UMC Utrecht
Julius Center

Systematic reviews of prognostic studies 3

meta-analytical approaches in systematic reviews of prognostic studies

Thomas PA Debray, Karel GM Moons
for the Cochrane Prognosis Review Methods Group

Keele University
Cochrane
University Medical Center Utrecht

Conflict of interest

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

Basic & Advanced Courses

MSC EPIDEMIOLOGY

Systematic Reviews, Meta Analysis
Clinical Epidemiology & Statistics

Face to Face & Online

Accessible from all over the world
More than 50 courses…

- **Systematic Reviews of Randomised Intervention Studies**
- **Systematic Reviews of Diagnostic Studies**
- **Systematic Reviews of Prognostic Studies**
- **Meta-Analysis with Individual Participants Data**
- **Numerous courses on primary studies** (randomised trials, side effects, diagnostic accuracy, prognosis and etiology studies)
- ….and many more

www.msc-epidemiology.eu **www.msc-epidemiology.online**

UMC Utrecht Julius Center

Universiteit Utrecht

elevate

Prediction

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

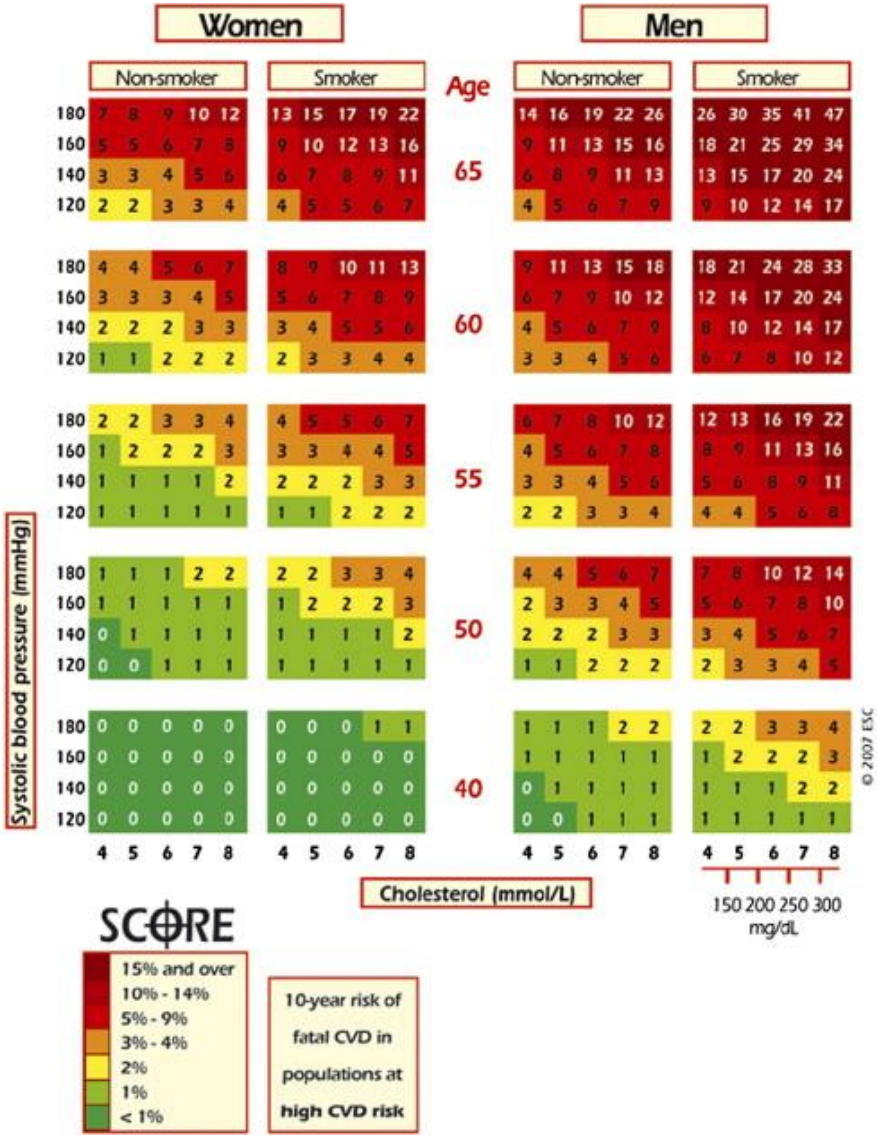
- Turn available information (predictors) into a statement about the probability:
  … diagnosis
  … prognosis

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

Prediction models

APGAR
Test Scoring

	Score 0	Score 1	Score 2
Appearance	Blue all over	Blue only at extremities	No blue coloration
Pulse	No pulse	<100 beats/min.	>100 beats/min.
Grimace	No response to stimulation	Grimace or feeble cry when stimulated	Sneezing, coughing, or pulling away when stimulated
Activity	No movement	Some movement	Active movement
Respiration	No breathing	Weak, slow, or irregular breathing	Strong cry



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

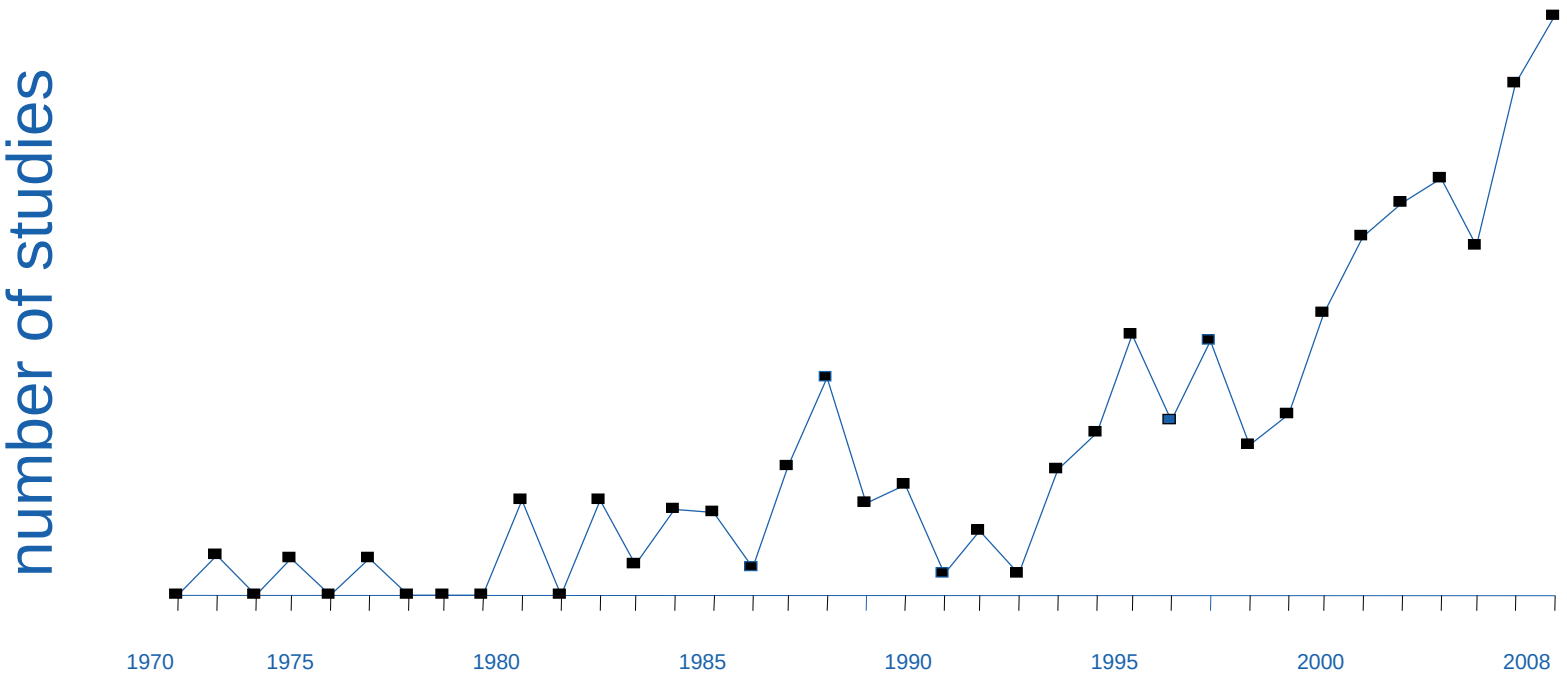
- Average/overall prognosis: 'What is the most likely course (outcome) of people with this health condition?'
- Prognostic factors: 'What factors are associated with that outcome?
- Prognostic (prediction) models: 'Are there risk groups who are likely to have different outcomes?'
- Treatment selection/factors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

Focus this workshop: MA of prediction model studies

BOTH: PROGNOSTIC AND DIAGNOSTIC

Why focus on prediction models?

Steyerberg 2009



Three phases of Prediction Modelling

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

1. Developing a prediction model
2. Validate (+update) the model in other subjects
3. Quantify model's impact on doctor's decision making and patient outcome (cost-effectiveness)

What is big difference between 3 versus 1-2?

Focus on 1-2

External validation

What is it?

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

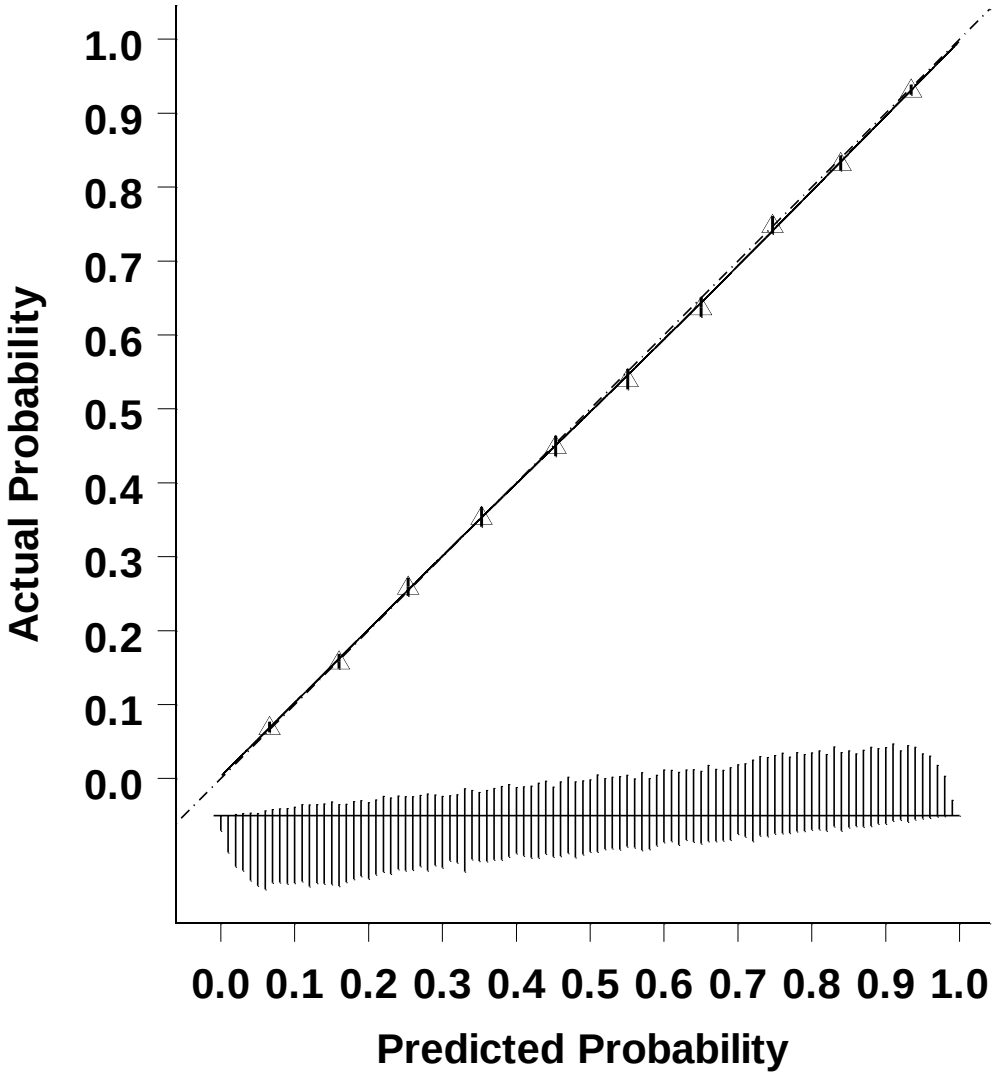
Why do we need it?

- 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?

Prediction model performance measures

UMC Utrecht

Calibration plot – good model?

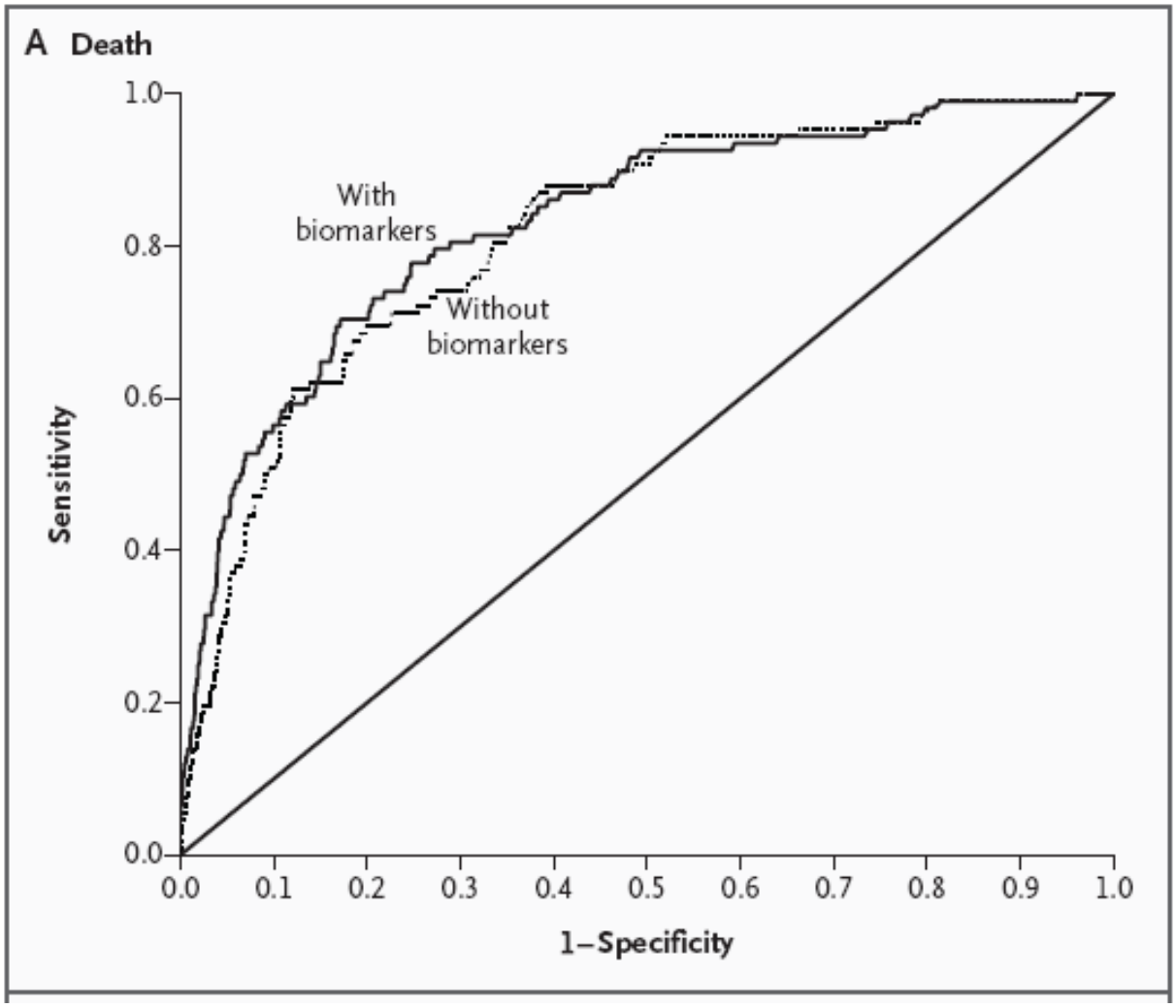


Ideal calibration

O:E = 1

Slope = 1

Model to predict cardiovascular outcomes – added value biomarkers?



AUC 0.76

AUC 0.77

Wang TJ, et al. NEJM

Prediction model performance measures

- Calibration
  - plot (for specific time point in case of survival models)
  - Ratio of observed and expected events
  - Calibration slope, calibration-in-the-large

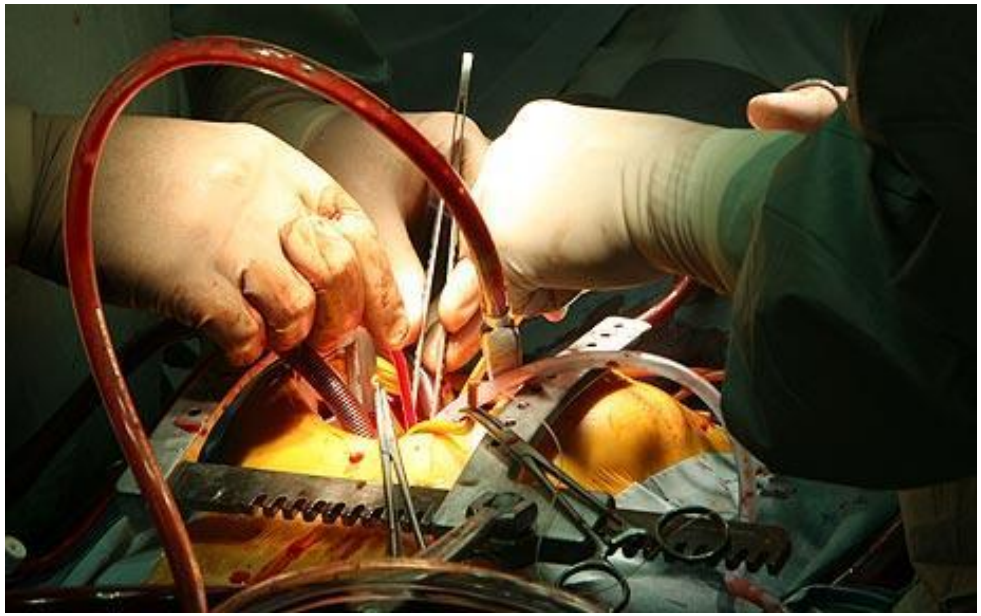
- Discrimination
  - C-statistic (ROC area for logistic regression)

- (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

Example

Predicting mortality after cardiac surgery

- Cardiac surgery in high-risk population
- Need for risk stratification
- Establish risk profile of cardiac surgical patients using **multivariable prediction models**



Example
Predicting mortality after cardiac surgery

Development of EuroSCORE model

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☆

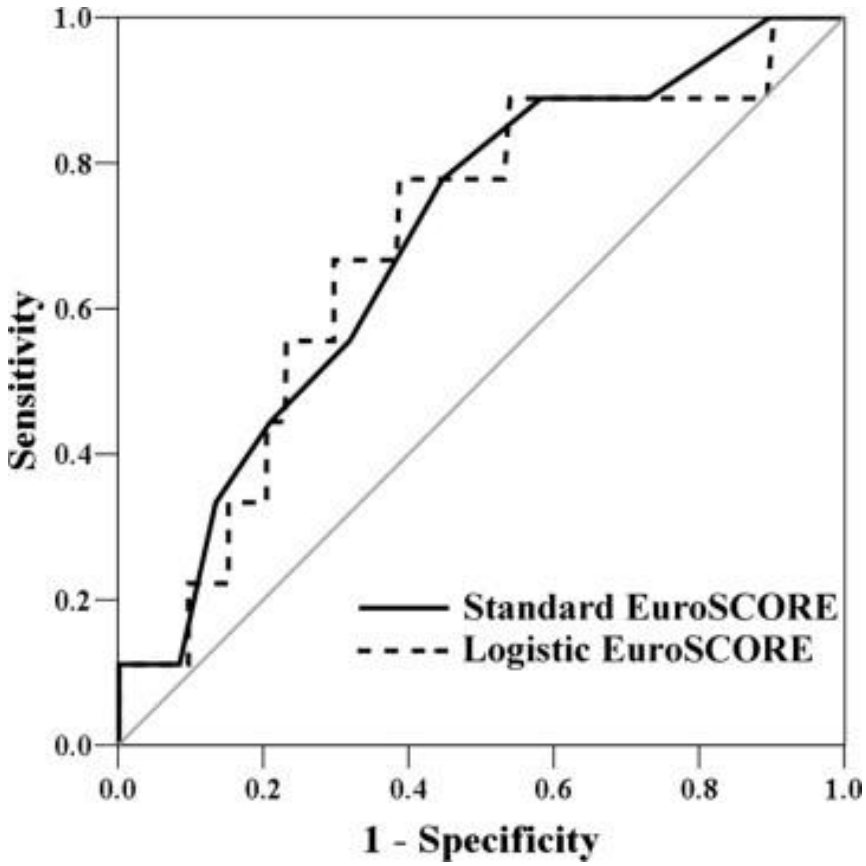
F. Roques*, S.A.M. Nashef, P. Michel, E. Gauducheau, C. de Vincentiis, E. Baudet, J. Cortina, M. David, A. Faichney, F. Gabrielle, E. Gams, A. Harjula, M.T. Jones, P. Pinna Pintor, R. Salamon, L. Thulin

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

Example
External validation of EuroSCORE

Discrimination



What c-statistic does the ROC curve indicate?

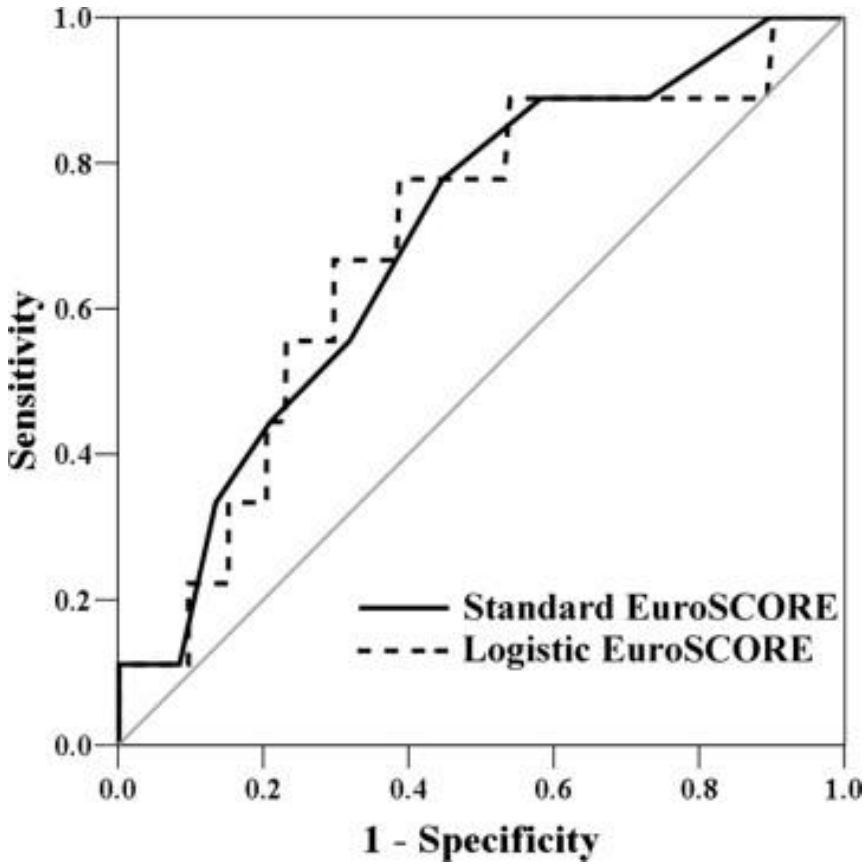
(a) 0.75 – 1.00

(b) 0.60 – 0.75

(c) < 0.60

Example
External validation of EuroSCORE

Discrimination



What c-statistic does the ROC curve indicate?

(a) 0.75 – 1.00

(b) 0.60 – 0.75 (0.71)

(c) < 0.60

Example
External validation of EuroSCORE

Calibration

Expected mortality (%) versus observed in-hospital mortality

Score	N	Expected	Observed
0-2	201	1.4	0.5
3-5	309	4.0	1.0
6-8	181	6.8	2.2
>= 9	66	10.5	3.0

How well does the standard EuroSCORE calibrate?

(a) Good

(b) Poor, due to over-prediction

(c) Poor, due to under-prediction

Example
External validation of EuroSCORE

Calibration

Expected mortality (%) versus observed in-hospital mortality

Score	N	Expected	Observed
0-2	201	1.4	0.5
3-5	309	4.0	1.0
6-8	181	6.8	2.2
>= 9	66	10.5	3.0

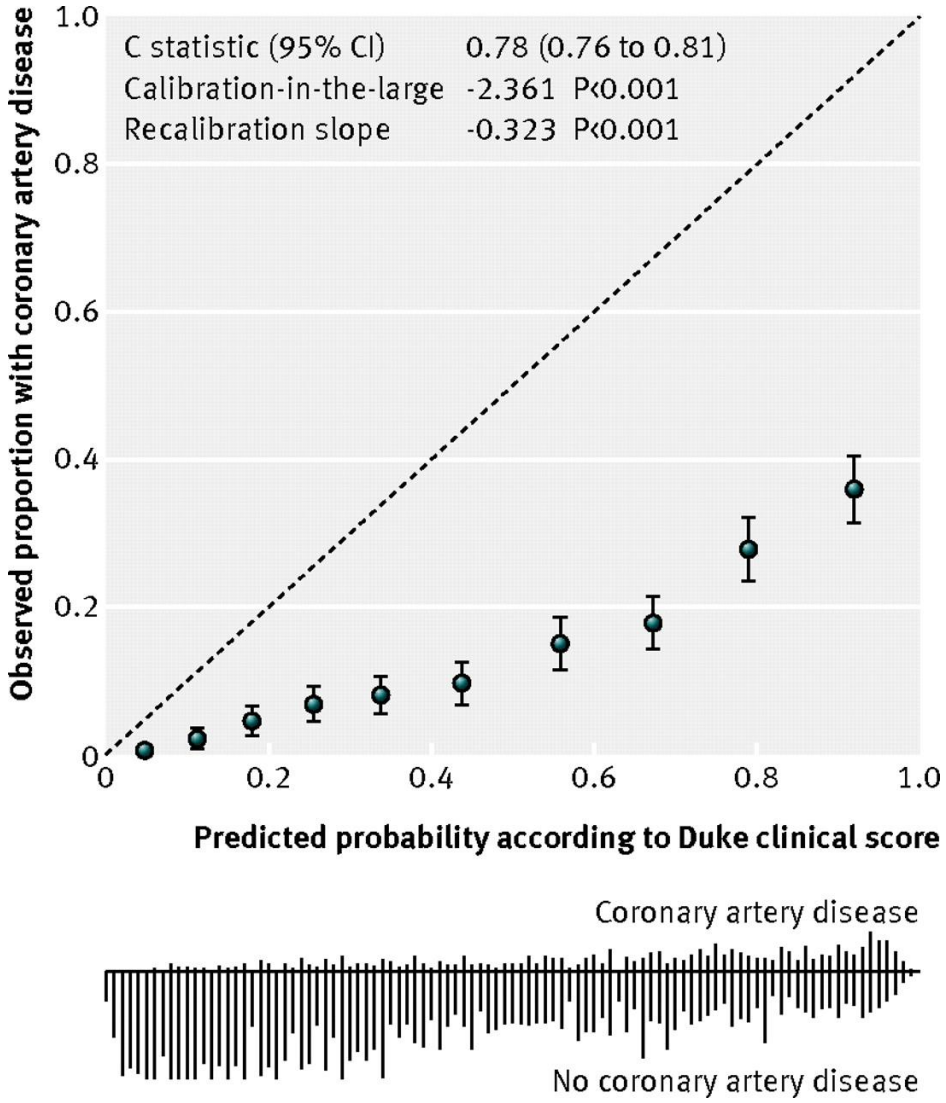
How well does the standard EuroSCORE calibrate?

(a) Good

(b) Poor, due to over-prediction

(c) Poor, due to under-prediction

Calibration plot – good model?



Ref: Genders et al. Prediction model to estimate presence of coronary artery disease: retrospective pooled analysis of existing cohorts. BMJ 2012

Caveats in prediction modeling research

- Lack of external validation studies
- Lack of head-to-head comparisons
- Lack of data to tailor the model to local circumstances

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.

We need systematic review and meta-analysis of validation studies

Aims

- Summarize model performance
- Investigate generalizability

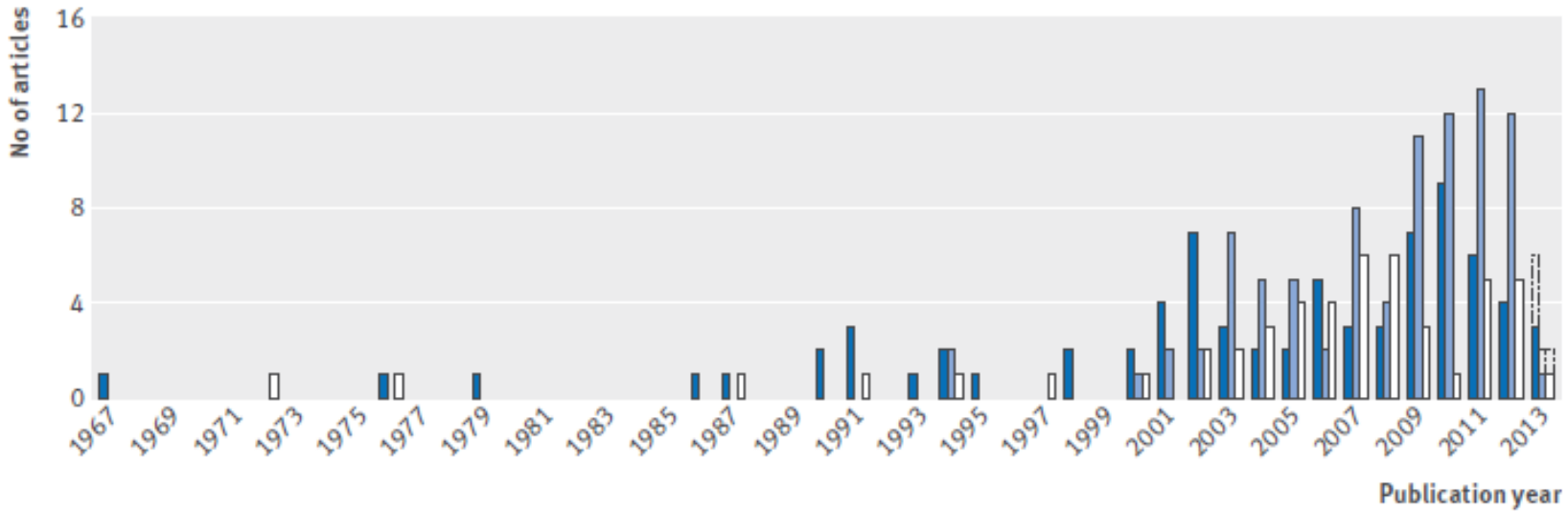
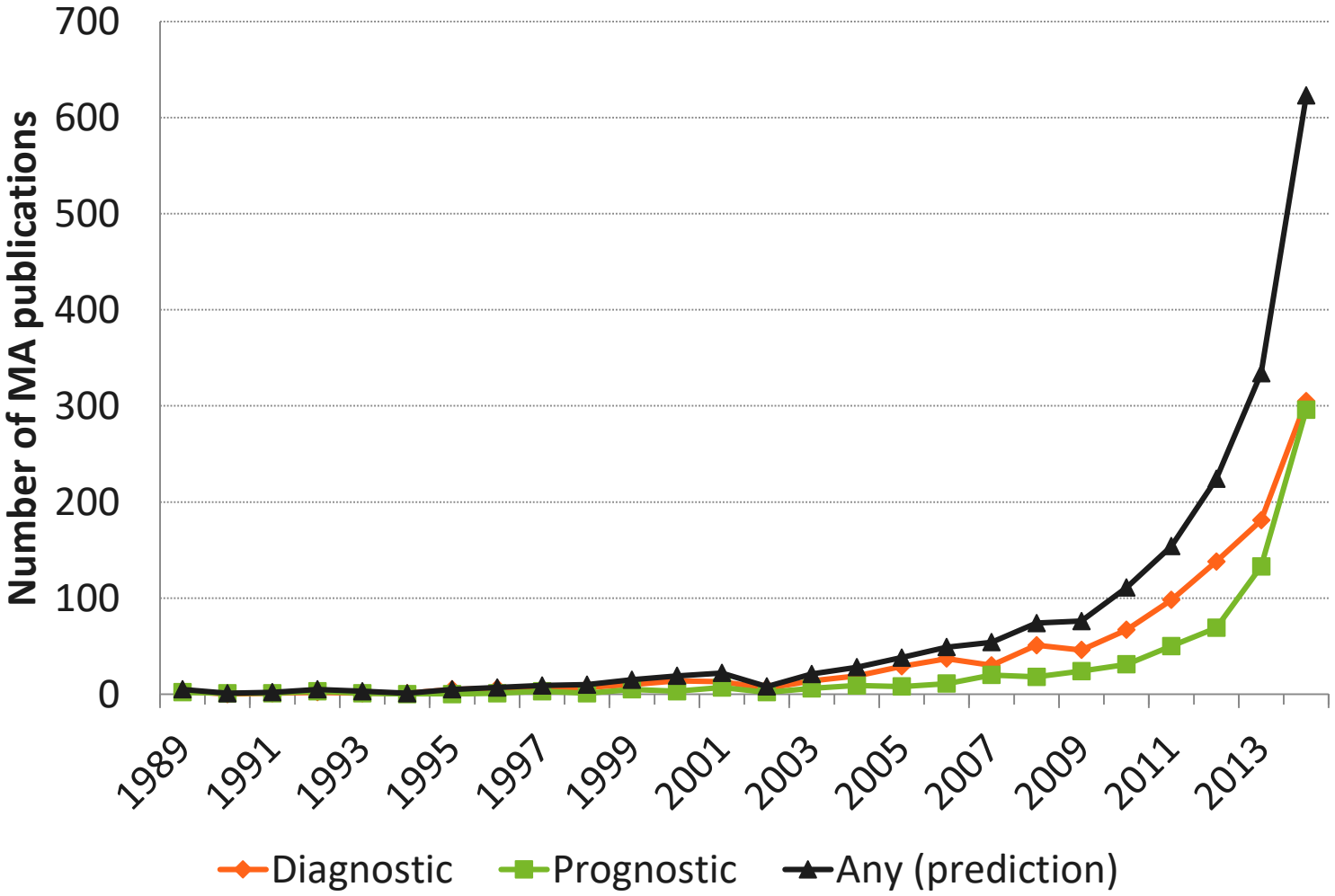


Fig 2 | Numbers of articles in which only one or more models were developed (dark blue), only one or more models were externally validated (light blue), or one or more models were developed and externally validated (white), ordered by publication year (up to June 2013). Predictions of the total numbers in 2013 are displayed with dotted lines

Systematic review and meta-analysis of prediction models increasingly popular



Recommended steps

1. Formulating the review question
2. Formulating the search strategy
3. Critical appraisal (CHARMS & PROBAST)
4. Quantitive data extraction
   - Discrimination
   - Calibration
5. Meta-analysis
6. Reporting (TRIPOD)

Focus of today

Systematic review and meta-analysis of prediction model performance

Illustration: 22 validations of the the additive European system for cardiac operative risk evaluation (EuroSCORE)

Important: The previous additive and logistic EuroSCORE models are out of date. A new model has been prepared from fresh data and is launched at the 2011 EACTS meeting in Lisbon. The model is called EuroSCORE II - we strongly advise that you use this model - available here. If you really wish to calculate the older "additive" or "logistic" EuroSCORE you can use it below.

Patient-related factors			Cardiac-related factors		
Age (years)	0	0	Unstable angina[6]	No	0
Gender	Select	0	LV function	Select	0
Chronic pulmonary disease[1]	No	0	Recent MI[7]	No	0
Extracardiac arteriopathy[2]	No	0	Pulmonary hypertension[8]	No	0
Neurological dysfunction[3]	No	0	**Operation-related factors**		
Previous Cardiac Surgery	No	0	Emergency[9]	No	0
Creatinine > 200 µmol/ L	No	0	Other than isolated CABG	No	0
Active endocarditis[4]	No	0	Surgery on thoracic aorta	No	0
Critical preoperative state[5]	No	0	Post infarct septal rupture	No	0
Standard EuroSCORE	0				
Note: Logistic is now default calculator	Calculate Clear				

Step 1
Formulating the review question and protocol

UMC Utrecht

Step 1
Formulating the review question and protocol

- Describe rationale, objectives, design, methodology and statistical considerations of the systematic review
- Define the PICOTS

Extensively discussed in the CHARMS workshop!

Step 1
Formulating the review question and protocol

Predictive performance of the EuroSCORE

Population	Patients undergoing coronary artery bypass grafting
Intervention	The (additive) EuroSCORE model
Comparator	Not applicable
Outcome(s)	All cause mortality
Timing	30 days, predicted using peri-operative conditions
Setting	risk stratification in the assessment of cardiac surgical results

Step 2
Formulating the search strategy

UMC Utrecht

Step 2
Formulating the search strategy

- Use information from the PICOTS
- Combine with existing search filters
- Evaluate citations of the development paper

Tools: electronic databases, conference abstracts, hand searching, online registers

Step 3
Critical appraisal

UMC Utrecht

Step 3
Critical appraisal

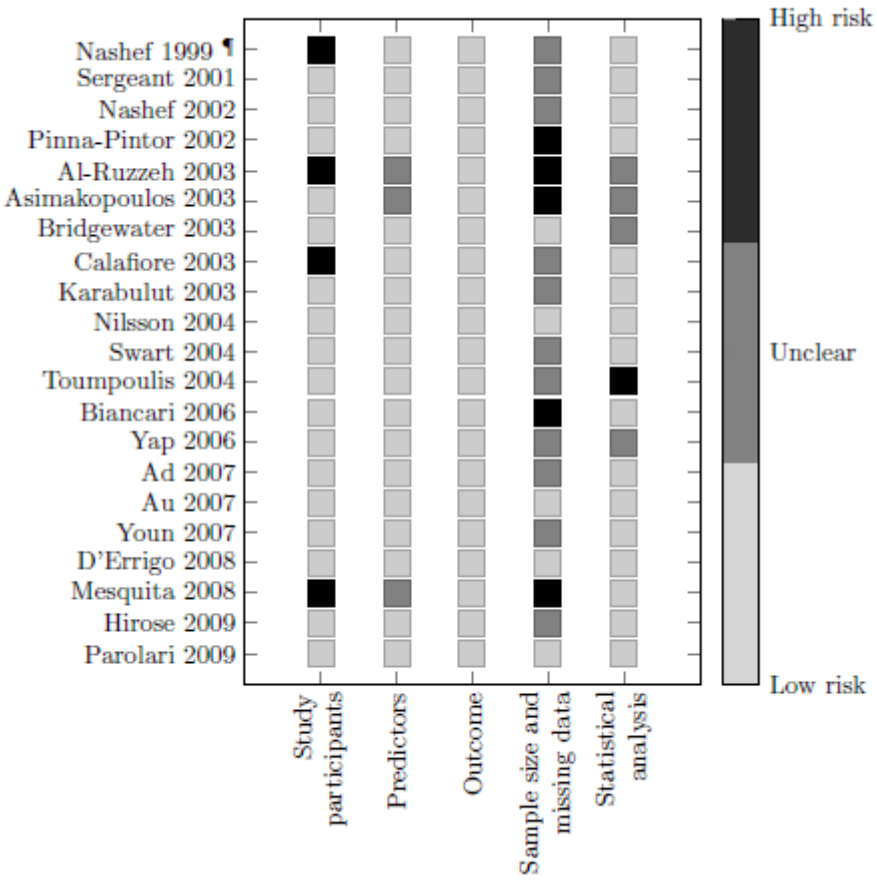
Evaluate **bias and applicability** of each validation study

- CHARMS checklist
- PROBAST (2o17) – see previous workshop

Decide whether studies should be excluded due to low quality and/or applicability with respect to the current review

Step 3
Critical appraisal

Predictive performance of the EuroSCORE



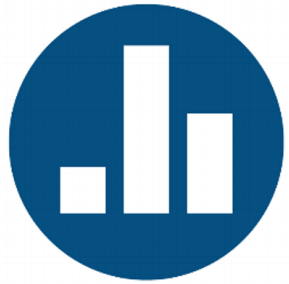
Step 4

Quantitative data extraction and preparation

UMC Utrecht

Step 4

Quantitative data extraction and preparation



What statistics can we summarize when reviewing external validation studies?

Step 4

Quantitative data extraction and preparation

What statistics can we summarize?

- Overall performance
- Model discrimination
- Model calibration

Step 4

Quantitative data extraction and preparation

Measures of overall performance

- Explained variation (R^2)
- Brier score

However, studying the discriminative ability and calibration of a model is often more meaningful than an overall performance measure when we want to appreciate the quality of model predictions for individuals.

Ref: Steyerberg. Clinical prediction models: a practical approach to development, validation and updating. Springer 2009.

Step 4

Quantitative data extraction and preparation

Discrimination

Quantifies the model's ability to distinguish between events and non-events

- Summary statistics
  - Concordance (c) index
  - Area under the ROC curve (AUC)
  - Discrimination slope
- Visual inspection
  - Receiving Operating Characteristics (ROC) curve

Step 4

Quantitative data extraction and preparation

Calibration

Agreement between observed outcomes and predictions

- Summary statistics
  - O:E statistic (#observed events / #predicted events)
  - Calibration-in-the-large
  - Calibration slope
- Visual inspection
  - Calibration plot

Step 4

Quantitative data extraction and preparation

Common problems in data extraction

- Selective reporting
- Inconsistent measures of model performance
- Incomplete assessments (e.g. calibration)
- Missing estimates of precision (e.g. standard error)

Step 4
Quantitative data extraction and preparation

Table 2. Summary of performance data and reporting

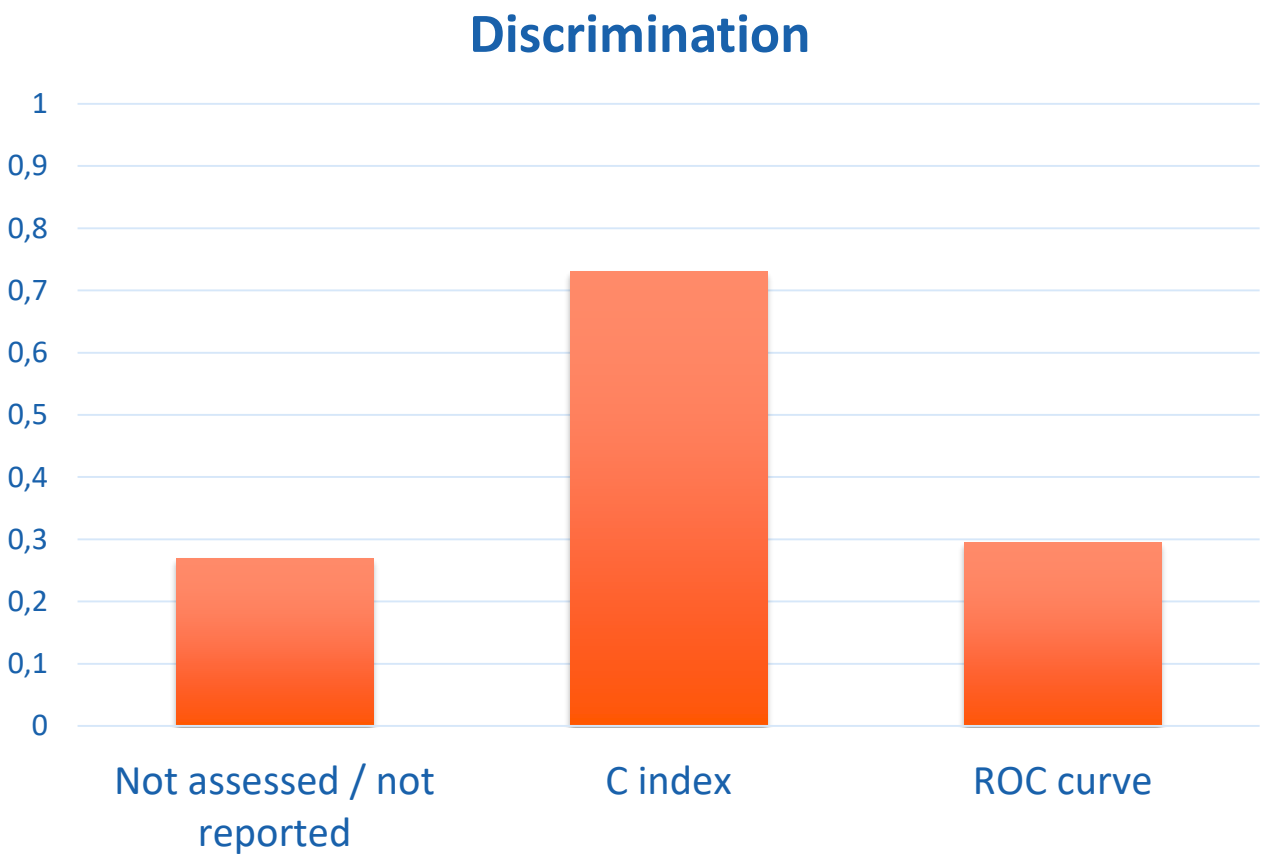
Study	Discrimination; *c*-statistic (95% CI)		Calibration
	Development	Validation	
Ando et al. [32]	Not reported	0.841 (0.799–0.894)	Not assessed
Bang et al. [37]	Not reported	0.88 and 0.71	Not assessed
Chien et al. [33]	0.768 (0.738–0.798)	0.667 (0.631–0.703)	Hosmer–Lemeshow, $P > 0.1$
Fisher and Taylor [38]	Not assessed	Not assessed	Not assessed
Halbesma et al. [34]	0.84 (0.82–0.86)	0.84 (bootstrap)	Not assessed
Hemmelgarn et al. [40]	0.59	0.59	Hosmer–Lemeshow, $\chi^2 = 0.77$
Hippisley-Cox and Coupland [31] (chronic kidney disease)	Not reported	0.877 and 0.875 (women) 0.878 and 0.875 (men)	Calibration plot
Hippisley-Cox and Coupland [31] (end-stage kidney disease)	Not reported	0.843 and 0.818 (women) 0.846 and 0.839 (men)	Calibration plot
Keane et al. [41]	Not assessed	Not assessed	Not assessed
Kshirsagar et al. [35]	0.70	0.70	Hosmer–Lemeshow, $P > 0.2$
Tangri et al. [36]	0.917 (0.901–0.933)	0.841 (0.825–0.857)	Nam and D'Agostino, $\chi^2 = 19$ calibration plot
Thakkinstian et al. [39]	0.770	0.741	0.045[a]

Ref: Collins et al. A systematic review finds prediction models for chronic kidney were poorly reported and often developed using inappropriate methods. JCE 2012.

Step 4

Quantitative data extraction and preparation

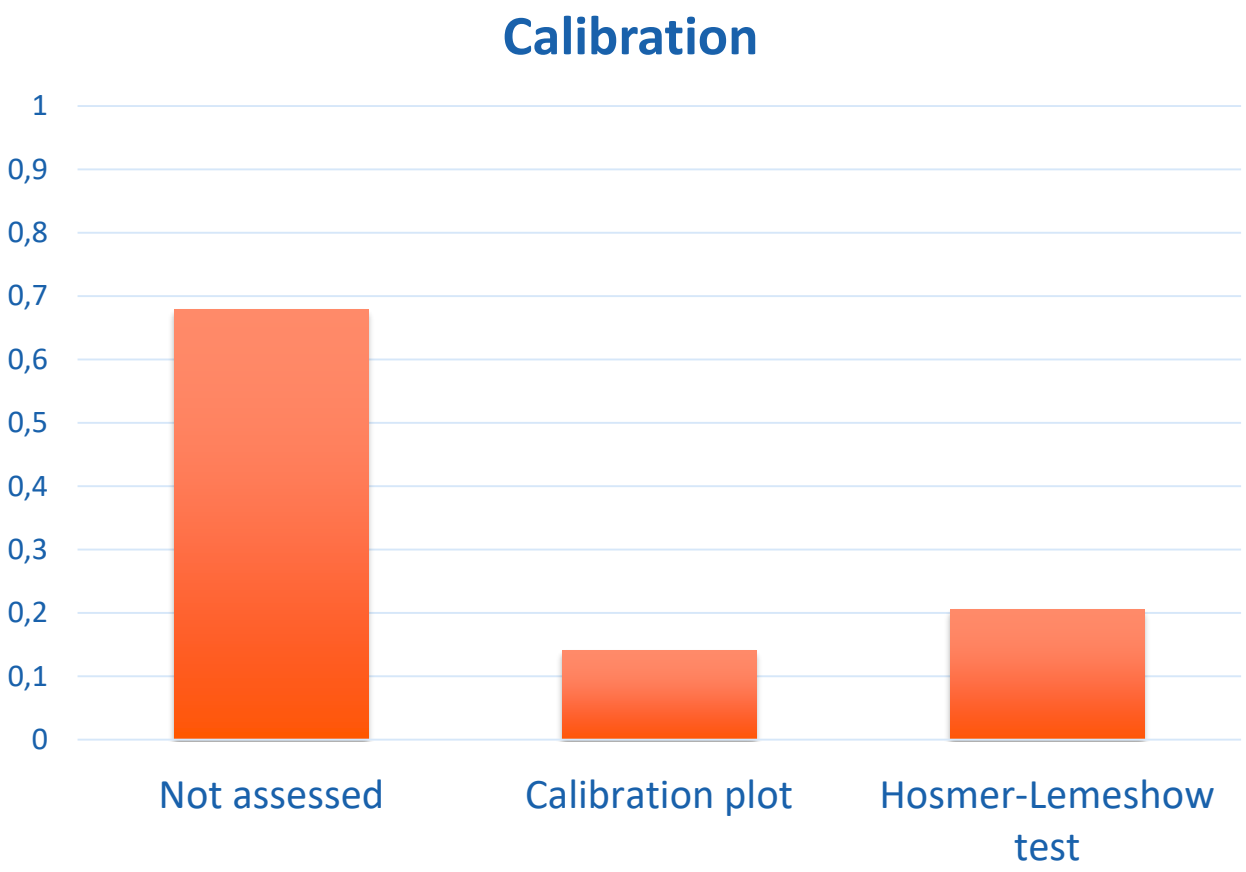
Reporting of performance measures



Step 4

Quantitative data extraction and preparation

Reporting of performance measures



Step 4
Quantitative data extraction and preparation

Dealing with incomplete reporting

- C-statistic
  - Directly related to Somer´s D statistic
  - Can be approximated from the distribution of the LP
  - Can be approximated from the log-odds ratio of the LP
  - Can be approximated from Cohen´s effect size
- O:E ratio (or E:O ratio)
  - Can be calculated from O and E, or, from Pr(O) and Pr(E)
  - Can be derived from calibration-in-the-large
  - Can often be derived from calibration plots and/or tables

Step 4

Quantitative data extraction and preparation

Dealing with incomplete reporting

- SE of the C-statistic
  - can be retrieved from 95% CI
  - can be estimated from #events, #non-events and total sample size
- SE of the O:E ratio
  - Can be retrieved from 95% CI
  - Can be retrieved from p-values
  - Can be approximated from O and/or Pr(O)

Step 4

Quantitative data extraction and preparation

Other information to extract

- Information on case-mix variation
  - Mean & standard deviation of key subject characteristics
  - Mean & standard deviation of the linear predictor
- Information on key study characteristics
  - Location
  - Standards w.r.t. treatments, patient referral, …

Step 4

Quantitative data extraction and preparation

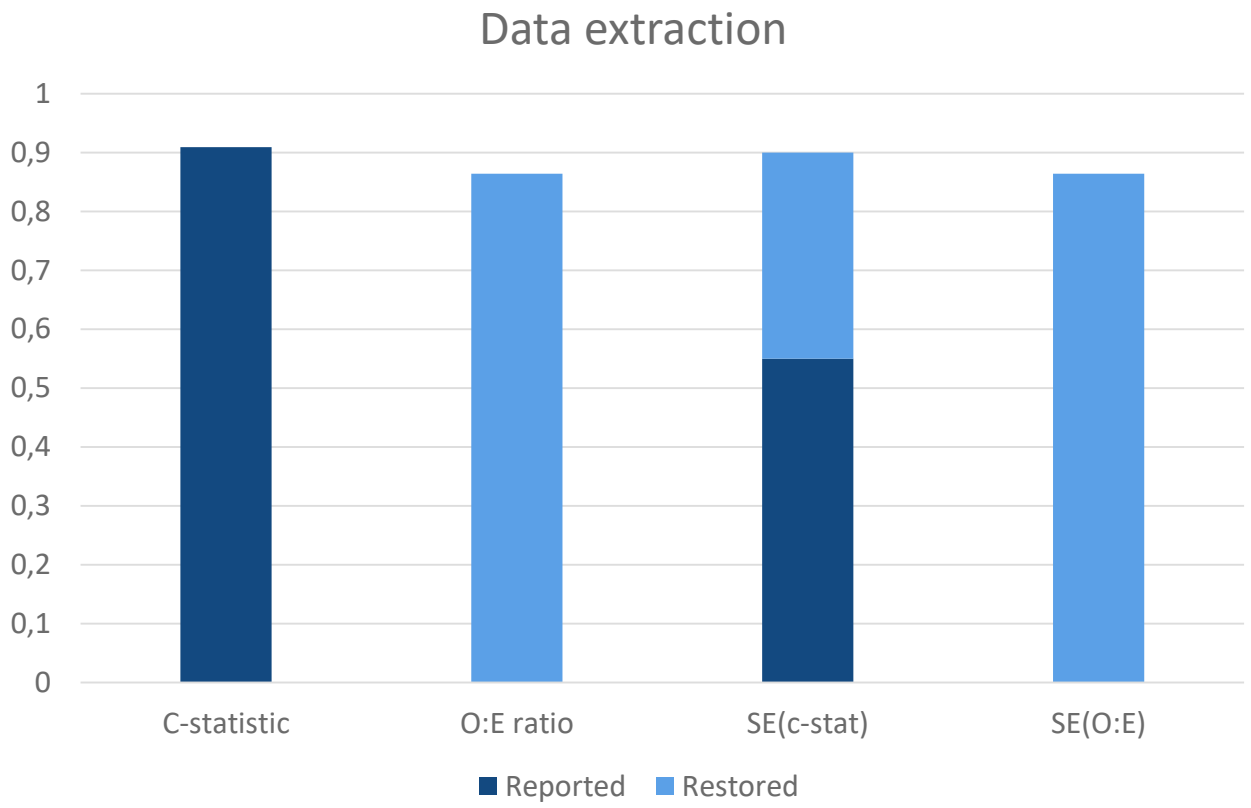
Predictive performance of the EuroSCORE

- C-statistic
  - Summary statistic reported in 20 validations
  - SE approximated for 7 studies
- O:E
  - Relevant information obtained for 21 validations
- Case-mix
  - Distribution of the LP obtained for 15 validation studies

Step 4

Quantitative data extraction and preparation

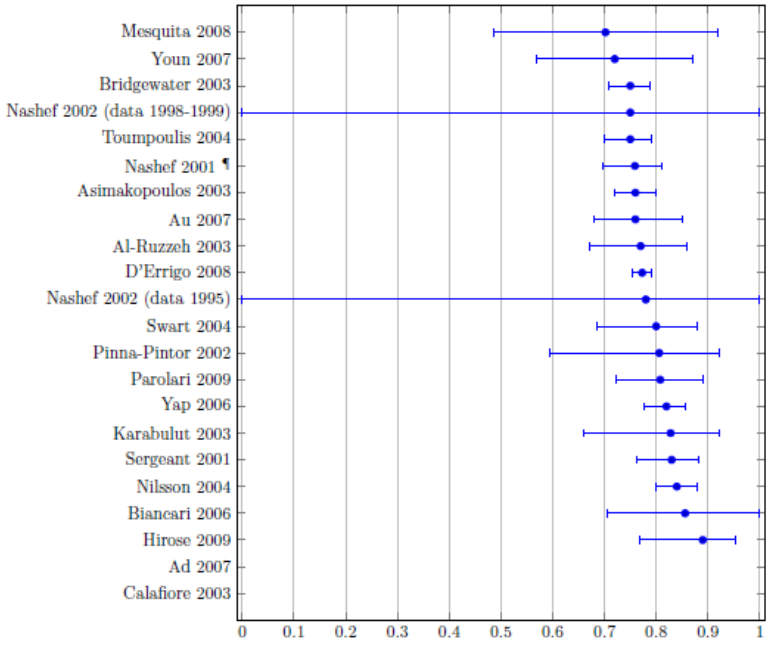
Predictive performance of the EuroSCORE



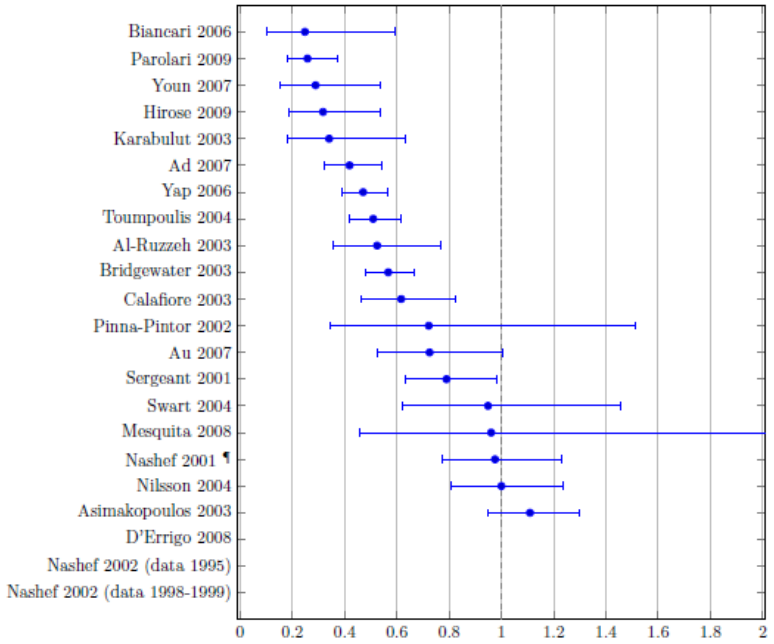
Step 4

Quantitative data extraction and preparation

Predictive performance of the EuroSCORE



(i) Forest plot of the study-specific c-statistics. All 95% confidence intervals were estimated on the logit scale.



(ii) Forest plot of the study-specific total O:E ratios. When missing, 95% confidence intervals were approximated on the log scale using the equations from Appendix 7.

Step 5
Meta-analysis

UMC Utrecht

Step 5
Meta-analysis

Recap

- Fixed effect meta-analysis
  - The model's *true* predictive accuracy is the same for all validation studies
  - Variation in predictive accuracy only appears due to chance
- Random effects meta-analysis
  - The model's *true* predictive accuracy differs across validation studies
  - Variation in predictive accuracy arises from sampling error and between-study heterogeneity

Step 5
Meta-analysis

Fixed or random effects?

- Assumption of homogeneity (fixed effect) often unrealistic because validation studies typically differ in design, execution and case-mix variation
- Ignoring heterogeneity leads to an overly precise summary result
- Summary estimates of predictive accuracy have limited usefulness when there is strong heterogeneity

Step 5
Meta-analysis

Other considerations

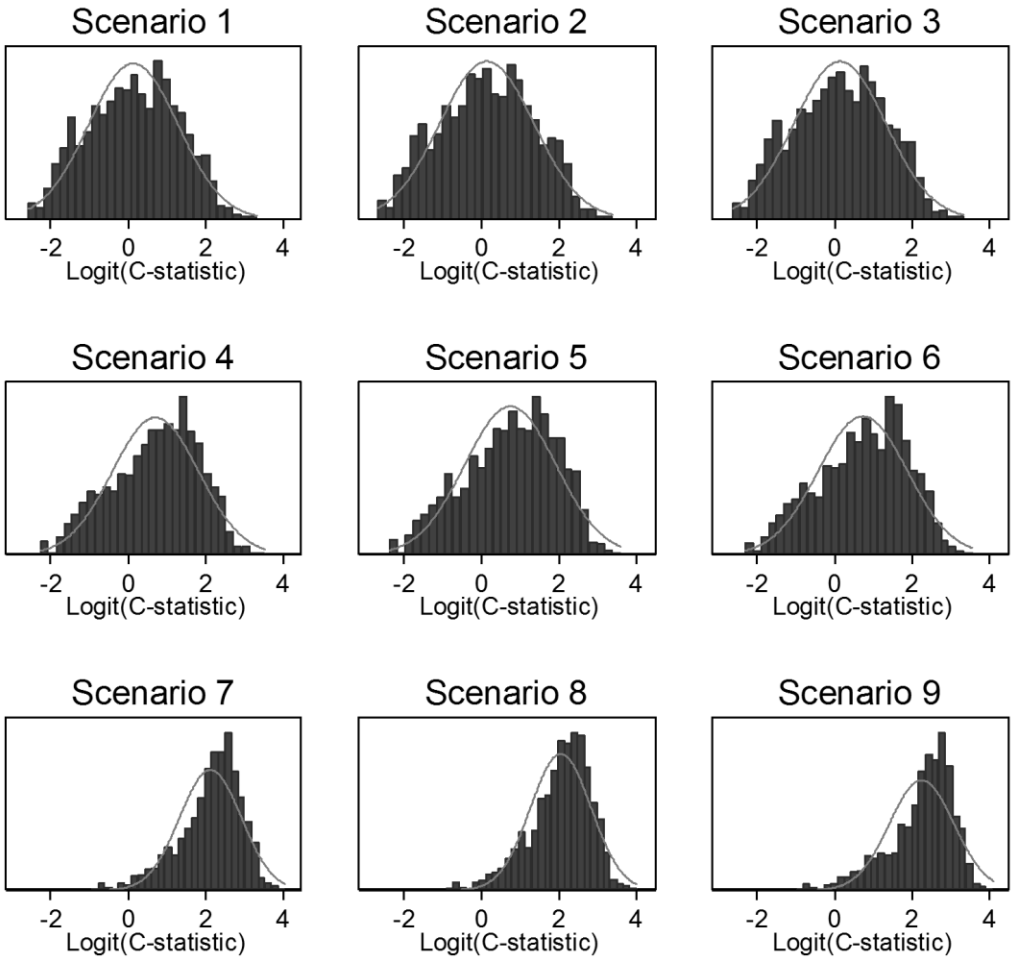
- Traditional meta-analysis methods assume normality of performance statistics within and across studies
- Normality assumption often challenged because:
  - Some performance measures are bounded: c-statistic (between 0 and 1), total O:E ratio (between 0 and +Inf)
  - Central Limit Theorem not applicable in small samples
- Potentially leading to misleading estimates of uncertainty, and to biased summary estimates

Step 5
Meta-analysis

Recommendations

- Allow for random effects
- Rescaling of C-statistics using **logit** transformation
- Rescaling of total O:E ratios using **log** transformation
- No rescaling needed for calibration slope or calibration-in-the-large
- Apply restricted maximum likelihood estimation
- Use Hartung-Knapp-Sidik-Jonkman method for deriving 95% confidence intervals

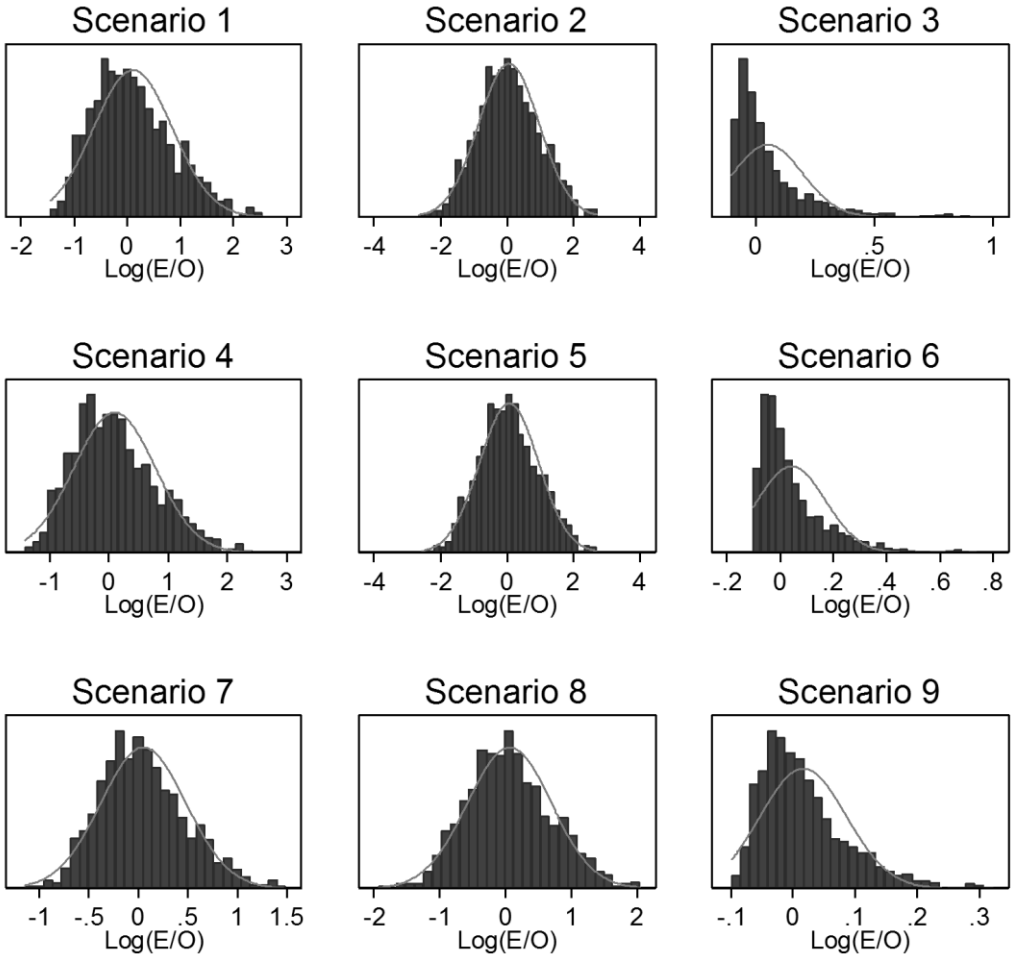
Step 5
Meta-analysis



Histograms for the **logit(c-statistic)** in hypothetical validation studies with strong variation in *true* predictor effects

Ref: Snell KIE et al, Prediction model performance across multiple studies: which scale to use for the c-statistic and calibration measures? *In preparation*

Step 5
Meta-analysis



Histograms for the total **log(E:O)** ratio in hypothetical validation studies with strong variation in baseline risk

Ref: Snell KIE et al, Prediction model performance across multiple studies: which scale to use for the c-statistic and calibration measures? *In preparation*

Step 5
Meta-analysis

Random effects models

- C-statistic

$$\text{logit}(c_i) \sim N\left(\mu_{\text{discr}}, \text{var}(\text{logit}(c_i)) + \tau^2_{\text{discr}}\right)$$

- Total O:E ratio

$$\log(O:E_i) \sim N\left(\mu_{\text{cal}}, \text{var}(\log(O:E_i)) + \tau^2_{\text{cal}}\right)$$

Step 5
Meta-analysis

Formula

- Discrimination:
  - $\text{logit}(c_i) = \log\left(\frac{c_i}{1-c_i}\right)$
  - $\text{var}(\text{logit}(c_i)) = \frac{\text{var}(\boldsymbol{c_i})}{(\boldsymbol{c_i}(\boldsymbol{1}-\boldsymbol{c_i}))^{\boldsymbol{2}}}$
- Calibration
  - $\log(O{:}\,E_i) = \log\left(\frac{O_i}{E_i}\right)$
  - $\text{var}(\log(O{:}\,E_i)) = \frac{\boldsymbol{1}-\boldsymbol{Po_i}}{\boldsymbol{O_i}} \approx \frac{\boldsymbol{1}}{\boldsymbol{O_i}}$

Step 5
Meta-analysis

Guidance paper

More information for obtaining relevant estimates of predictive performance and uncertainty soon to appear:

Debray TPA et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ. *Provisionally accepted*.

Step 5
Meta-analysis

Quantifying heterogeneity

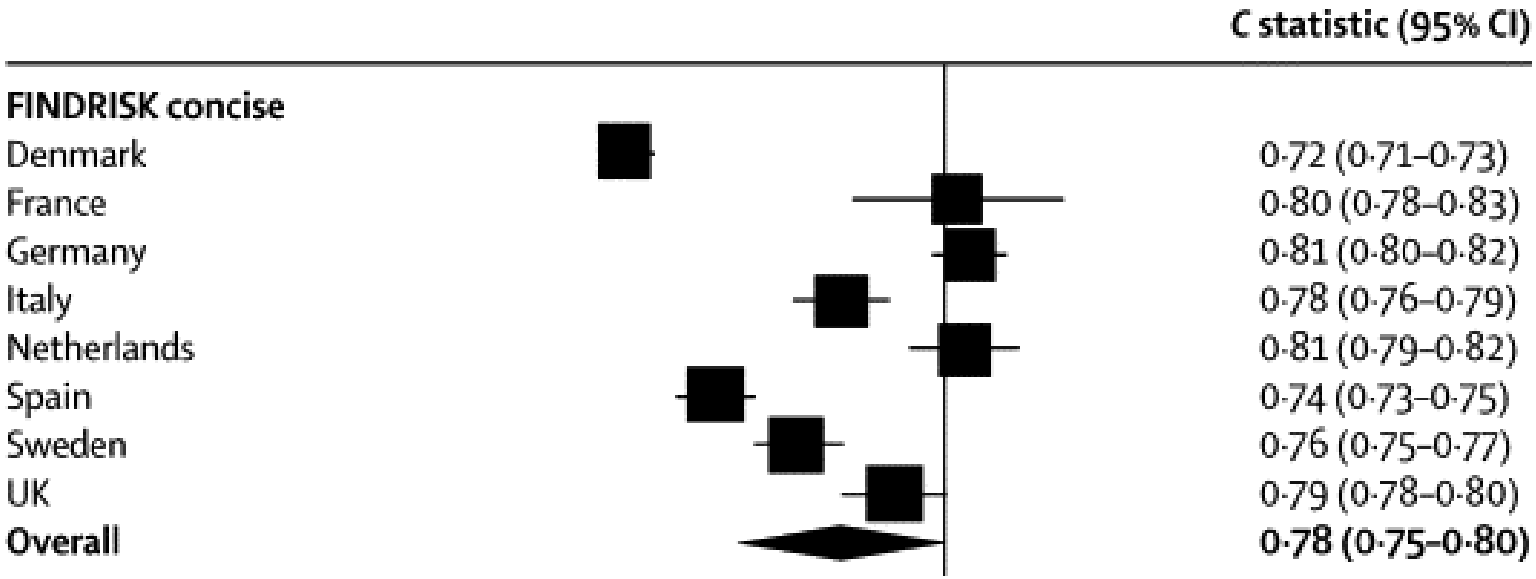
I^2 statistic

- Describes the percentage of total variation across studies that is due to heterogeneity rather than chance
- A value of 0% indicates no observed heterogeneity, larger values show increasing heterogeneity (max: 100%)

Relevance of I^2 depends on the precision of individual studies

Large I^2 values may be misleading and vice versa

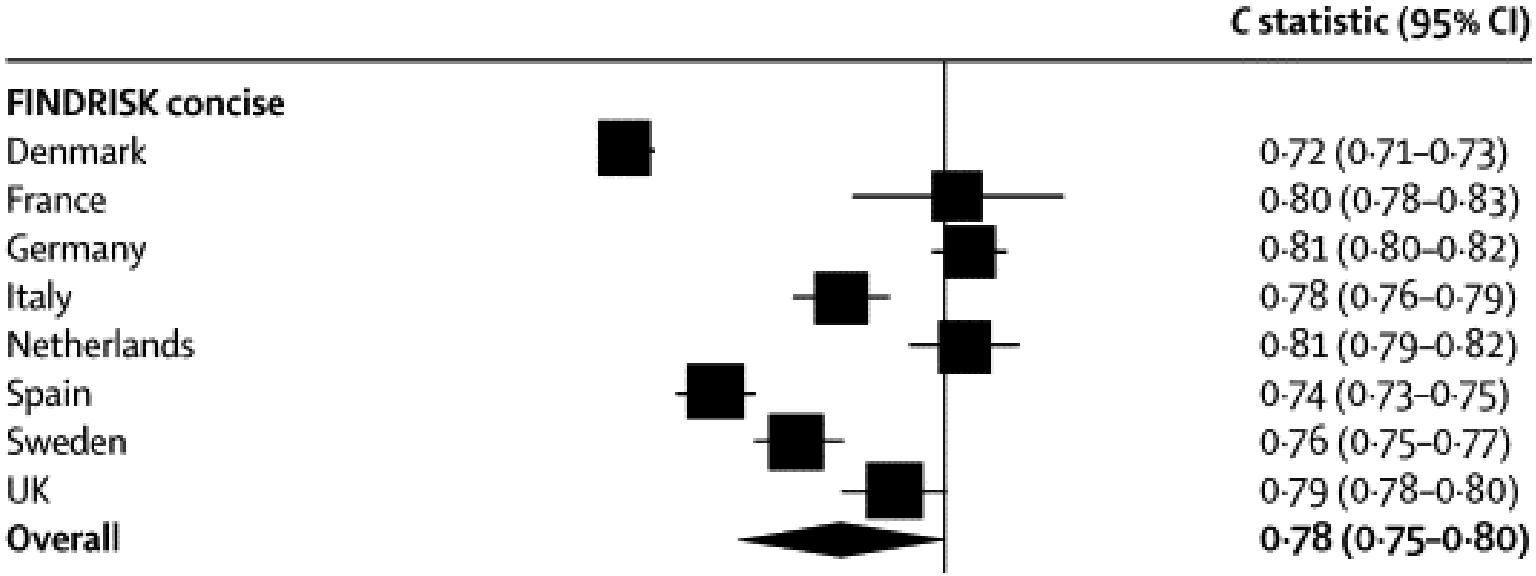
Step 5
Meta-analysis



Ref: Kengne et al. Non-invasive risk scores for prediction of type 2 diabetes (EPIC-InterAct): a validation of existing models. Lancet Diabetes Endocrinol 2014.

Step 5
Meta-analysis

$I^2 = 98\%$



Ref: Kengne et al. Non-invasive risk scores for prediction of type 2 diabetes (EPIC-InterAct): a validation of existing models. Lancet Diabetes Endocrinol 2014.

Step 5
Meta-analysis

Quantifying heterogeneity

Prediction interval

- Combines the standard error of the summary estimate with the estimate for between-study variability
- Typically based on a T distribution
- Provides a range for the potential predictive accuracy in a new validation study
- Ideally calculated from 10 or more validation studies

Step 5
Meta-analysis

Quantifying heterogeneity

Probability of "good" performance

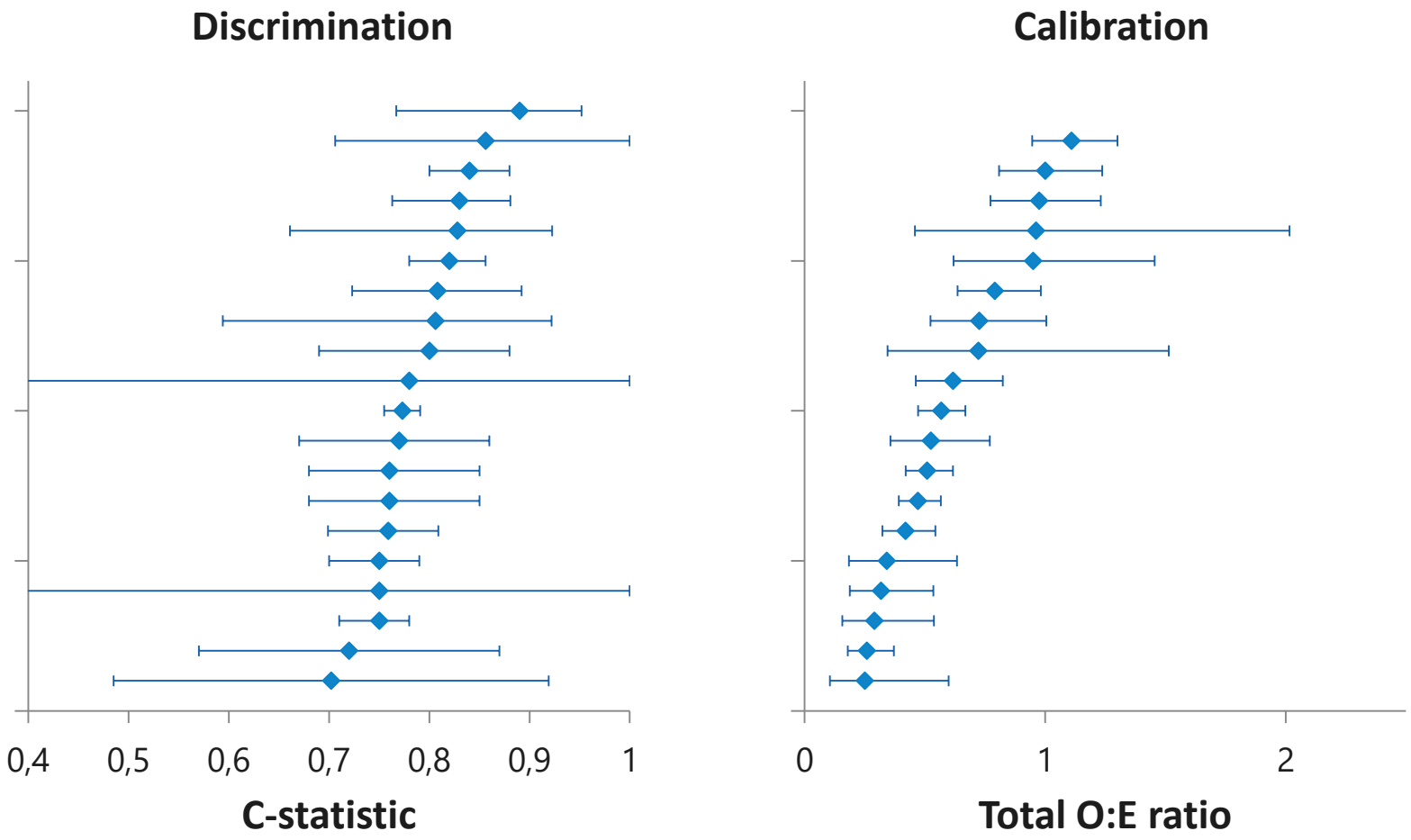
- Calculate the likelihood of achieving a certain c-statistic and/or total O:E ratio in a new validation study
- Rough indication of model generalizability

Ref: Snell et al. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model . JCE 2015.

Step 5

Meta-analysis

Results for EuroSCORE



Step 5
Meta-analysis

Results for EuroSCORE

Meta-analysis	N	Summary	95% CI	95% PI
C-statistic	18	0.78	0.76 – 0.80	0.73 – 0.83
O:E ratio	19	0.55	0.43 – 0.69	0.20 – 1.53

Probability of "good" discrimination ($c > 0.75$) = **89%**

Probability of "good" calibration ($0.8 \leq O:E \leq 1.2$) = **15%**

Step 6

Investigating heterogeneity across studies

UMC Utrecht

Step 6

Investigating heterogeneity across studies

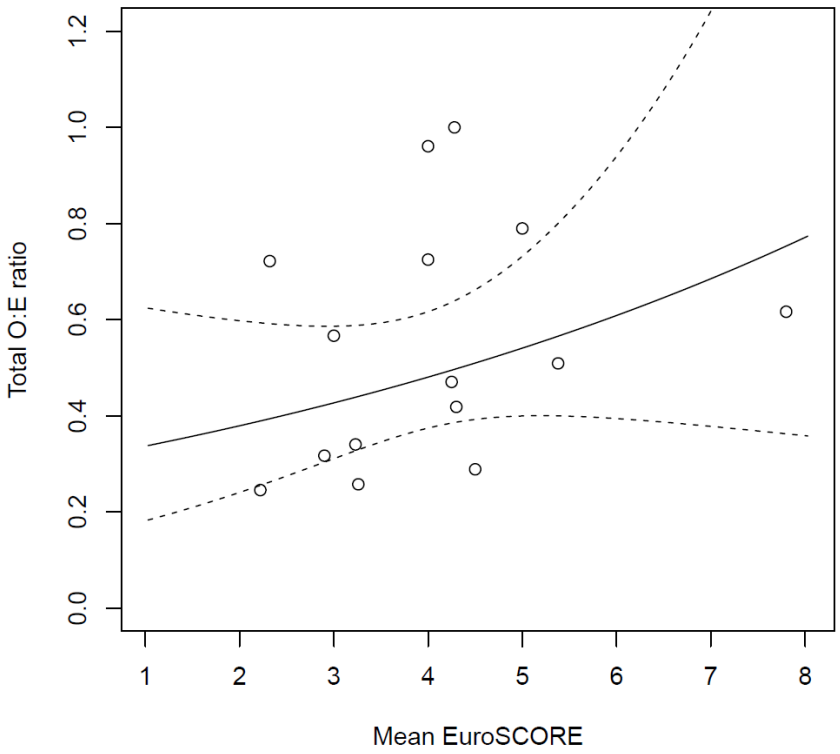
- Summary estimates of limited value in presence of strong heterogeneity
- Heterogeneity in model performance should be expected
  - C statistic may vary due to differences in “true” regression coefficients and/or due to differences in case-mix
  - Total O:E ratio may vary due to differences in outcome prevalence
- Need for meta-regression / subgroup analysis

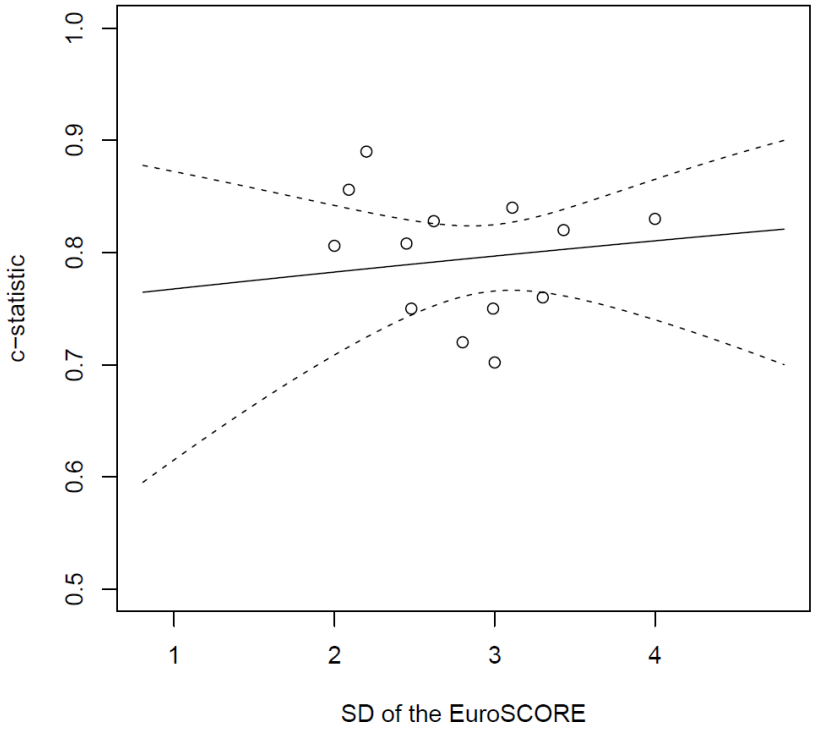
Step 6

Investigating heterogeneity across studies

Meta-analysis of EuroSCORE performance

Adjustment for case-mix variation





Step 7

Sensitivity analyses

UMC Utrecht

Step 7
Sensitivity analyses

Evaluate the robustness of drawn conclusions

- Influence of low(er) quality validation studies
- Influence of key modelling assumptions
- …

Step 7
Sensitivity analyses

Results for EuroSCORE

Meta-analysis	ROB	M	Summary	95% CI	95% PI
C-statistic	All	18	0.78	0.76 – 0.80	0.73 – 0.83
	Low	4	0.80	0.73 – 0.85	0.66 – 0.89
O:E ratio	All	19	0.55	0.43 – 0.69	0.20 – 1.53
	Low	3	0.57	0.10 – 3.33	0.02 – 19.15

Step 7
Sensitivity analyses

Multivariate meta-analysis

- Joint pooling of model discrimination and calibration
- Borrow information across different performance measures within and across studies
- Make joint inferences on different aspects of model performance in new populations

ELSEVIER

Journal of Clinical Epidemiology

Volume 69, January 2016, Pages 40–50

Original Article

Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model

Kym I.E. Snell[a], Harry Hua[b], Thomas P.A. Debray[c, d], Joie Ensor[e], Maxime P. Look[f], Karel G.M. Moons[c, d], Richard D. Riley[e]

Step 8
Reporting

UMC Utrecht

Step 8
Reporting

Relevant guidelines

- PRISMA
- TRIPOD
- GRADE

Closing remarks

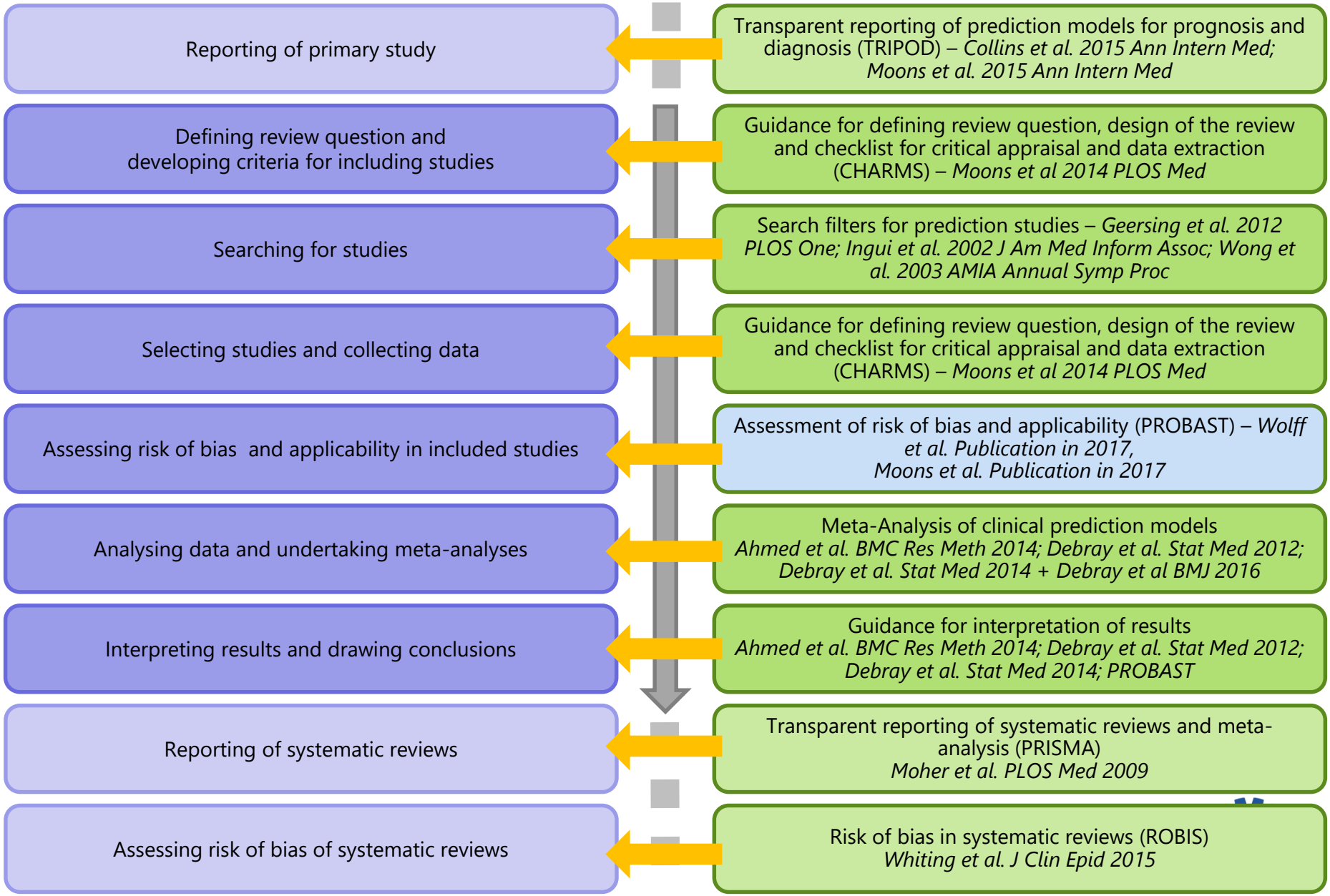
UMC Utrecht

Concluding remarks

- Many similarities to other types of meta-analysis, however,
  - Data extraction more difficult
  - Heterogeneity more common
  - Summary estimates less meaningful

- Recommendations
  - Need for better reporting
  - Need for (minimal set of) standard performance measures
  - Need for IPD

Conducting systematic reviews of prediction model studies



Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 - http://handbook.cochrane.org/

Handy tools / Papers

- Debray TPA et al. **A new framework to enhance the interpretation of external validation studies of clinical prediction models**. J Clin Epidemiol. 2015.

- Debray TPA et al. **A guide to systematic review and meta-analysis of prediction model performance**. BMJ. Provisionally accepted.

- Snell KIE et al. **Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model**. Journal of Clinical Epidemiology. 2015 May;69:40–50.

- Snell KIE et al. **Prediction model performance across multiple studies: which scale to use for the c-statistic and calibration measures**? *In preparation*

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)

Advanced topics

UMC Utrecht

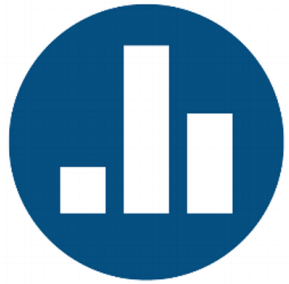
Recall - In case no own (validation) IPD set

Options

1. SR and MA of a specific prediction model across multiple 'model-validation-studies'
   → Investigate heterogeneity in model performance

2. SR and MA of a specific predictor when added to a specific model across multiple 'added-value-studies'
   → Investigate heterogeneity in the added value of a certain predictor

Option 2. SR and MA of specific model across multiple added-value studies



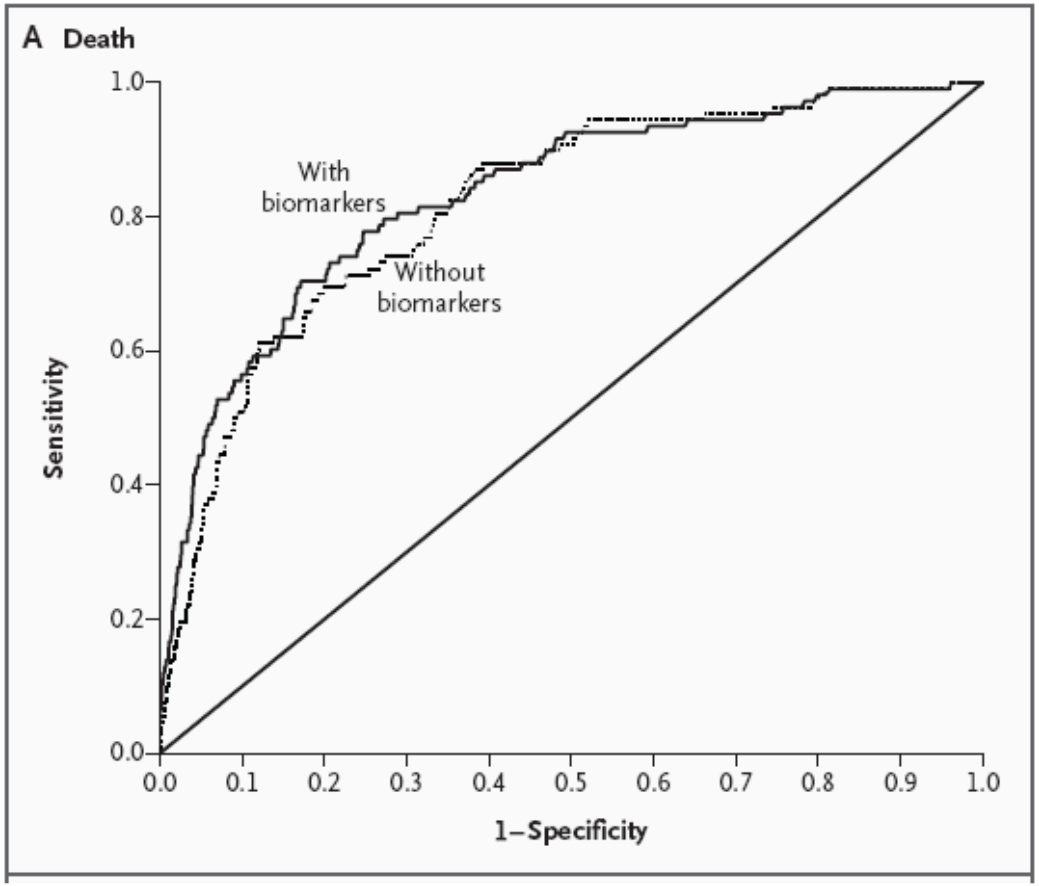
What statistics can we summarize when reviewing added-value studies?

2. SR and MA of specific model across multiple added-value studies

What statistics can we meta-analyze?

- Change in overall performance
- Change in model discrimination
- Change in model calibration
- Model reclassification
- Adjusted regression coefficients

Model to predict cardiovascular outcomes – added value biomarkers?



AUC 0.76

AUC 0.77

Ref: Wang et al. Multiple biomarkers for the prediction of first major cardiovascular events and death. NEJM 2006

Example

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

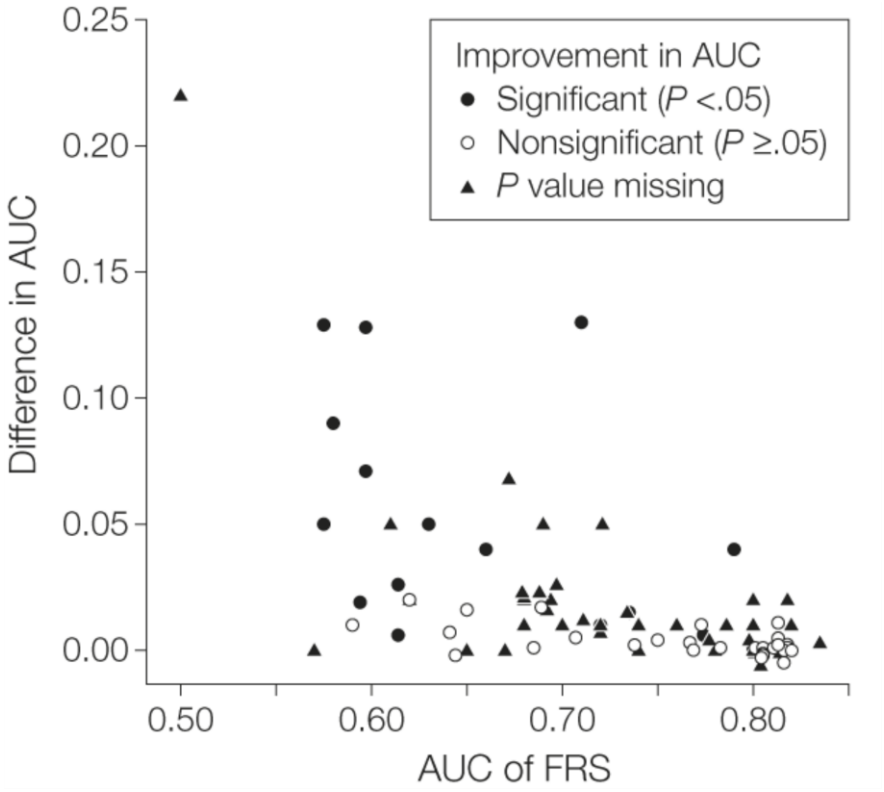
Systematic review of studies that …

- … 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

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

Reported test statistics:

- AUC of FRS alone
- AUC of FRS with additional predictor(s)
- Δ AUC



Meta-analysis of discriminative improvement

- Pooling of Δ AUC statistic can be achieved using the same methods as for pooling AUC of a specific model!

- It is well known that measures of discrimination are insensitive to detecting (small) improvements in model performance when a new marker is added to a model that already includes important predictors

Meta-analysis of model reclassification

Compare alternative models or evaluate addition of a new predictor

- Requires probability thresholds

Procedures

- Two by two tables
  → diagnostic test accuracy MA procedures
- Net reclassification index (NRI)
  → beyond this lecture

Reclassification without probability thresholds

Integrated Discrimination Improvement (IDI) integrates the NRI over all possible cut-offs for the probability of the outcome

- Equivalent to the difference in discrimination slopes of 2 models
- Equivalent to the difference in Pearson R^2 measures
- Equivalent to the difference in scaled Brier scores

So, we are back to meta-analysis of change in overall performance or discrimination

Meta-analysis of adjusted regression coefficients

- Added value studies often correct for *similar* well-known predictors
- It is possible to pool adjusted log-odds (or log-hazard) ratio
- Methods similar to intervention research!

Interpretation of pooled estimates less straightforward

Take home messages

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

Basic & Advanced courses

in Systematic Reviews, Meta Analysis, Clinical Epidemiolgy and Statistics

MSC EPIDEMIOLOGY

Face to Face & Online

- Systematic Reviews of Randomised Intervention Studies
- Systematic Reviews of Diagnostic Studies
- Systematic Reviews of Prognostic Studies
- Meta Analysis with Individual Participants Data
- Clinical Trials and Drug Risk Assessment
- Diagnostic Research
- Prognostic Research
- Missing Data

www.msc-epidemiology.eu

www.msc-epidemiology.online

UMC Utrecht Julius Center

Universiteit Utrecht

elevate