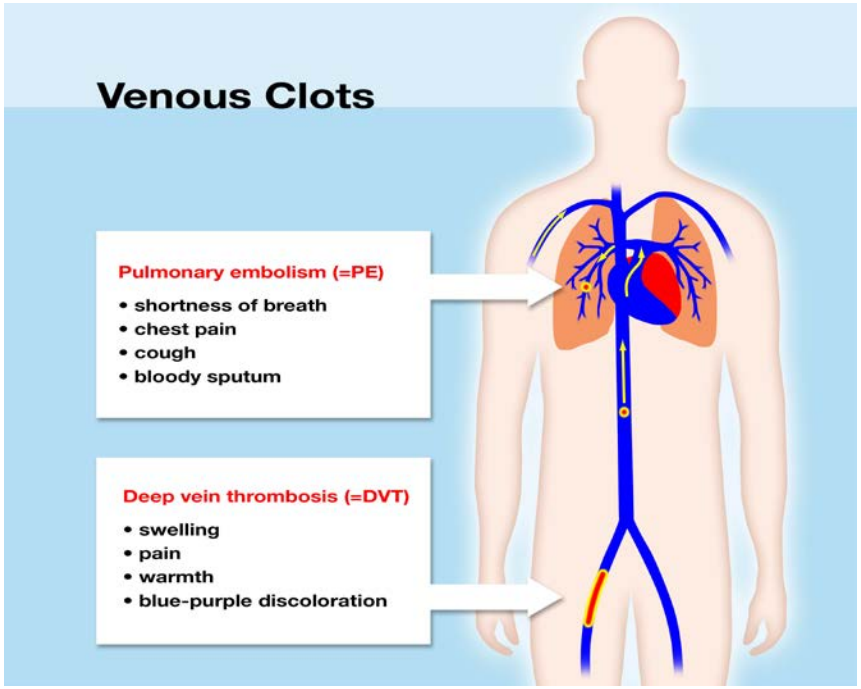
Models with similar predictors

- Univariate meta-analysis
  - Pool predictor effects separately
- Multivariate meta-analysis
  - Simultaneous pooling of all predictor effects
- Multivariate meta-analysis + Bayesian inference
  - Pooled predictor effects from the literature are used as prior information for the predictor effects in the validation sample

Meta-analysis of prediction models for diagnosing deep vein thrombosis

Diagnosis of **deep vein thrombosis** (DVT)

- Blood clot that forms in a vein in the body (lower leg/thigh)
- If blood clot breaks off -> blood stream -> lungs -> blockage
- Pulmonary embolism, preventing oxygenation of blood
- Potentially causing death



Meta-analysis of prediction models for diagnosing deep vein thrombosis

- Limited value of signs and symptoms (primary care)
- Most patients referred to secondary care
- Burden on patients and health care budgets

Need for developing **multivariable prediction models**

- Predict presence of DVT in suspected patients
  - Patient history and physical examination
  - Biomarker test results: D-dimer test
- Primary care versus secondary care

Meta-analysis of prediction models for diagnosing deep vein thrombosis

WELLS Score (DVT) (*)

• Active cancer (treatment ongoing or within previous 6 months, or palliative treatment)	1
• Paralysis, paresis, or recent plaster immobilization of the lower extremities	1
• Recently bedridden for 3 days or more, or major surgery within the previous 12 weeks requiring general or regional anesthesia	1
• Localized tenderness along the distribution of the deep venous system	1
• Entire leg swollen	1
• Calf swelling > 3 cm compared to asymptomatic leg (measuring 10 cm below tibial tuberosity)	1
• Pitting edema confined to the symptomatic leg	1
• Nonvaricose collateral superficial veins	1
• Previously documented DVT	1
• Alternative diagnosis at least as likely as DVT	-2

≤ 0 : LOW pretest probability
1 or 2 : MODERATE pretest probability
≥ 3 : HIGH pretest probability

Wells PS, et al. N Engl J Med 2003; 349: 1227-35
Anderson DR, et al. J Thromb Haemost 2003; 1: 645-51

(*) In patients with symptoms in both legs, the more symptomatic leg is used

Meta-analysis of prediction models for diagnosing deep vein thrombosis

Characteristics	Hamilton	Modified Wells
Plaster immobilization of lower limb	2	1
Active malignancy (within 6 months or current)	2	1
Strong clinical suspicion of deep venous thrombosis by the emergency physicians without other diagnostic possibilities	2	–
Bed rest (>3 days) or recent surgery (within 4 weeks)	1	1
Male sex	1	–
Calf circumference >3 cm on affected side (measured 10 cm below tibial tuberosity)	1	1
Erythema	1	–
Localized tenderness along the distribution of the deep venous system	–	1
Entire leg swollen	–	1
Pitting edema confined to the symptomatic leg	–	1
Collateral superficial veins (nonvaricose)	–	1
Previously documented deep vein thrombosis	–	1
Alternative diagnosis at least as likely as deep vein thrombosis	–	−2
Unlikely versus likely cutoff score	2 or less	1 or less

Meta-analysis of prediction models for diagnosing deep vein thrombosis

Diagnostic variables	Odds ratio	Regression coefficient*	p-value	Points for the rule
Male gender	1.80 (1.36 – 2.16)	0.59	<0.001	1
Oral contraceptive use	2.12 (1.32 – 3.35)	0.75	0.002	1
Presence of malignancy	1.52 (1.05 – 2.44)	0.42	0.082	1
Recent surgery	1.46 (1.02 – 2.09)	0.38	0.044	1
Absence of leg trauma	1.82 (1.25 – 2.66)	0.60	0.002	1
Vein distension	1.62 (1.19 – 2.20)	0.48	0.002	1
Calf difference ≥ 3 cm	3.10 (2.36 – 4.06)	1.13	<0.001	2
D-dimer abnormal	20.3 (8.25 – 49.9)	3.01	<0.001	6
Constant		–5.47		

DVT= deep vein thrombosis; *=natural logarithm of the odds ratio; D-dimer abnormal for VIDAS ≥ 500 ng/ml and Tinaquant ≥ 400 ng/ml. Probability of DVT as estimated by the final model =1/(1+exp-(-5·47 + 0·59*male gender + 0·75*OC use + 0·42*presence of malignancy + 0·38*recent surgery + 0·60*absence of leg trauma + 0·48*vein distension + 1·13*calf difference ≥ 3cm + 3·01*abnormal D-dimer)).

Meta-analysis of prediction models for diagnosing deep vein thrombosis

TABLEAU II

Analyse multivariée : modèle de régression logistique final prédisant la présence d'une thrombose veineuse profonde

Variable	p	odds ratio	coefficient
Immobilisation médicale dans le mois précédent (alitement > 48 h ou paralysie)	0,07	1,9 (1,0–3,7)	0,62
Contraception oestroprogestative	0,02	4,0 (1,2–12,9)	1,38
Antécédent personnel de MVTE	0,02	2,1 (1,1–4,0)	0,74
Cancer évolutif	<0,01	7,3 (2,4–22,1)	1,99
Diminution du ballant du mollet	0,01	2,3 (1,3–4,1)	0,83
Diagnostic alternatif au moins aussi probable	<0,01	0,1 (0,1–0,3)	−2,08

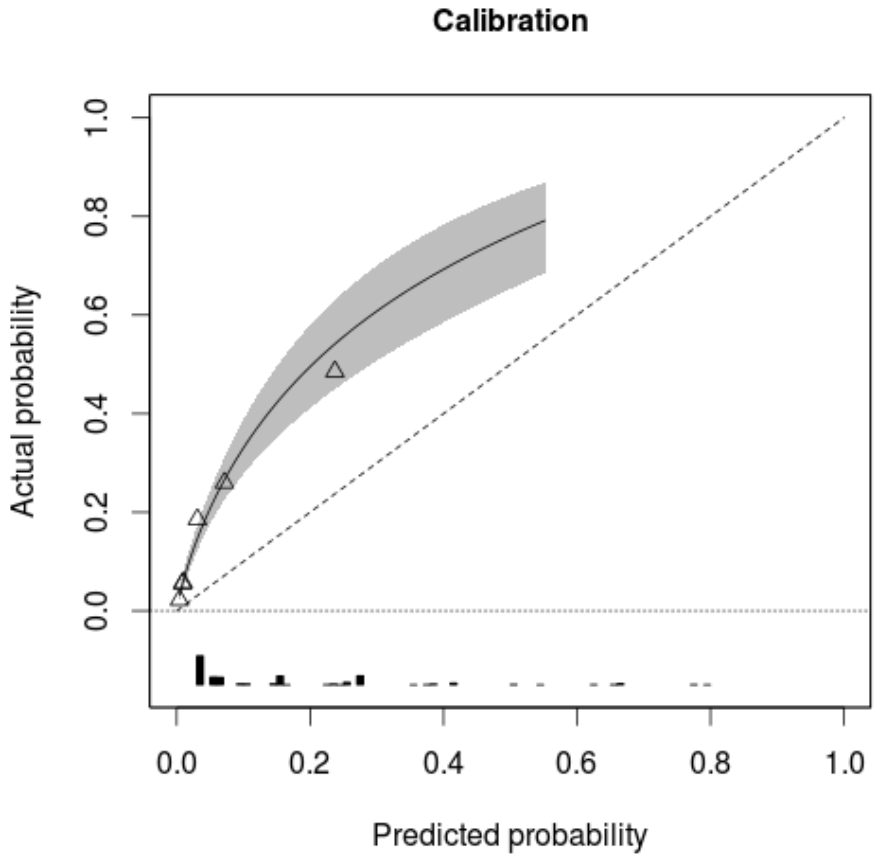
External validation of prediction models for diagnosing deep vein thrombosis

- Prospective management study
- 300 primary care practices in 3 regions of the Netherlands (Amsterdam, Maastricht, Utrecht)
- Outcome: incidence of symptomatic venous thromboembolism during 3-month follow-up
- 1028 patients with clinically suspected DVT
- 131 patients eventually diagnosed with DVT

External validation of prediction models for diagnosing deep vein thrombosis

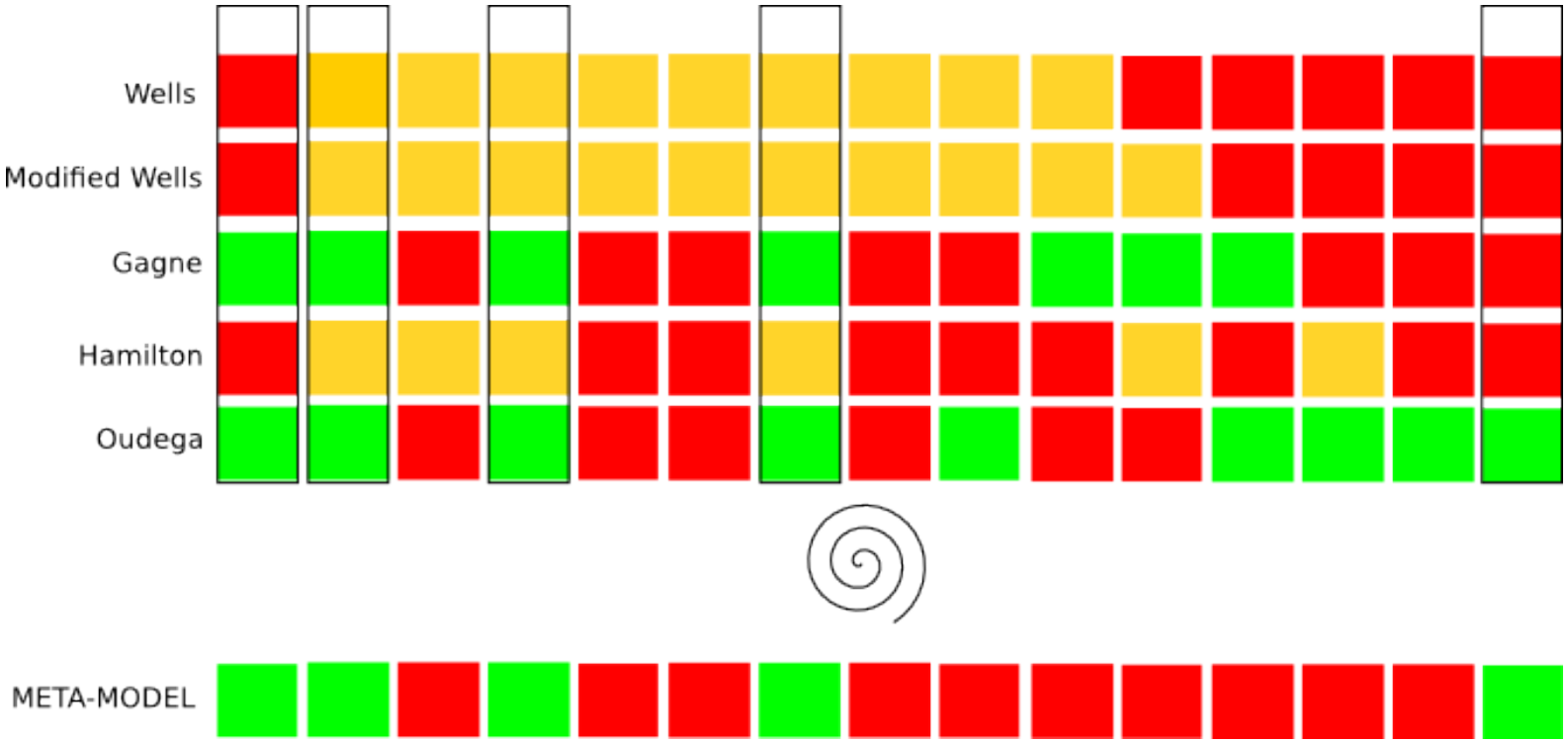
Gagne model

- AUC = 0.81
- O:E= 3.47
- Slope = 0.85

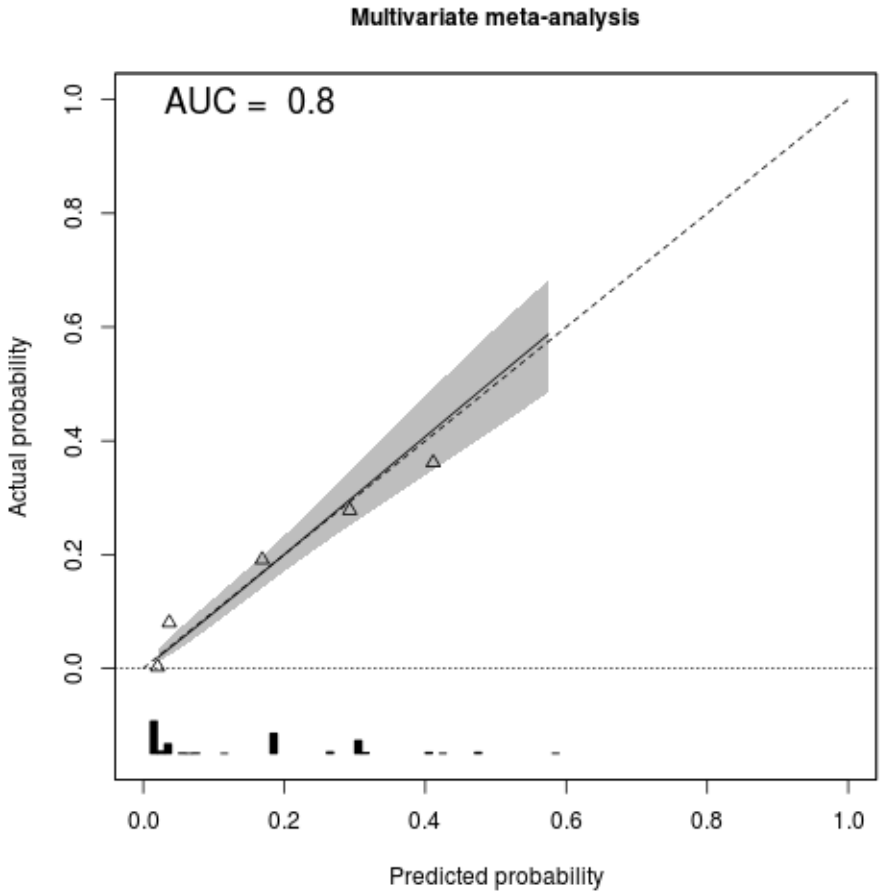


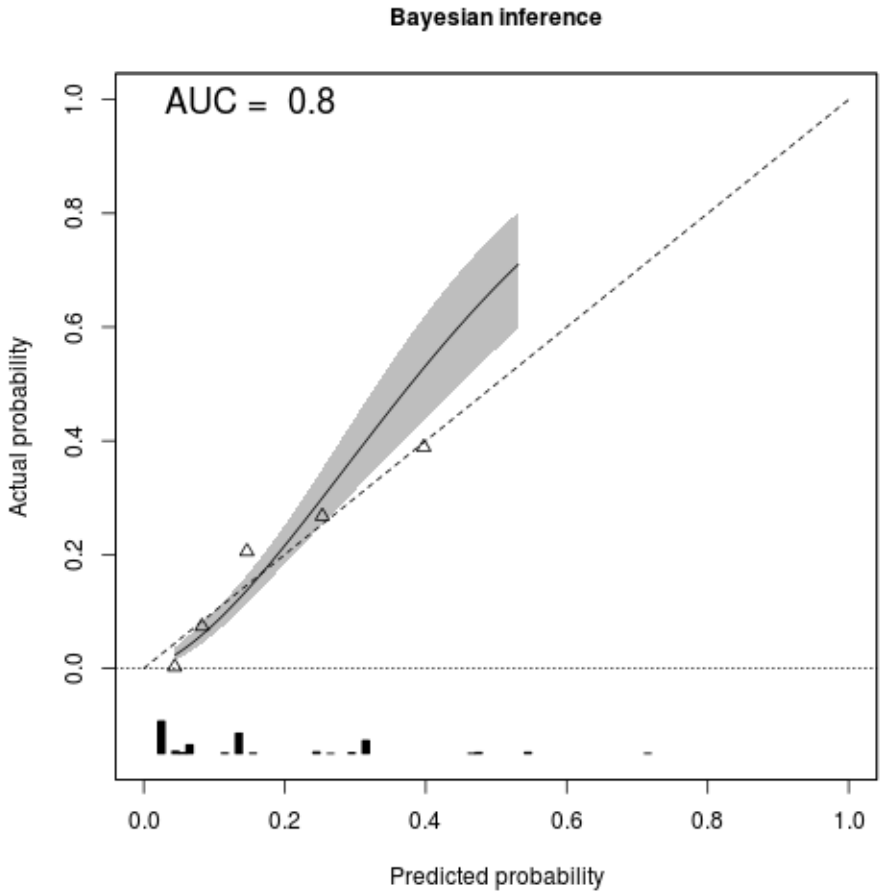
Meta-analysis of prediction models for diagnosing deep vein thrombosis

Focus on 4 common core predictors (+ intercept term)



Internal validation of meta-model for diagnosing deep vein thrombosis





External validation of meta-model for diagnosing deep vein thrombosis

- Primary care setting (N=791)
- Performance meta-models
  - AUC = 0.73 (MMA); 0.74 (Bayesian Inf.)
  - O:E = 0.822 (MMA); 0.904 (Bayesian Inf.)
  - Slope = 1.203 (MMA); 1.363 (Bayesian Inf.)
- Performance best literature model (Oudega)
  - AUC = 0.77
  - O:E = 0.723
  - Slope = 0.899

Meta-analysis of prediction models in case of own IPD set

Models with similar predictors

- Fewer predictors
- Slight decrease in model discrimination (as compared to best literature model)
- Adjusted for validation sample (baseline risk)

Implementation difficult when literature models differ much in terms of included predictors

Meta-analysis of prediction models

Two types

1. In case no own (validation) IPD set – aggregate data only

2. **In case own (validation) IPD set – combination of aggregate and IPD**
   - **a.** Models with similar predictors
   - **b. Models with different predictors**

Meta-analysis of prediction models in case of own IPD set

Models with different predictors

Statistics
in Medicine

Research Article

Meta-analysis and aggregation of multiple published prediction models

Thomas P.A. Debray[1,*], Hendrik Koffijberg[1], Daan Nieboer[2], Yvonne Vergouwe[2], Ewout W. Steyerberg[2] and Karel G.M. Moons[1]

Article first published online: 14 JAN 2014

DOI: 10.1002/sim.6080

Issue

Statistics in Medicine

Early View (Online Version of Record published before inclusion in an issue)

SEARCH

In this issue

Advanced > Saved Searches >

ARTICLE TOOLS

Get PDF (329K)

Save to My Profile

E-mail Link to this Article

Export Citation for this Article

Get Citation Alerts

Request Permissions

Share

Additional Information (Show All)

How to Cite | Author Information | Publication History

Meta-analysis of prediction models in case of own IPD set

Models with different predictors

Aims

- Avoid focus on similar predictors
- Improve performance over best literature model
- Adjust for between-study heterogeneity in baseline risk and predictor effects

Methods

- Model averaging
- Stacked regressions

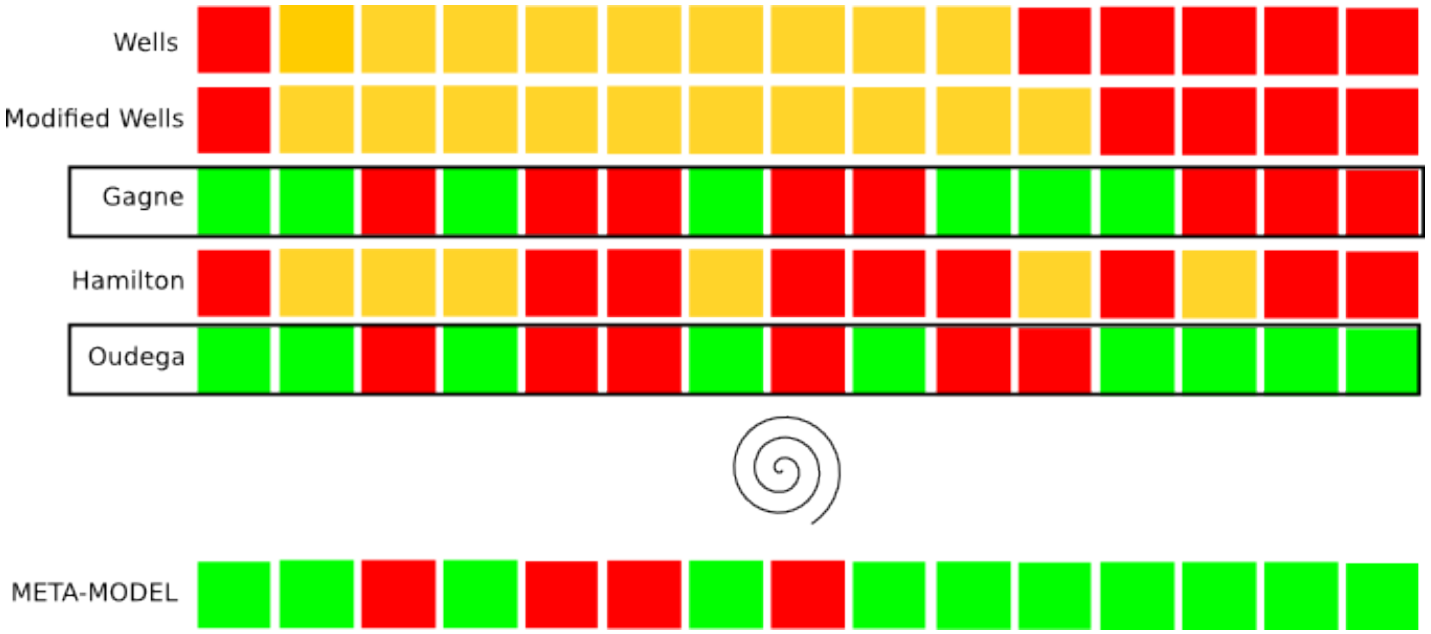
Model averaging

Required steps

1. Update literature models to validation sample
2. Calculate predictions for each subject, for each model
3. Evaluate performance literature models
4. Calculate weights based on model fit and updating complexity (BIC)
5. Obtain (weighted) average predictions
6. Calculate summary model

Model averaging of prediction models for diagnosing deep venous thrombosis

- Update intercept and common slope of all models
- Achieved weights:
  0.998 (Oudega), 0.002 (Gagne), 0 (other models)

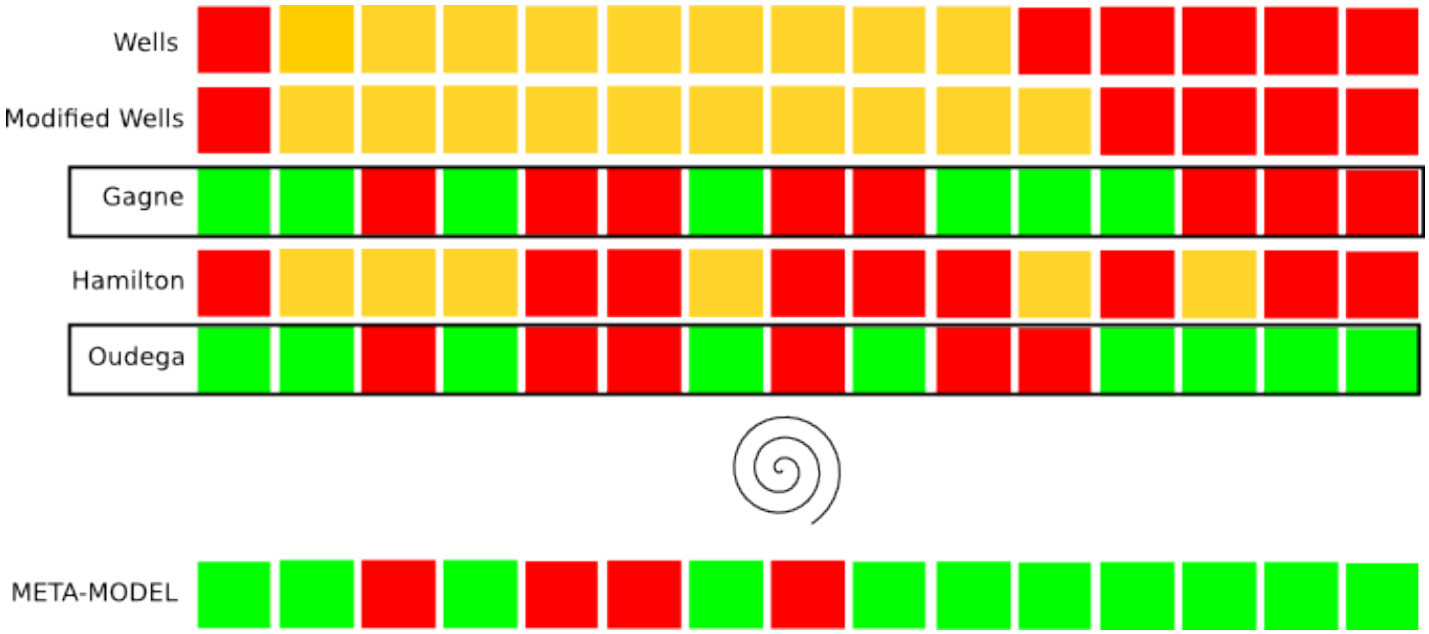


Stacked regressions

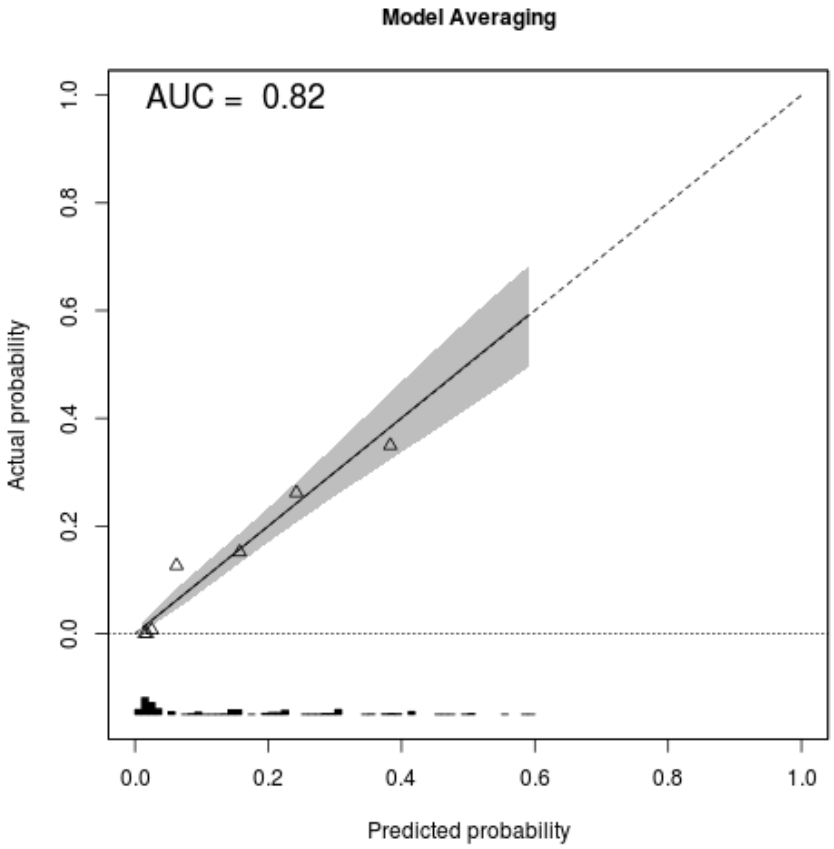
- Weight predictions from literature models
- Discard models with little (added) value
- Update common intercept and overall slope
- No distinct steps, one straightforward estimation procedure
- Borrows less information from validation sample (as compared to model averaging)

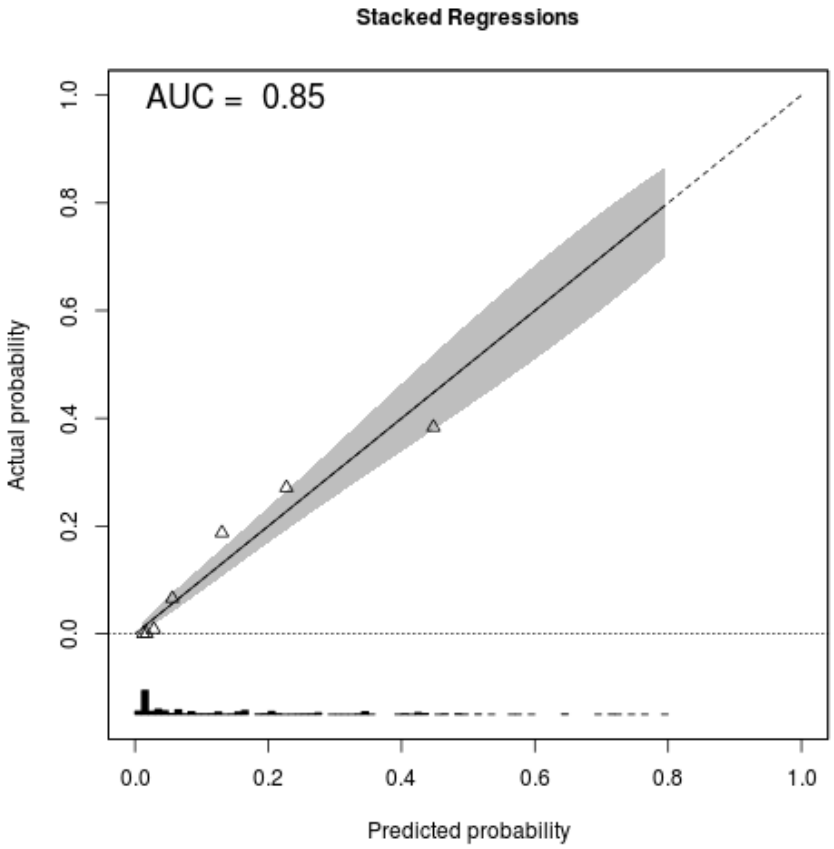
Stacked regressions of prediction models for diagnosing deep venous thrombosis

- Achieved weights:
  1.01 (α), 0.537 (Oudega), 0.497 (Gagne), 0 (other models)



Internal validation of meta-models for diagnosing deep vein thrombosis





External validation of meta-models for diagnosing deep vein thrombosis

- Primary Care (N=791)
  - Best literature model: AUC = 0.77, slope = 1.13
  - Model Averaging: AUC = 0.77, slope = 1.13
  - Stacked Regressions: AUC = 0.74, slope = 0.82
- Secondary Care (N=1756)
  - Best literature model: AUC = 0.84, slope =1.29
  - Model Averaging: AUC = 0.86, slope = 1.29
  - Stacked Regressions: AUC = 0.88, slope = 1.33

Take home messages

- Strong focus on model (re-)development
  - Little efforts on model validation
  - Model performance often worse than anticipated

- Model (re-)development only useful when…
  - … large (validation) sample available
  - … existing literature models too heterogeneous with target population (i.e. differences beyond intercept and common slope)

Take home messages

Model updating recommended in many settings

Problems:

- Which literature model should be updated/used?
- How extensively should the model be updated?
- How to account for evidence from other models?

Take home messages

Systematic review & meta-analysis of prediction models

- Novel paradigm for model development/validation
- Model aggregation versus selective updating
- Better use of prior knowledge, but only if relevant for target population

Take home messages

Two types

1. In case no own (validation) IPD set
   Summarize performance of existing model(s)

2. In case own (validation) IPD set – combination of aggregate and IPD
   a. Combine models with similar predictors
   b. Combine models with different predictors

Take home messages

Methods

- Pooling of individual predictor effects
  - Allows to simplify existing models whilst achieving similar performance
  - Difficult when studies adjust for different co-variates
  - Susceptible to bias and heterogeneity

- Model averaging & stacked regressions
  - Identify added value of existing models
  - Combine updating and aggregation
  - Outperform individual literature models

Handy Tools/Papers

- CHARMS paper – Plos Med 2014 (Moons et al)
- TRIPOD paper (Collins et al, 14 journals)
- PROBAST –Robert Wolff et al (2015)

Workshop aftercare

- Questions about workshop?
- Assistant needed with review of studies of prognosis studies?
- Please contact:
  - PMG Coordinator: Alexandra Hendry (Alexandra.Hendry@sswahs.nsw.gov.au)
  - PMG Co-convenor: Karel Moons (K.G.M.Moons@umcutrecht.nl)