UMC Utrecht
Julius Center

Quantitative synthesis and meta-analytical approaches in syst. reviews of prognostic studies

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)

Keele University

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?
Design, protocol and data extraction using the CHARMS checklist in systematic reviews of prognostic studies	Carl Moons Lotty Hooft	**4 October, Sunday** **16.00 to 17.30**
Assessing risk of bias in studies of prognostic factors using the QUIPS tool	Jill Hayden Carl Moons	**5 October, Monday** **14.00 to 15.30**
Assessing risk of bias in studies of prediction models using the PROBAST tool	Robert Wolff Penny Whiting Carl Moons	**5 October, Monday** **16.00 to 17.30**
Quantitative synthesis and Meta-analytical approaches in systematic reviews of prognostic studies	Thomas Debray Carl Moons	**6 October, Tuesday** **11.00 to 12.30**

Overview Cochrane Prognostic Methods Group (PMG) Workshops

PMG Workshop	Facilitators	When?
Using GRADE in systematic reviews of studies on overall prognosis	Alfonso Iorio Elizabeth Matovinovic Jill Hayden	**7 October, Wednesday 14.00 to 15.30**
IPD Workshop	**Facilitators**	**When?**
Individual Participant Data (IPD) Meta-analysis of prediction modelling studies	Thomas Debray Hans Reitsma	**7 October, Wednesday 14.00 to 15.30**

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'?

Four main types of prognosis studies

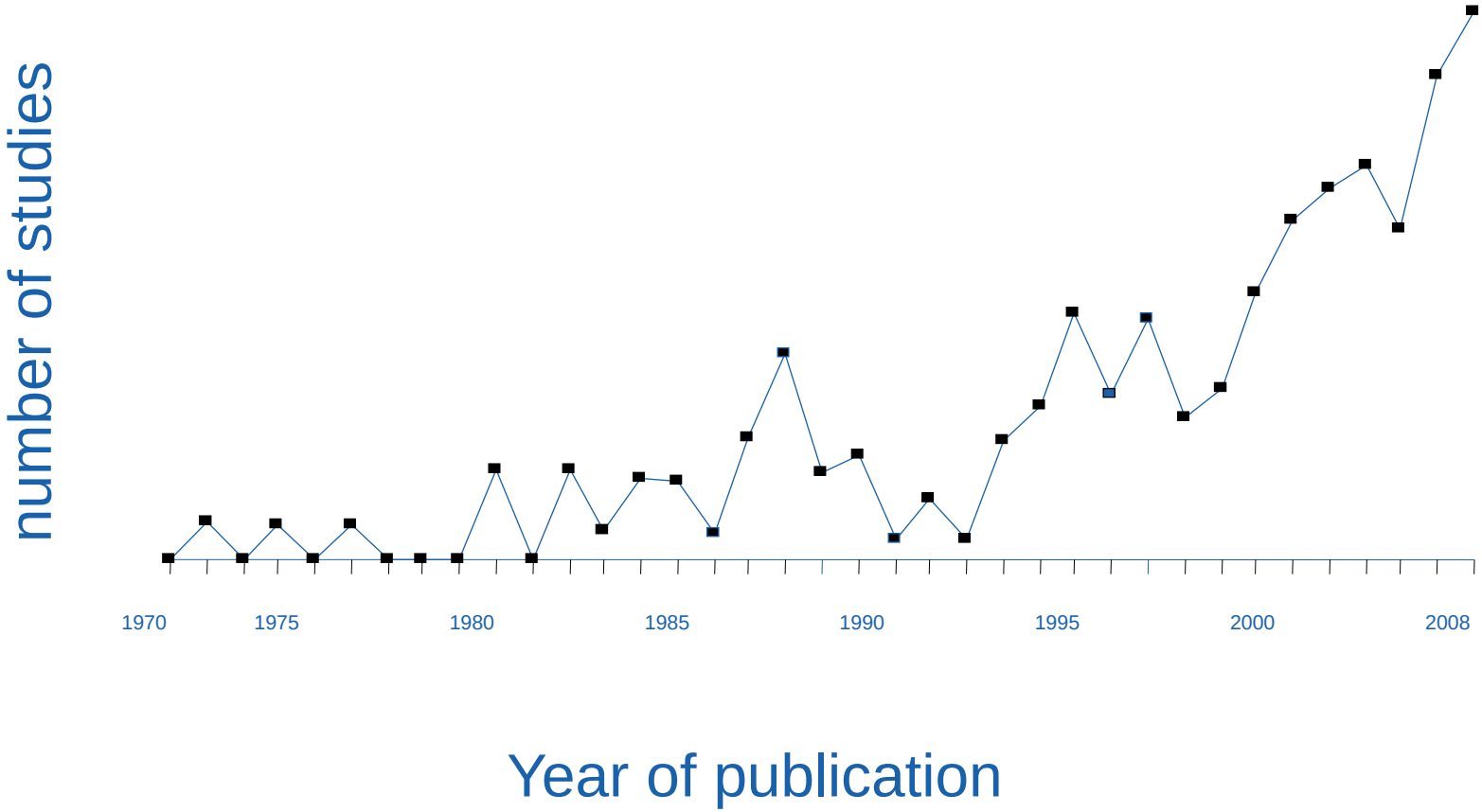
PROGRESS series 2013: BMJ and Plos Med

- 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?

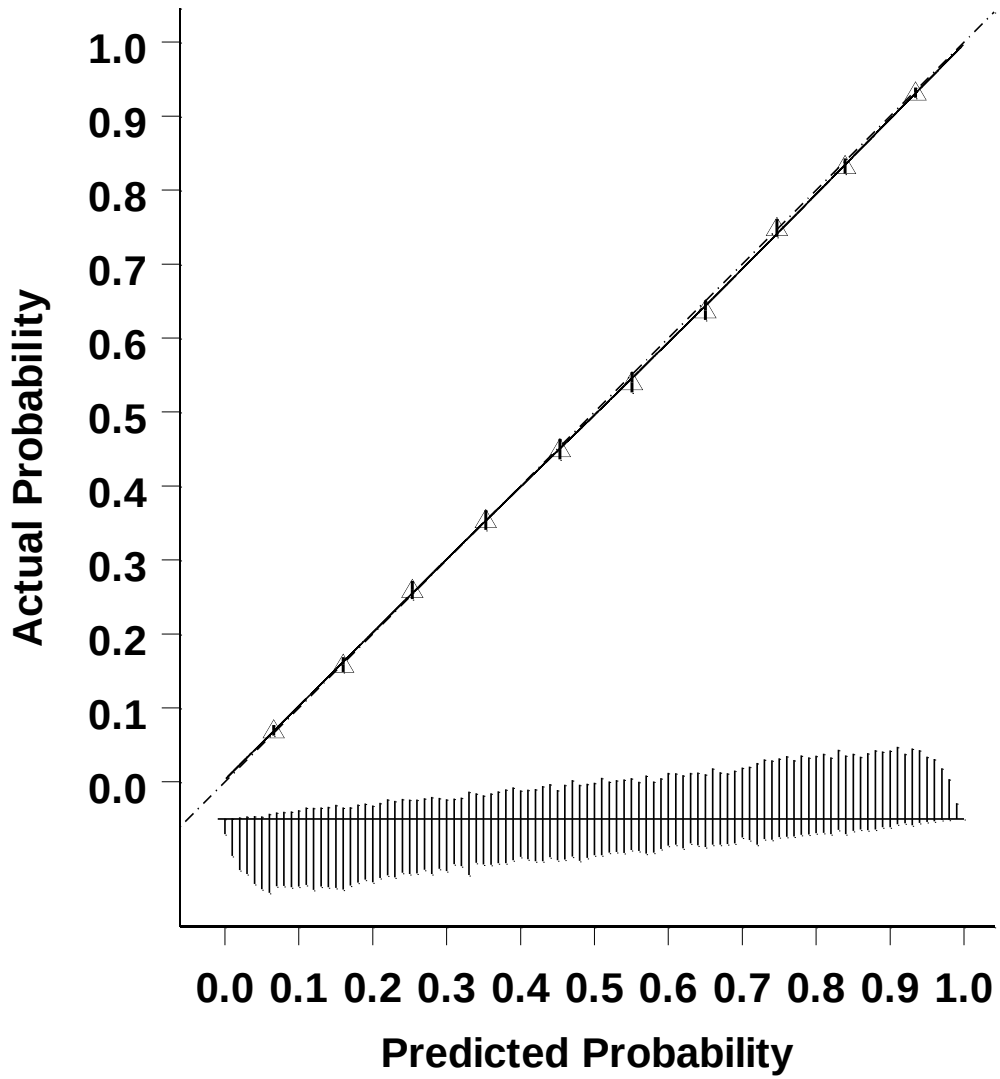
Prediction model performance measures

- 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?

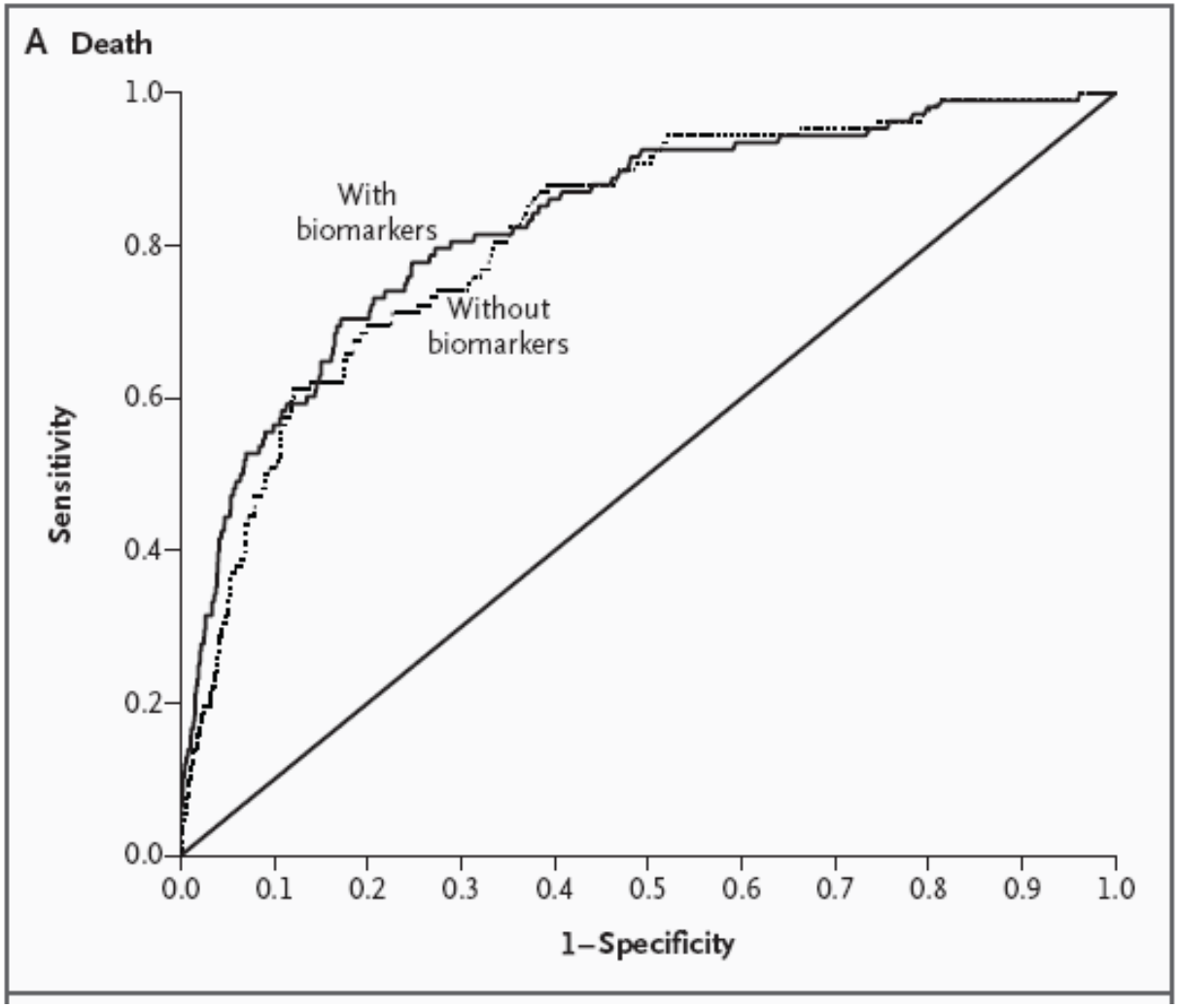


Ideal calibration

O:E = 1

Slope = 1

Model to predict cardiovascular outcomes – added value biomarkers?



AUC 0.76

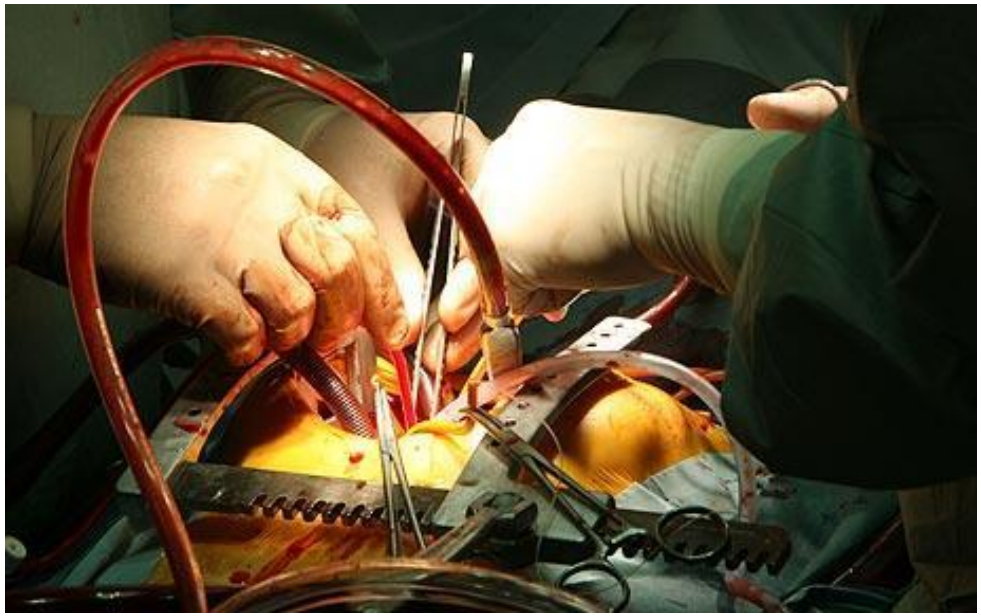
AUC 0.77

Wang TJ, et al. NEJM

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

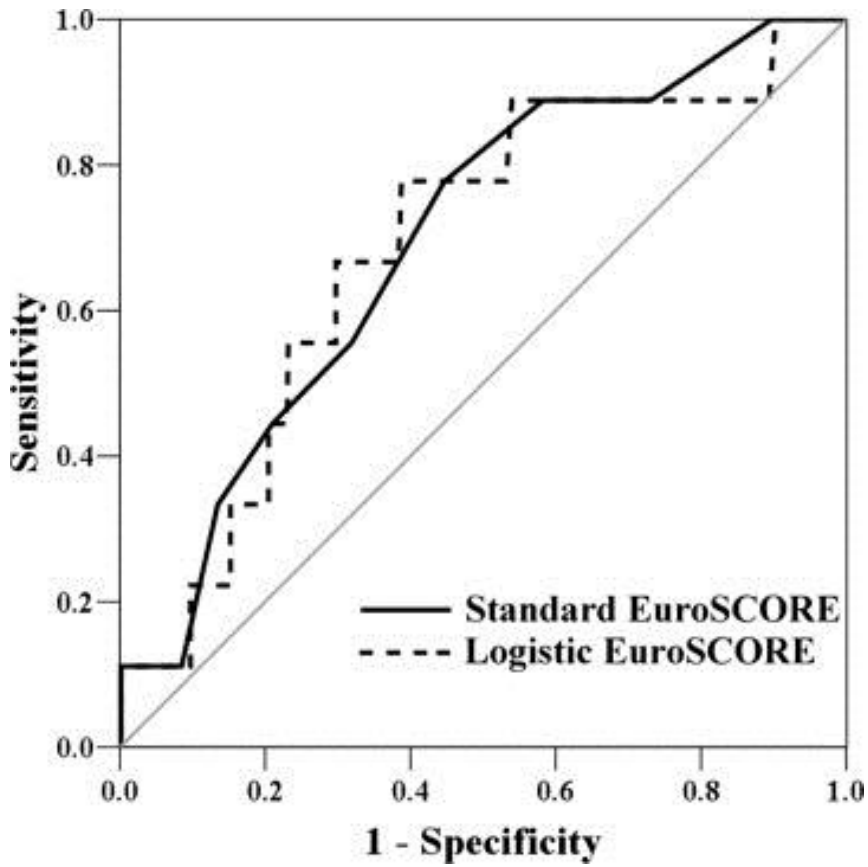
Service de chirurgie cardiovasculaire, CHU de Fort de France, 97200 Martinique, France

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

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

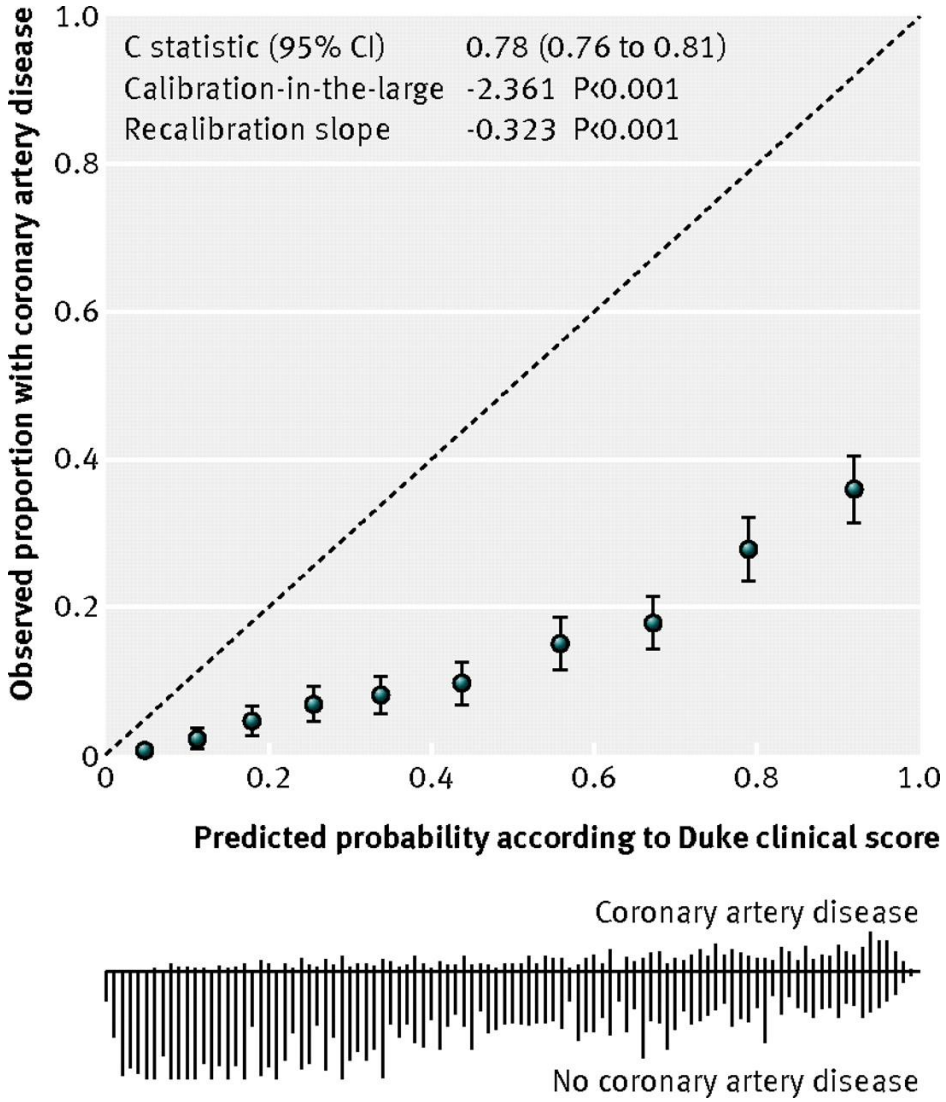
How well does the standard EuroSCORE calibrate?

(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

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

ELSEVIER

CrossMark

Journal of Clinical Epidemiology

Journal of Clinical Epidemiology 68 (2015) 279–289

ORIGINAL ARTICLES

A new framework to enhance the interpretation of external validation studies of clinical prediction models

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

[a]*Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Str. 6.131, PO Box 85500, 3508GA Utrecht, The Netherlands*

[b]*Department of Public Health, Erasmus Medical Center, Rotterdam, The Netherlands*

Accepted 30 June 2014; Published online 30 August 2014

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

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)

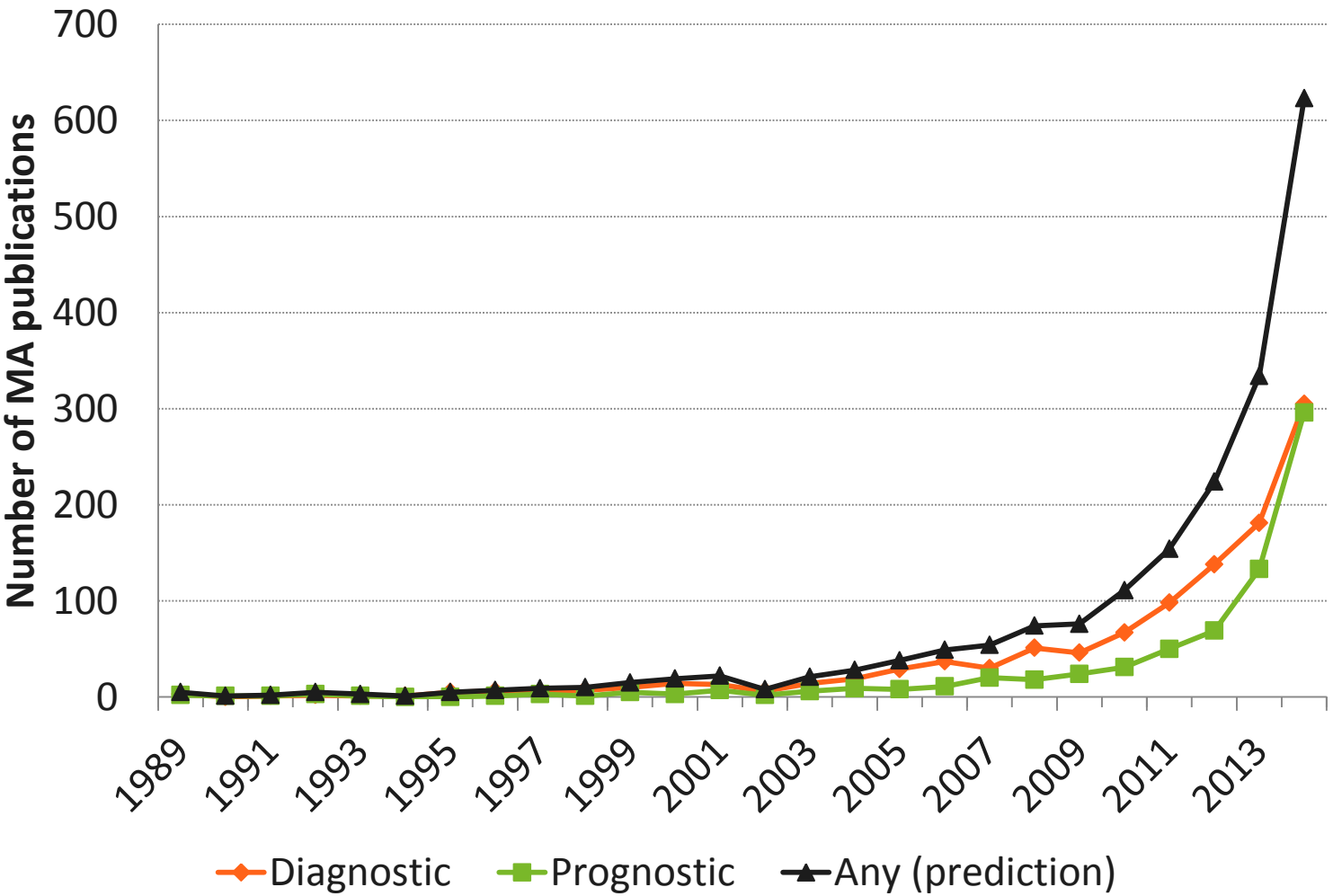
EDITORIALS

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 increasingly popular



Advantages of meta-analysis

- Increase precision
- Resolve inconsistencies
- Explore sources of heterogeneity, e.g.:
  - Under what conditions does a model yield adequate performance?
  - In which patient subgroups does a predictor provide added value to an existing model?
- Improve generalizability of a novel prediction model
- …

Three types of MA in prediction research

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

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

3. **In case of multiple IPD sets – IPD meta-analysis**

Three types of MA in prediction research

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

What statistics can we summarize when reviewing external validation studies?

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

What statistics can we summarize?

- Overall performance
- Model discrimination
- Model calibration

Overall performance

Statistics

- 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.

Discrimination

Quantifies the model's extent 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

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

What about other performance measures?

Model fit

- Maximum likelihood (and derivatives such as AIC, BIC) are not suitable for pooling as their magnitude depends on the sample size of individual studies
- Results from the Hosmer-Lemeshow test are also not suitable for pooling, as the test statistic again depends on the sample size and often remains unreported.

Meta-analysis principles

Recap

- Fixed effect meta-analysis
  - Assumes common performance for all studies
  - Variation in observed study estimates is due only to chance
- Random effects meta-analysis
  - Variation in observed performance is due to chance and between-study heterogeneity

Fixed or random effects?

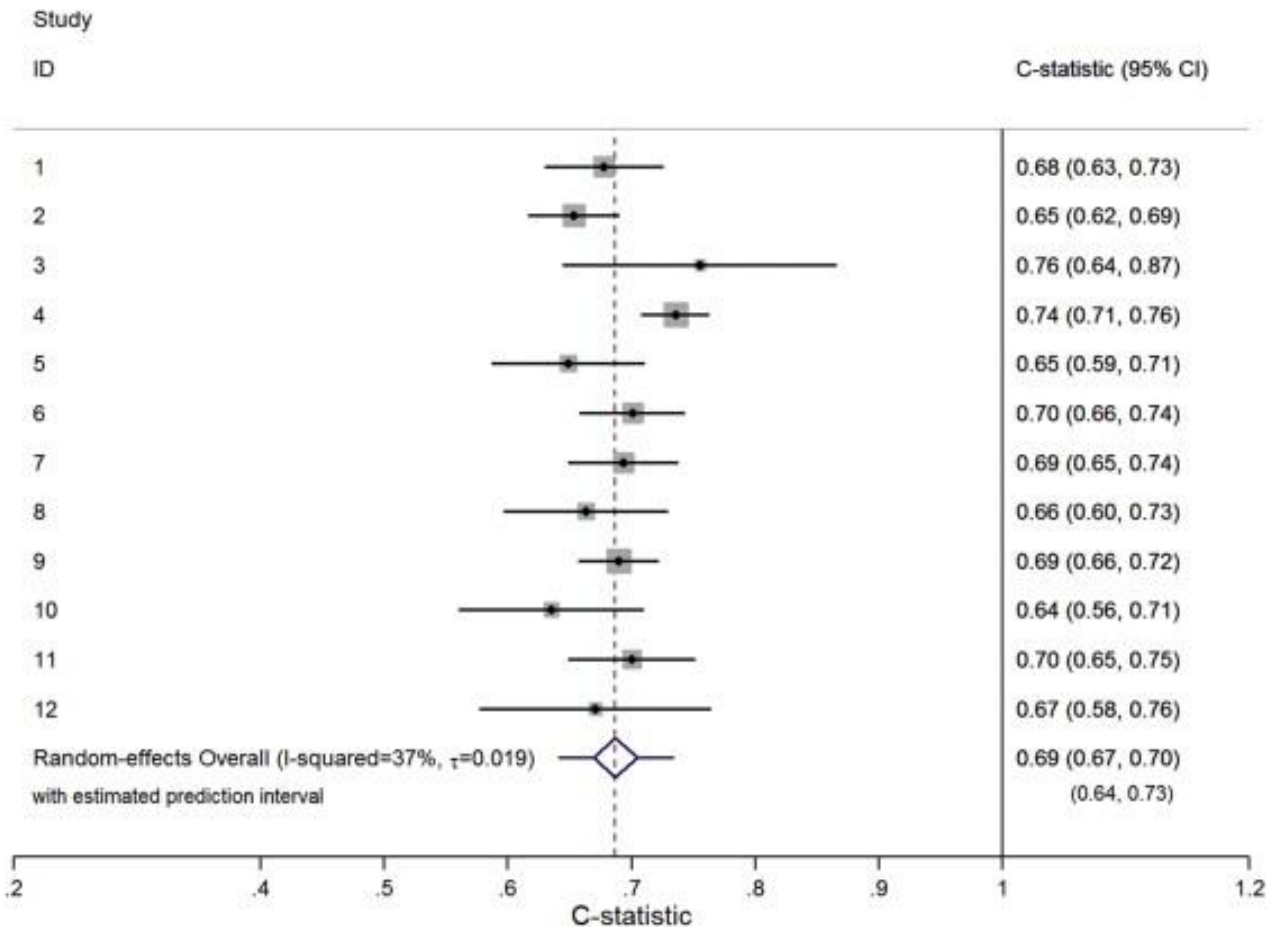
- Assumption of homogeneity (fixed effect) often unrealistic
- Ignoring heterogeneity leads to an overly precise summary result
- Summary estimates of performance have limited usefulness when there is strong heterogeneity

Recommendation: allow for random effects and calculate a prediction interval

$$\hat{\mu} \pm t_{k-2}\sqrt{\hat{\tau}^2 + SE(\hat{\mu})^2}$$

Ref: Riley et al. Interpretation of random effects meta-analyses. BMJ 2010.

Prediction interval



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

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%)
- I^2 can directly be compared between meta-analyses with different number of studies and different types of outcome data
- I^2 is preferable to a test for heterogeneity in judging consistency of evidence

Ref: Higgins et al. Measuring inconsistency in meta-analyses. BMJ 2003.

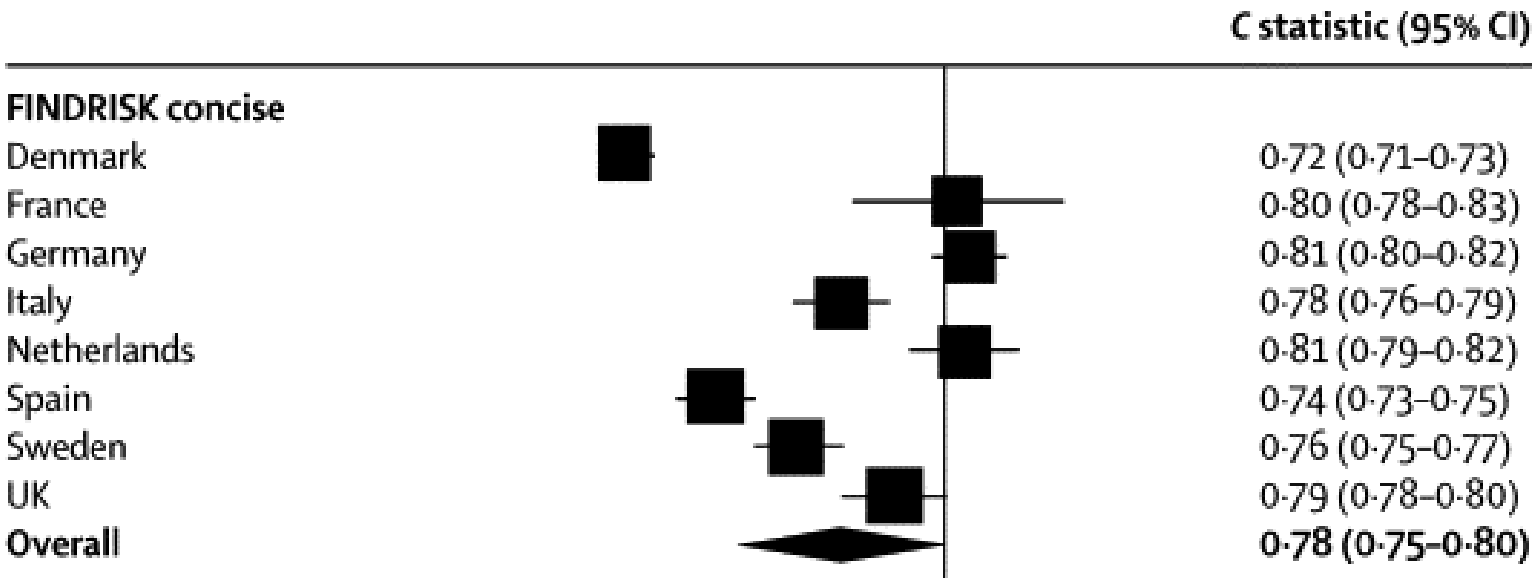
Quantifying heterogeneity

I^2 value	Guide to Interpretation
0% to 40%	Might not be important
30% to 60%	May represent moderate heterogeneity *
50% to 90%	May represent substantial heterogeneity *
75% to 100%	Considerable heterogeneity *

Importance of I^2 value depends on

- Magnitude and direction of effects
- Strength of evidence of heterogeneity
  - Chi-squared P value, or
  - I^2 confidence interval

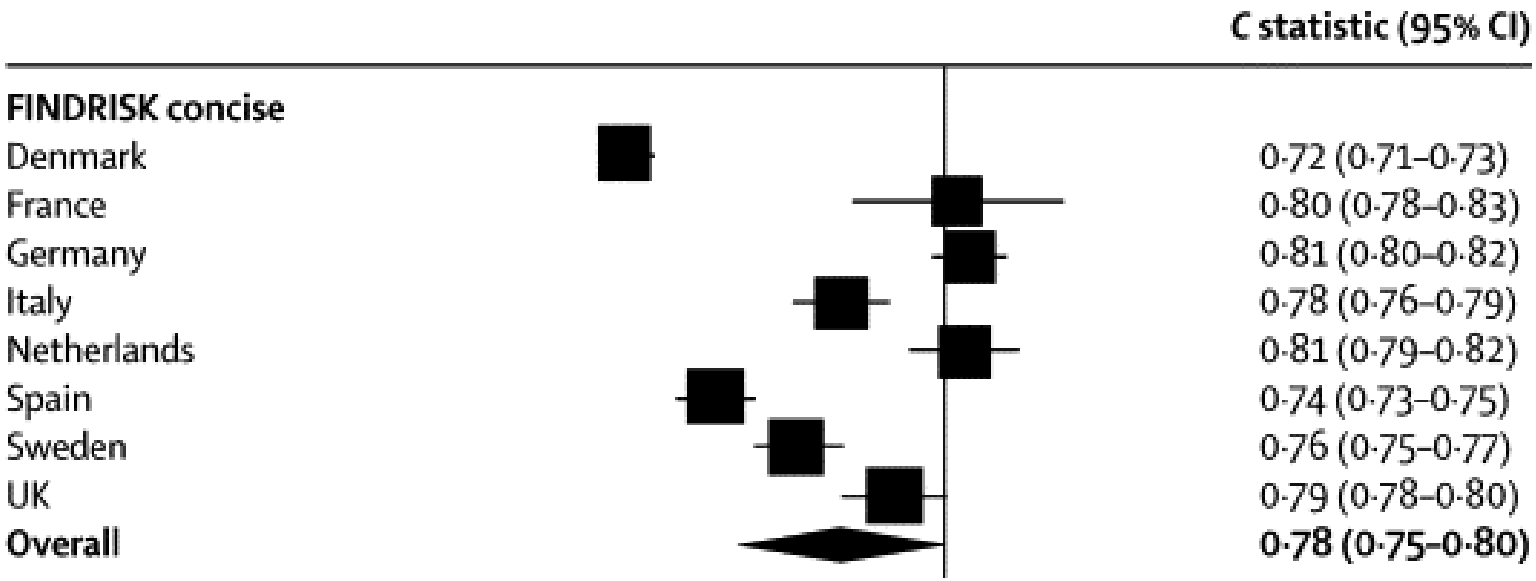
Quantifying heterogeneity



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.

Quantifying heterogeneity

$I^2 = 98\%$



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.

Example

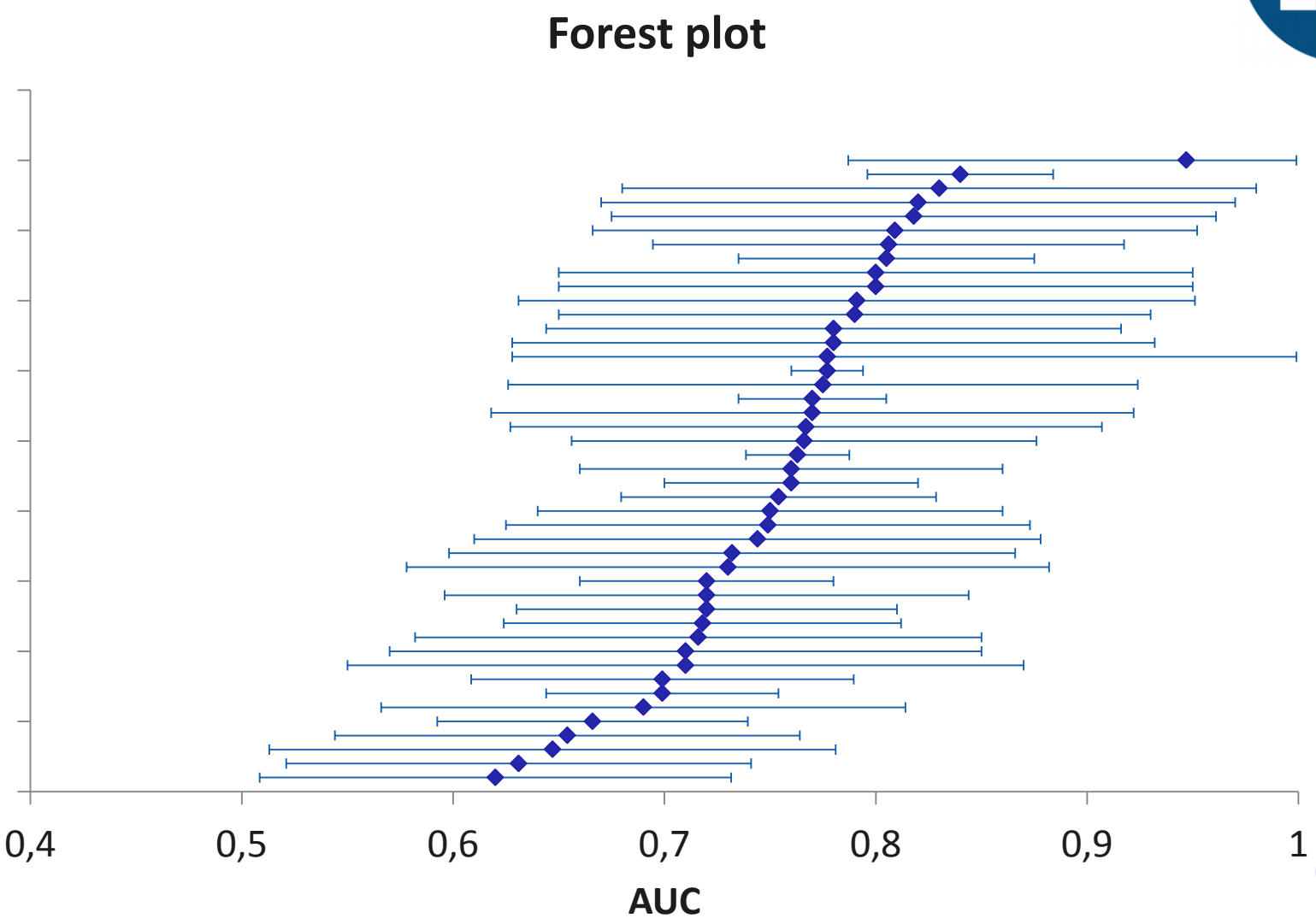
Meta-analysis of the EuroSCORE model

45 published validation studies with information on:

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

Ref: Siregar et al. Performance of the original EuroSCORE. Eur J Cardiothorac Surg 2012.

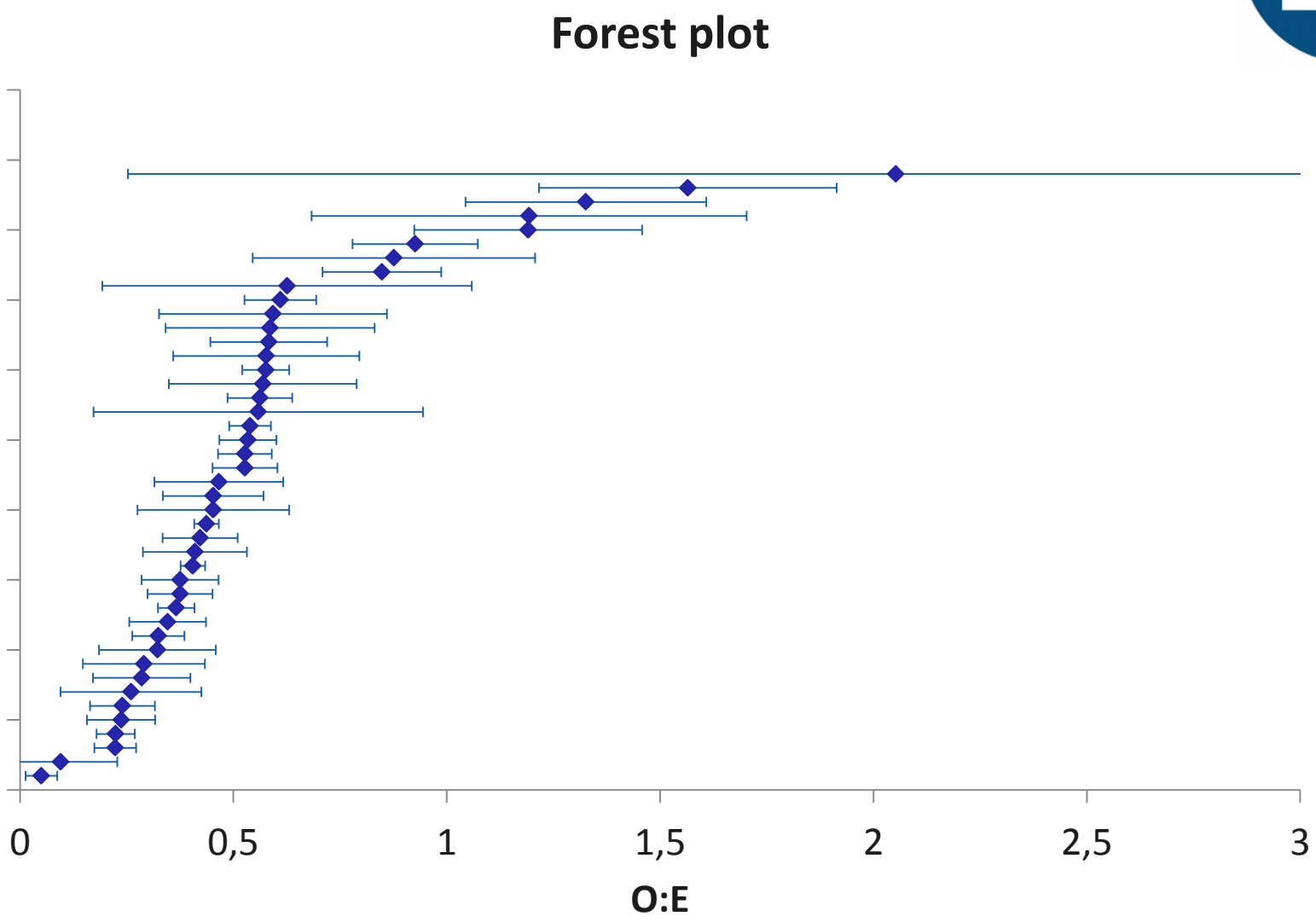
Discrimination



Results meta-analysis of AUC

- Pooled estimate: **0.7516**
- Standard error: 0.0089
- Std. dev. between studies (τ): 0.0318
- 95% confidence interval: 0.73 – 0.77
- 95% prediction interval: 0.69 – 0.82
- I^2 statistic: **32.3%**
- Cochran Q-test for heterogeneity: p-value = 0.0216

Calibration



Results meta-analysis of O:E

- Pooled estimate: **0.5205**
- Standard error: 0.0438
- Std. dev. between studies (τ): 0.2748
- 95% confidence interval: 0.43 – 0.61
- 95% prediction interval: 0.00 – 1.07
- I^2 statistic: **95.3%**
- Cochran Q-test for heterogeneity: p-value = 0.0000

Meta-regression 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

Results meta-regression of AUC

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)
- I^2 statistic: 1%
- Cochran Q-test for heterogeneity: p-value = 0.5299

Results meta-regression of O:E

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)
- I^2 statistic: 93.4%
- Cochran Q-test for heterogeneity: p-value = 0.0000

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

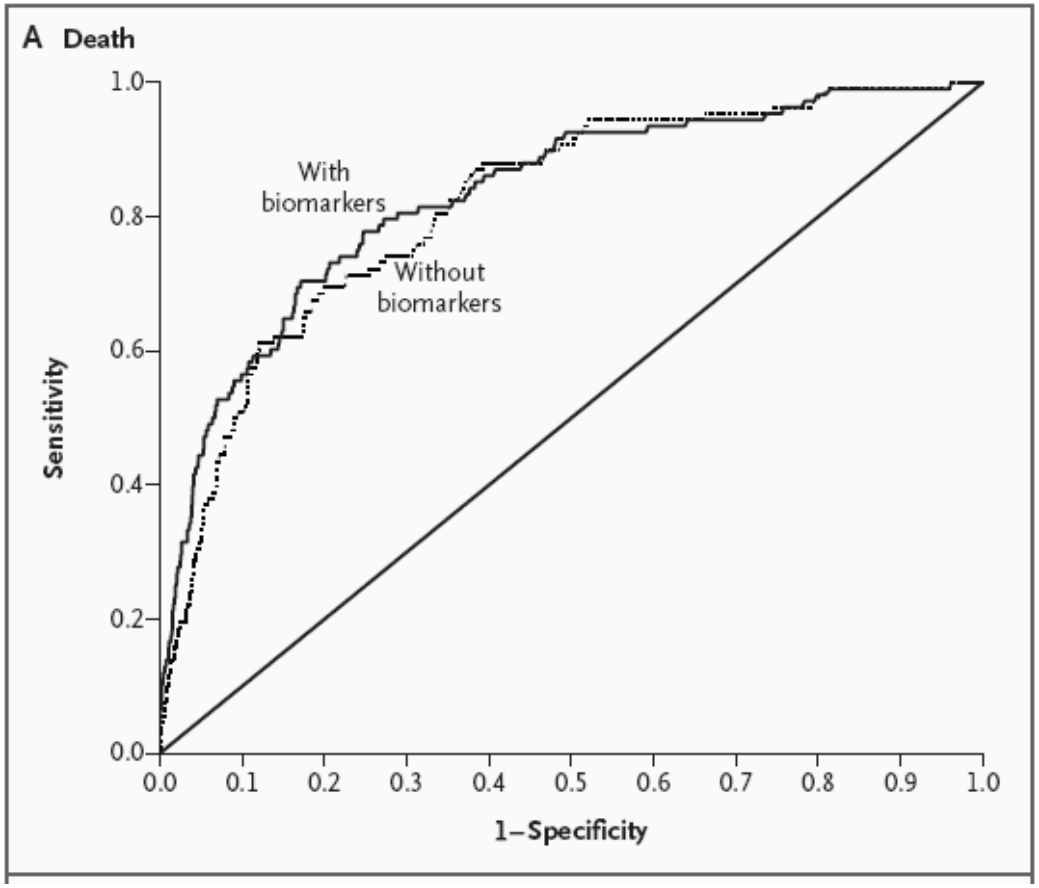
What statistics can we summarize when reviewing added-value studies?

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

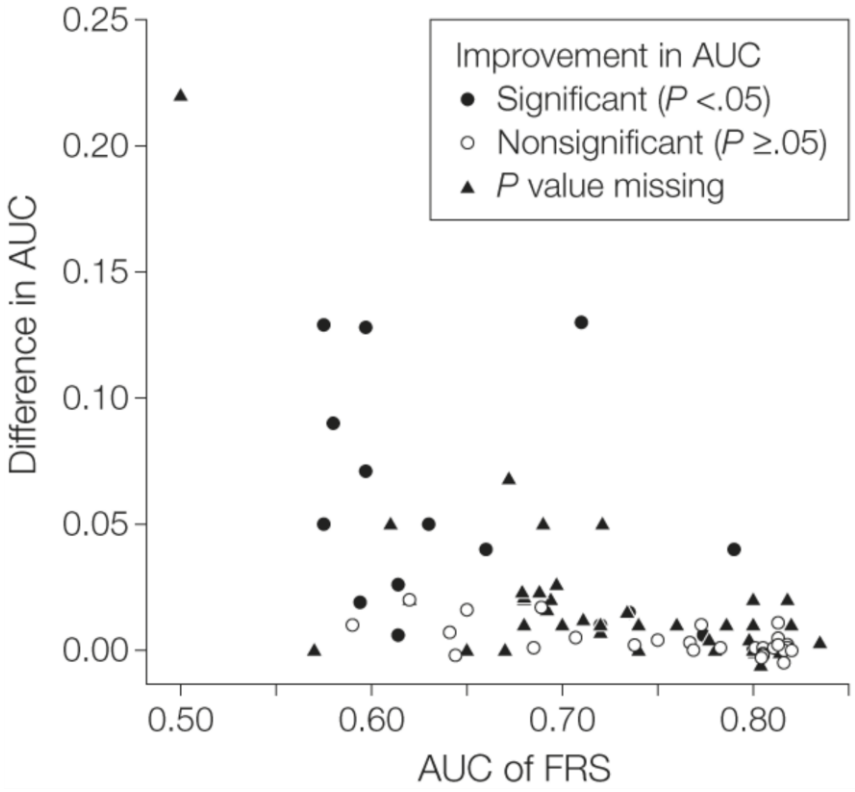
Systematic review of studies that …

- … 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

Recall: three types of MA

SR and MA of prediction models

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 data and IPD**

3. **In case of multiple IPD sets – IPD meta-analysis**

Combination of aggregate data and IPD

Three types of aggregate data

1. Reported univariable associations
2. Published prediction models with similar predictors
3. Published prediction models with different predictors

Goal

- Synthesize evidence on prognostic factors
- Combine evidence from aggregate data and IPD into a meta-model

Combination of aggregate data and IPD

Not discussed in this workshop

More information available online:

- Debray TPA *et al*. Incorporating published univariable associations in diagnostic and prognostic modeling. BMC Med Res Methodol 2012.
- Debray TPA *et al*. Aggregating published prediction models with individual participant data: a comparison of different approaches. Stat Med 2012.
- Debray TPA *et al*. Meta-analysis and aggregation of multiple published prediction models. Stat Med 2014.
- Steyerberg EW *et al*. Prognostic models based on literature and individual patient data in logistic regression analysis. Stat Med 2000.

Recall: three types of MA

SR and MA of prediction models

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 data and IPD**

3. **In case of multiple IPD sets – IPD meta-analysis**

IPD Meta-analysis

Discussed in workshop tomorrow (14h – Galerie 13/14)

More information available online:

- Debray *et al*. Individual Participant Data (IPD) Meta-analyses of Diagnostic and Prognostic Modeling Studies: Guidance on Their Use. PLOS Med 2015.
- Debray *et al*. A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. Stat Med 2013.
- Pennells *et al*. Assessing risk prediction models using individual participant data from multiple studies. Stat Med 2014.
- Royston *et al*. Construction and validation of a prognostic model across several studies, with an application in superficial bladder cancer. Stat Med 2004.

Advanced topics

Use of appropriate meta-analysis models

- Traditional meta-analysis methods assume normality of test statistics within and between studies
- Potential to misleading prediction intervals of model performance, and to biased summary estimates
- Alternative methods
  - Canonical transformations
  - Variance stabilizing transformations
  - Exact methods

Advanced topics
Canonical transformation

- Change the 'spacing' near the extremes
- Sample variance remains a function of the sample mean

Formula

- Discrimination: $\widehat{\boldsymbol{\theta}}_{\boldsymbol{j}} = \log\left(\frac{AUC_j}{1-AUC_j}\right)$ and $\widehat{\boldsymbol{\sigma}}_{\boldsymbol{j}}^{\boldsymbol{2}} = \frac{\mathrm{var}(\boldsymbol{AUC_j})}{\left(\boldsymbol{AUC_j}(\boldsymbol{1}-\boldsymbol{AUC_j})\right)^{\boldsymbol{2}}}$
- Calibration: $\widehat{\boldsymbol{\theta}}_{\boldsymbol{j}} = \log\left(\frac{O_j}{E_j}\right)$ and $\widehat{\boldsymbol{\sigma}}_{\boldsymbol{j}}^{\boldsymbol{2}} = \frac{\boldsymbol{1}}{\boldsymbol{O_j}}$

Ref: Van Klaveren et al. Assessing discriminative ability of risk models in clustered data. BMC Med Res Methodol 2014.

Advanced topics

Variance stabilizing transformation

Formula

- Discrimination: $\widehat{\boldsymbol{\theta}}_{\boldsymbol{j}} = \sin^{-1}\left(\sqrt{\boldsymbol{AUC_j}}\right)$ and $\widehat{\boldsymbol{\sigma}}_{\boldsymbol{j}}{}^{\boldsymbol{2}} = \frac{\mathrm{var}(\boldsymbol{AUC_j})}{4(\boldsymbol{1}-\boldsymbol{AUC_j})\boldsymbol{AUC_j}}$
- Calibration: $\widehat{\boldsymbol{\theta}}_{\boldsymbol{j}} = \sqrt{\frac{\boldsymbol{O_j}}{\boldsymbol{E_j}}}$ and $\widehat{\boldsymbol{\sigma}}_{\boldsymbol{j}}{}^{\boldsymbol{2}} = \frac{\boldsymbol{1}}{\boldsymbol{4E}}$

  Variance is now independent of estimated mean

Advanced topics

Approximate meta-analysis methods

Recommendations

- Estimates of calibration slope and calibration-in-the-large do not require transformation
- Estimates of AUC and O:E ratio should be transformed when using approximate methods
  - Canonical transformations are more reliable, but may still lead to bias in extreme scenarios
  - Further research warranted for variance stabilizing transformations

Advanced topics

Statistics of interest often poorly reported

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.

Advanced topics

How to obtain the O:E statistic?

- Extract O and E separately
  - Number of events
  - Risk tables
- Calculate E using mean values of patient characteristics
- Calculate O:E ratio from calibration-in-the-large

Restoring the standard error

- Transform reported confidence interval or p-values
- Use the Delta method (applying Poisson approximations)

Advanced topics

How to obtain the AUC?

The AUC is the most common measure in validation studies. However, it is often reported using inconsistent terminology.

- Area under the receiver operating characteristic curve
- Area under the ROC curve
- AUROC
- Concordance index (C index)
- Concordance statistic (C statistic; C-statistic)

Advanced topics
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

ARTICLE IN PRESS

ELSEVIER

Journal of Clinical Epidemiology ■ (2015) ■

Journal of Clinical Epidemiology

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,*]

[a]Public Health, Epidemiology and Biostatistics, School of Health and Population Sciences, Public Health Building, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

[b]School of Mathematics, Watson Building, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

[c]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Str. 6.131, PO Box 85500, 3508 GA Utrecht, The Netherlands

[d]Dutch Cochrane Centre, University Medical Center Utrecht, Str. 6.131, PO Box 85500, 3508 GA Utrecht, The Netherlands

[e]Research Institute for Primary Care and Health Sciences, Keele University, Staffordshire ST5 5BG, UK

[f]Department of Medical Oncology, Erasmus MC Cancer Institute, Erasmus University Medical Center, PO Box 2040, 3000 CA Rotterdam, The Netherlands

Accepted 8 May 2015; Published online xxxx

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?

Take home messages

Systematic review & meta-analysis of prediction models

- Step 1: summarize performance of existing models
- Step 2: identify which models are most promising for target population and interpret their generalizability
- Step 3: combine most promising models and tailor them to population at hand

When no relevant models are available, IPD is needed to develop a new model

Handy Tools/Papers

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

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)