UMC Utrecht
Julius Center

Meta-analysis of prediction models

combining aggregate and individual participant data

Thomas Debray, Hans Reitsma, Karel Moons, Richard Riley

for the Cochrane IPD Meta-analysis Methods Group

(Co-convenors: Jayne Tierney, Mike Clarke, Lesley Stewart, Maroeska Rovers)

Keele University

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

University Medical Center Utrecht

Conflict of interest

We have developed and validated several multivariable prediction models.

We performed several individual patient data meta-analyses, in addition to methodological work

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

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 (predictive factor)

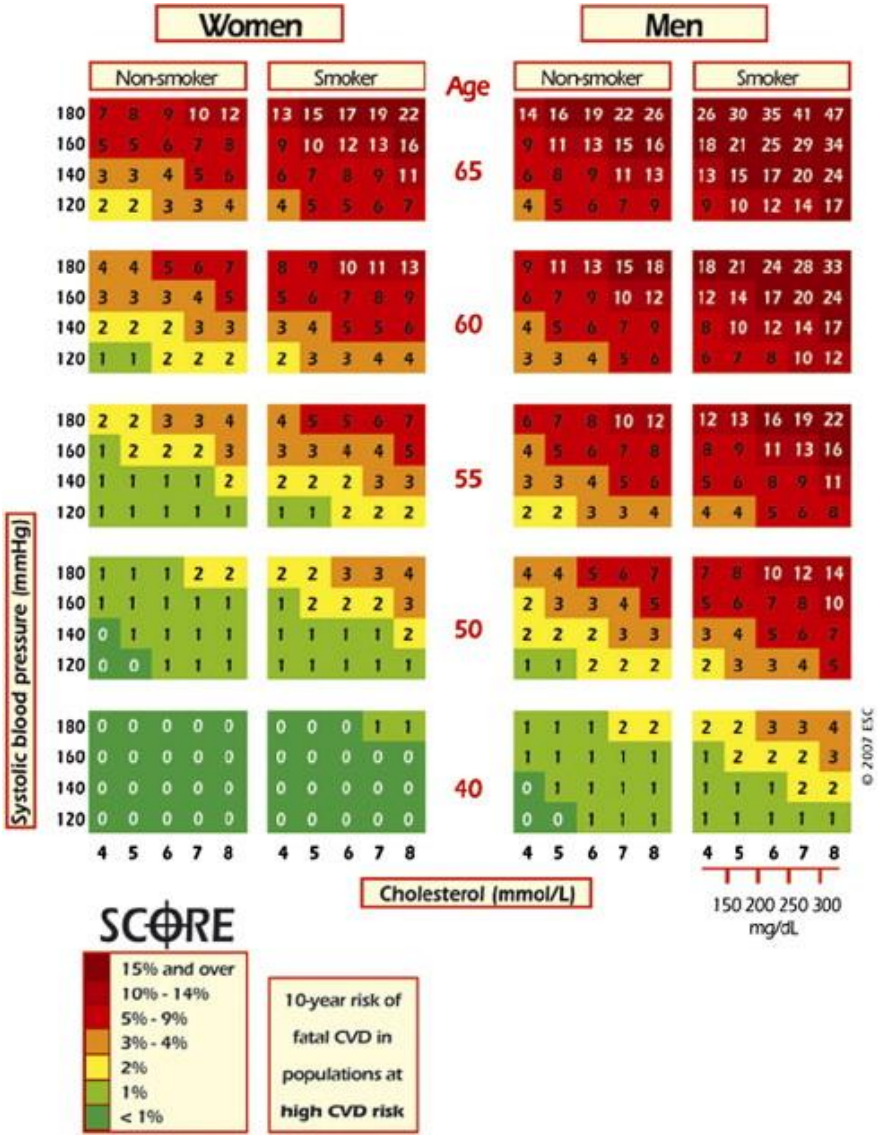
Four main types of prognosis studies
PROGRESS series 2013: BMJ and Plos Med

- 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

Prediction models

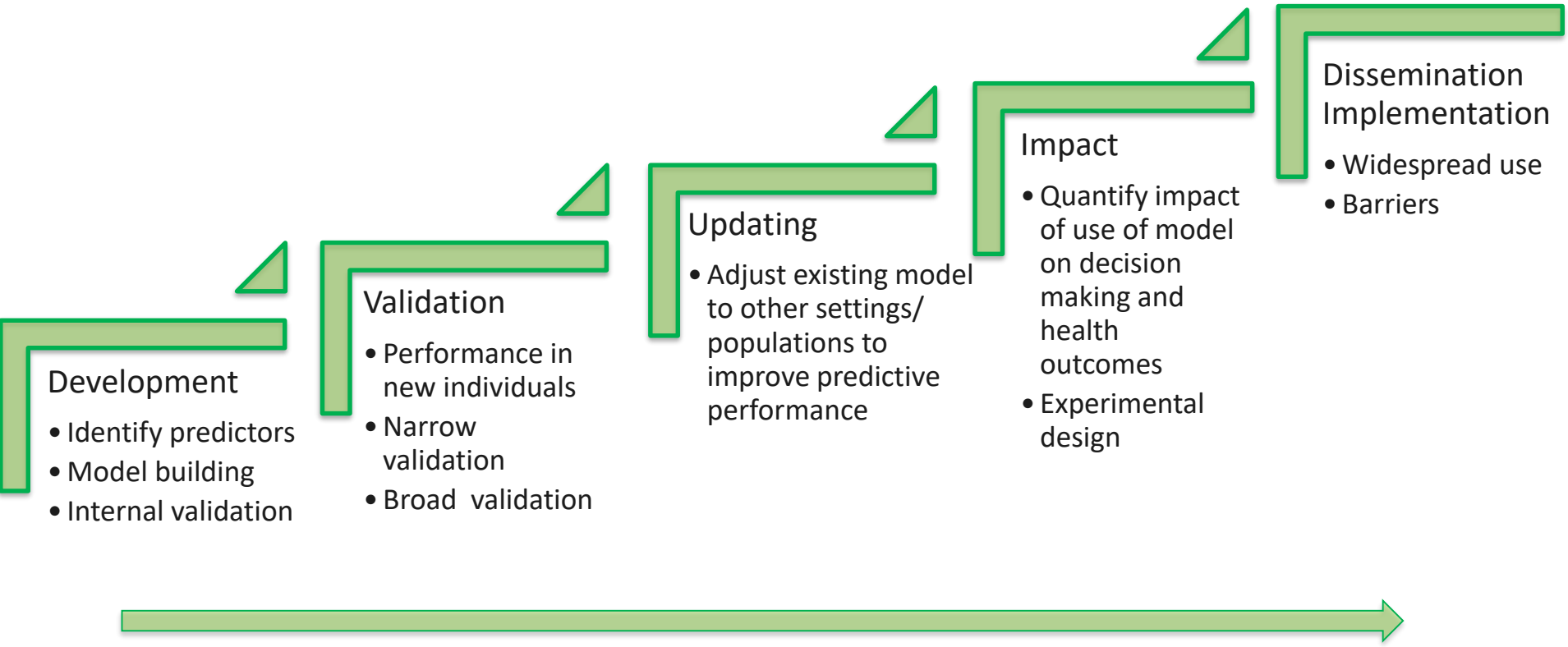
APGAR
Test Scoring

	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



Phases of prediction model evaluation

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



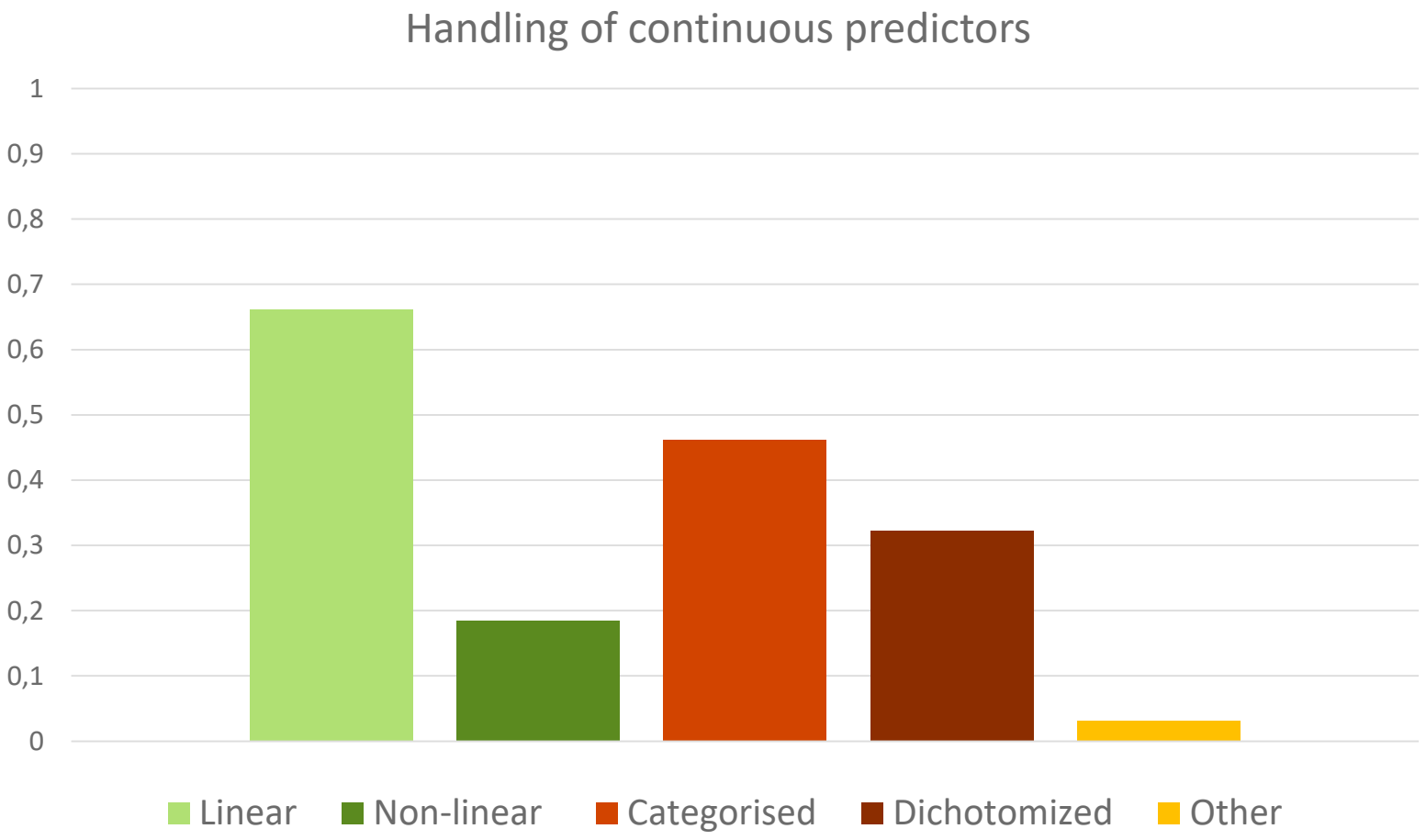
Increasing level of evidence for use of model in practice

Prediction models

Common problems

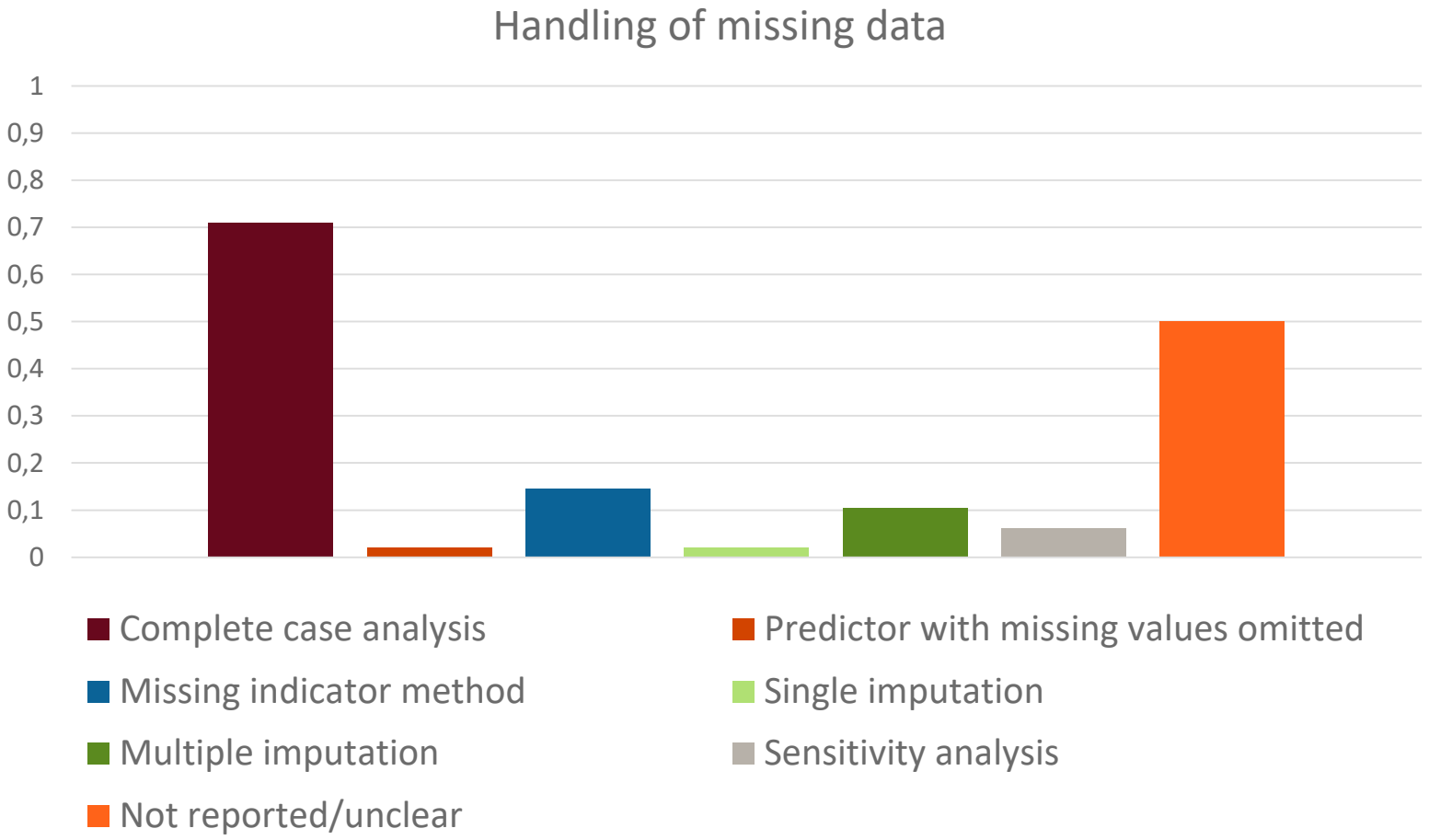
- Poor statistical methodology
- Poor predictive accuracy
  - Over-optimism
  - Lack of transportability

Problem 1: Poor methodology



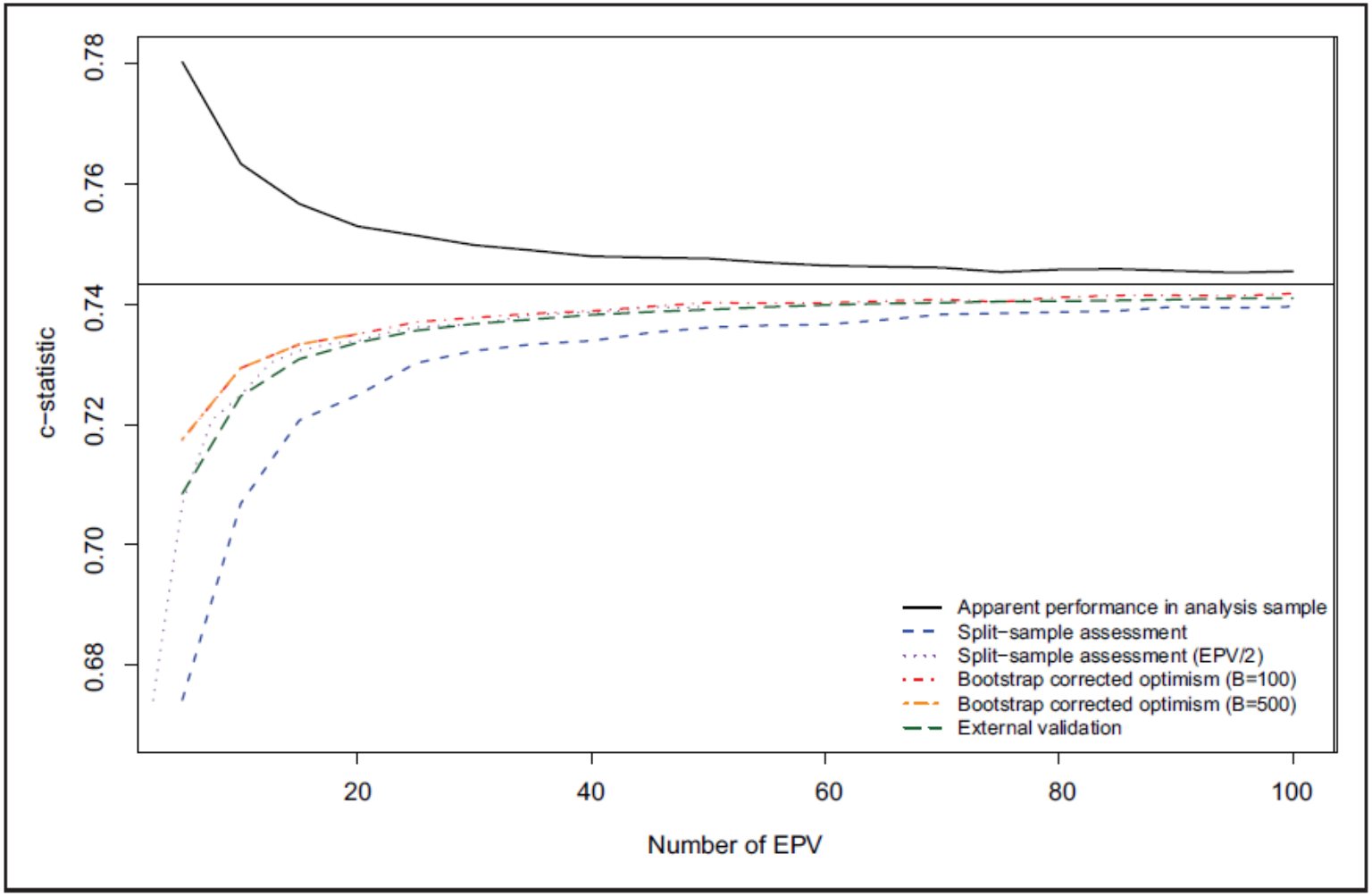
Ref: Bouwmeester W *et al.* Reporting and methods in clinical prediction research: a systematic review. PLoS Med. 2012.

Problem 1: Poor methodology



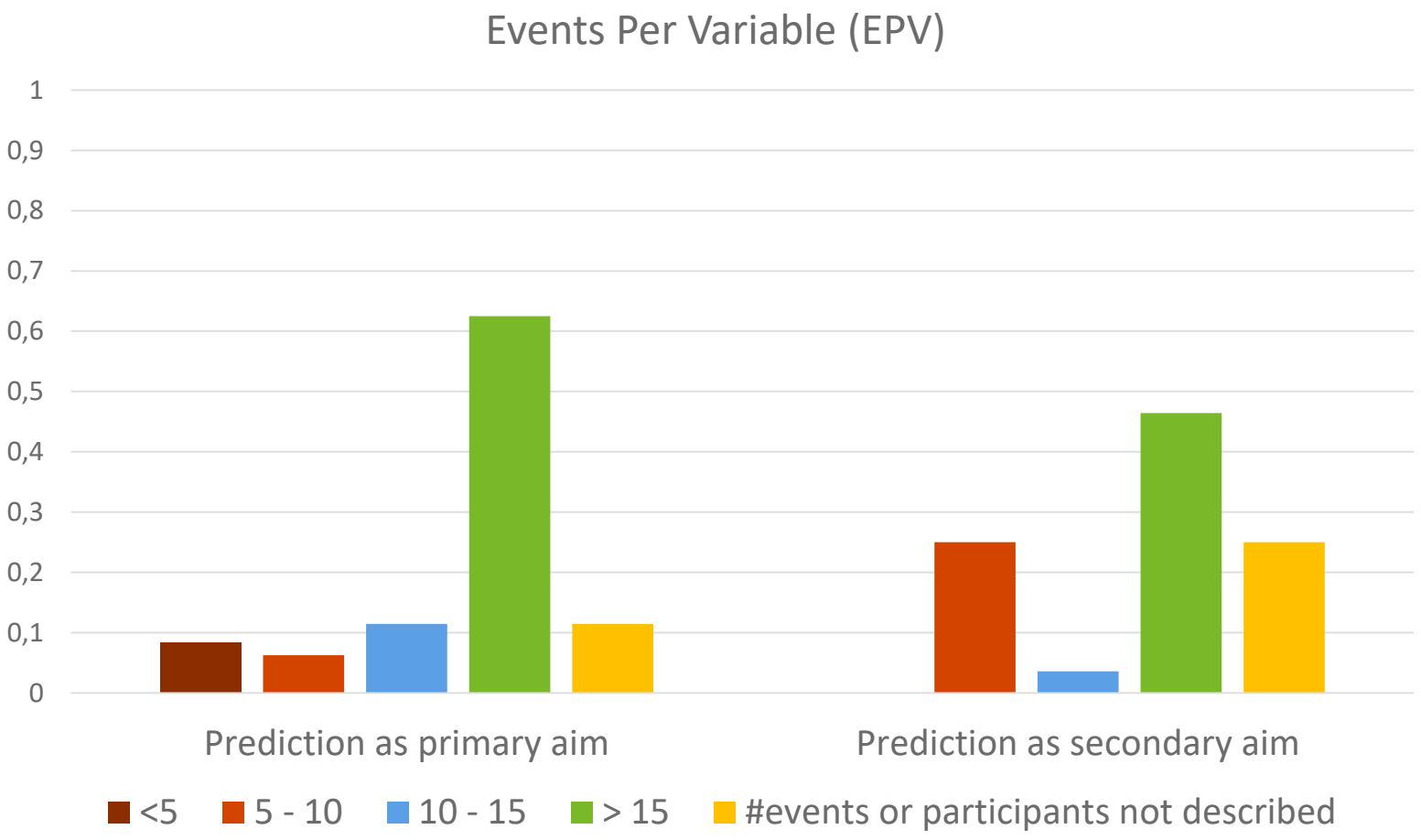
Ref: Bouwmeester W *et al.* Reporting and methods in clinical prediction research: a systematic review. PLoS Med. 2012.

Problem 2: Over-optimism



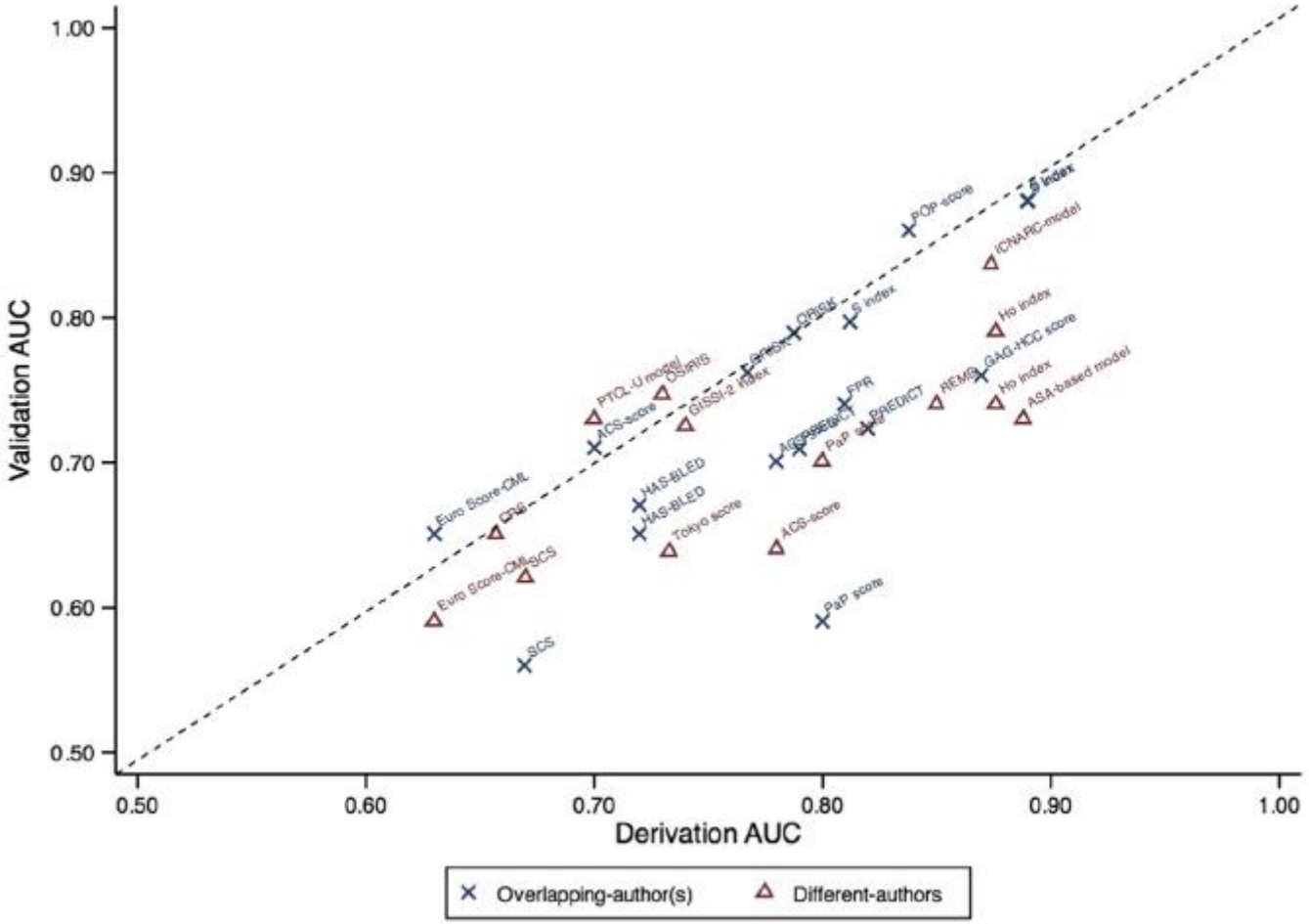
Ref: Austin PC *et al*. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. Stat Methods Med Res. 2014.

Problem 2: Over-optimism



Ref: Bouwmeester W *et al.* Reporting and methods in clinical prediction research: a systematic review. PLoS Med. 2012.

Problem 3: Lack of transportability



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

Numerous models for same target population + outcomes

Prior evidence not optimally used

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

Diagnosis of Deep Vein Thrombosis

Clinical feature	Score
Active cancer (treatment ongoing or within previous 6 months or palliative)	1
Paralysis, paresis, or recent plaster immobilisation of the lower extremities	1
Recently bedridden for more than 3 days or major surgery, within 4 weeks	1
Localised tenderness along the distribution of the deep venous system	1
Entire leg swollen	1
Calf swelling by more than 3 cm when compared with the asymptomatic leg (measured 10 cm below tibial tuberosity)	1
Pitting oedema (greater in the symptomatic leg)	1
Collateral superficial veins (non-varicose)	1
Alternative diagnosis as likely or greater than that of deep-vein thrombosis	−2

In patients with symptoms in both legs, the more symptomatic leg is used.

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)).

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

Numerous models for same target population + outcomes

Diagnosis of 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

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.

Numerous models for same target population + outcomes

Systematic review

1. Formulating review question
2. Formulating search strategy
3. Critical appraisal
4. Quantitative data extraction
5. Quantitative data synthesis

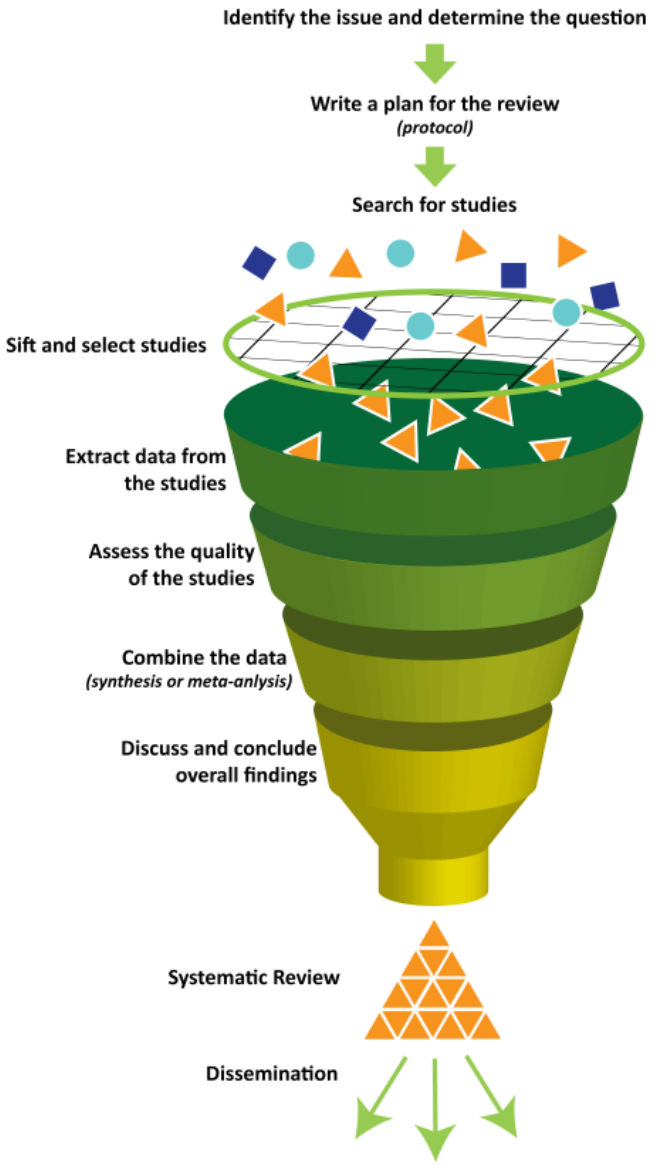


Image source: http://cccrg.cochrane.org/infographics

Numerous models for same target population + outcomes

<u>Focus of today</u>: using prior evidence when developing a new prediction model
= combining individual participant and aggregate data

Different types of aggregate data

- Univariable regression coefficients (or unadjusted odds/hazard ratios)
- Multivariable regression coefficients (or adjusted odds/hazard ratios)
- Complete regression models (or score charts)
- Regression trees, neural networks, …

Type 1
Incorporating unadjusted predictor effects

UMC Utrecht

Background

- Many publications for a particular clinical problem
  - Description of patient characteristics and outcome(s)
  - Reported information often sufficient to calculate an unadjusted regression coefficient

- Possible to take advantage of univariable literature data!

How could we make use of published unadjusted odds ratios when developing a logistic regression model?

Background

1987

Greenland S. Quantitative methods in the review of epidemiologic literature

2000

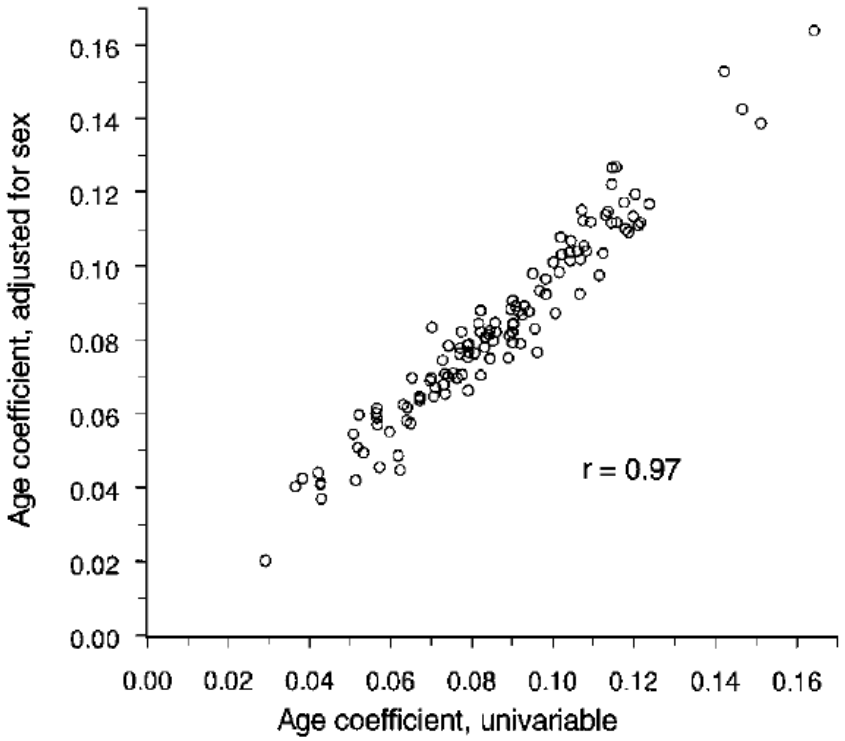
Steyerberg EW et al. Prognostic models based on literature and individual patient data in logistic regression analysis

2012

Debray TPA et al. Incorporating published univariable associations in diagnostic and prognostic modeling

Background

Strong correlation between *univariable* and *multivariable* regression coefficient



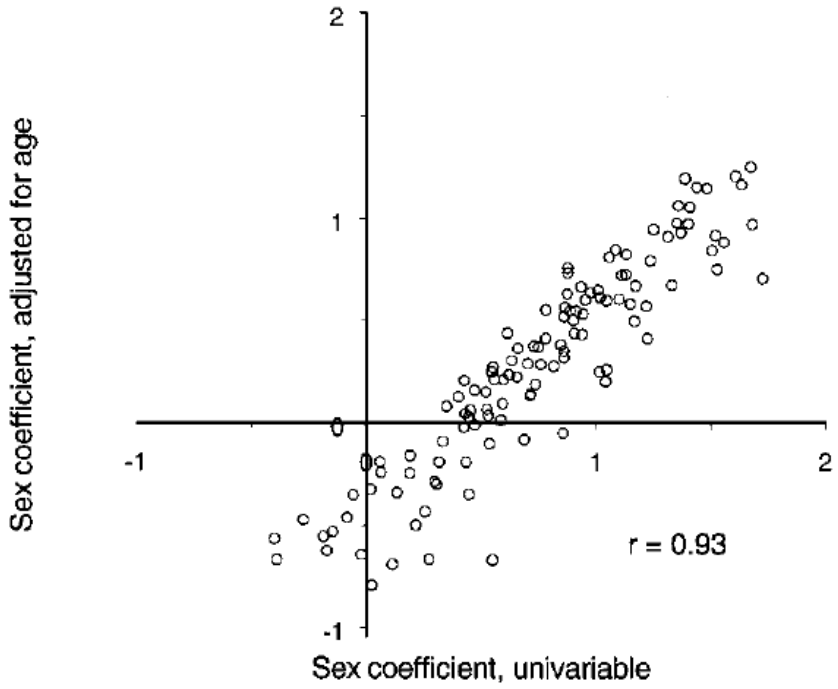


Figure 1. Univariable and multivariable regression coefficients in the two-predictor model consisting of age and sex estimated in 121 small subsamples of the GUSTO-I data set

The adaptation method

Required steps

1. Extract the *univariable* regression coefficients
2. Summarize the extracted coefficients
3. Estimate the change from *univariable* to *multivariable* coefficient in "own" data set
4. Use estimated change to transform the pooled *univariable* coefficient into a *multivariable* coefficient

The adaptation method

Steyerberg (2000)

- The *univariable* coefficients are summarized using fixed effects meta-analysis
- The updated *multivariable* coefficient and its SE can be approximated using simple equations:

$$\hat{\beta}_{m|L} = \hat{\beta}_{m|I} + c\left(\hat{\beta}_{u|L} - \hat{\beta}_{u|I}\right)$$

$$\widehat{\mathrm{Var}}\left(\hat{\beta}_{m|L}\right) = \widehat{\mathrm{Var}}\left(\hat{\beta}_{u|L}\right) + \left[\widehat{\mathrm{Var}}\left(\hat{\beta}_{m|I}\right) - \widehat{\mathrm{Var}}\left(\hat{\beta}_{u|I}\right)\right]$$

m = multivarible, u = univariable, L = literature, I = "own" IPD

The adaptation method

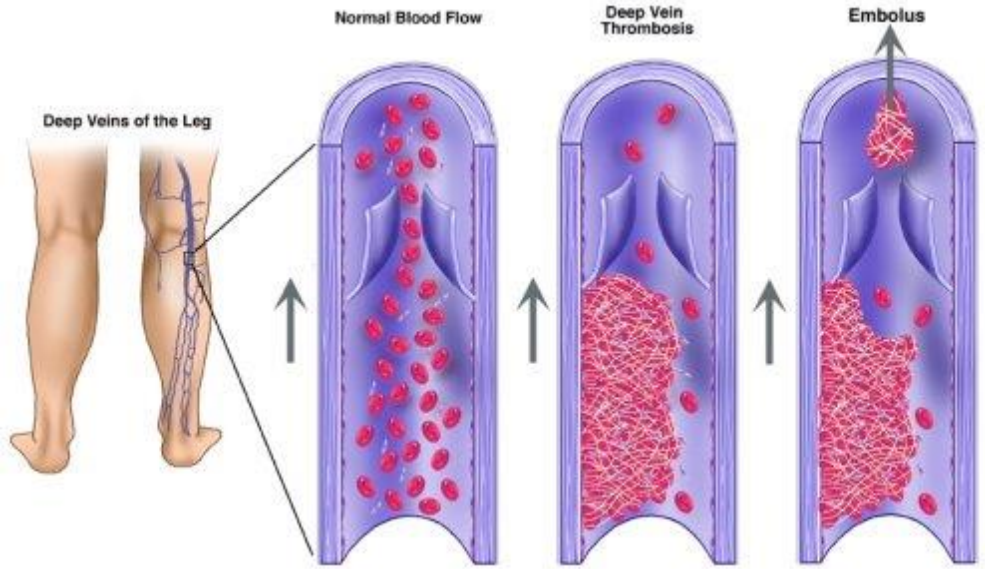
Extensions by Debray et al. (2012)

- Summarize *univariable* coefficients using random effects meta-analysis
- Apply penalization to estimate the change from *univariable* to *multivariable* coefficient
- Apply bootstrap procedure to estimate SE of updated *multivariable* regression coefficient

Case study

Diagnosis of Deep Vein Thrombosis

- Outcome: presence of DVT
- Population: patients suspected of DVT
- Literature: 7 studies reporting D-dimer test results
- Model development dataset (N=1295)
- External validation dataset (N=1756)



Case study

Diagnosis of Deep Vein Thrombosis

Model development

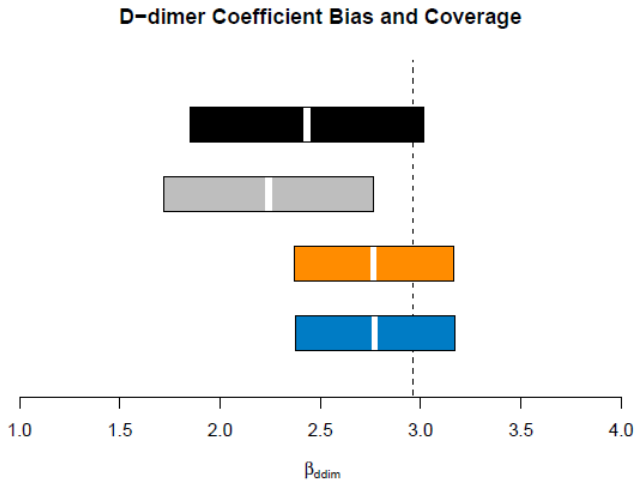
- Logistic regression in “own” IPD (●)
- Penalized logistic regression in “own” IPD (●)
- Steyerberg Adaptation method (●)
- Debray Adaptation method (●)

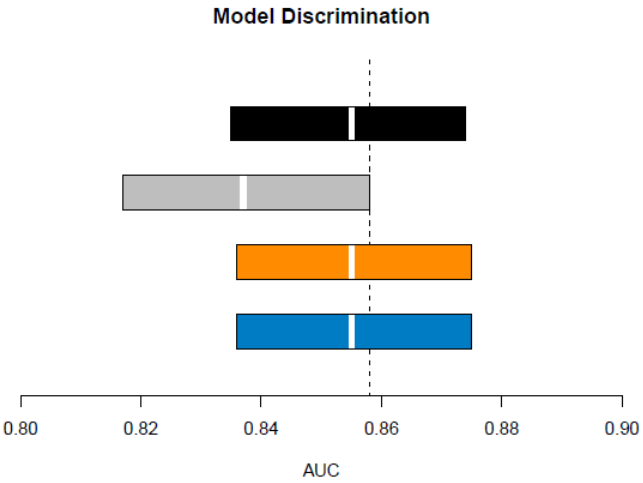
Model validation

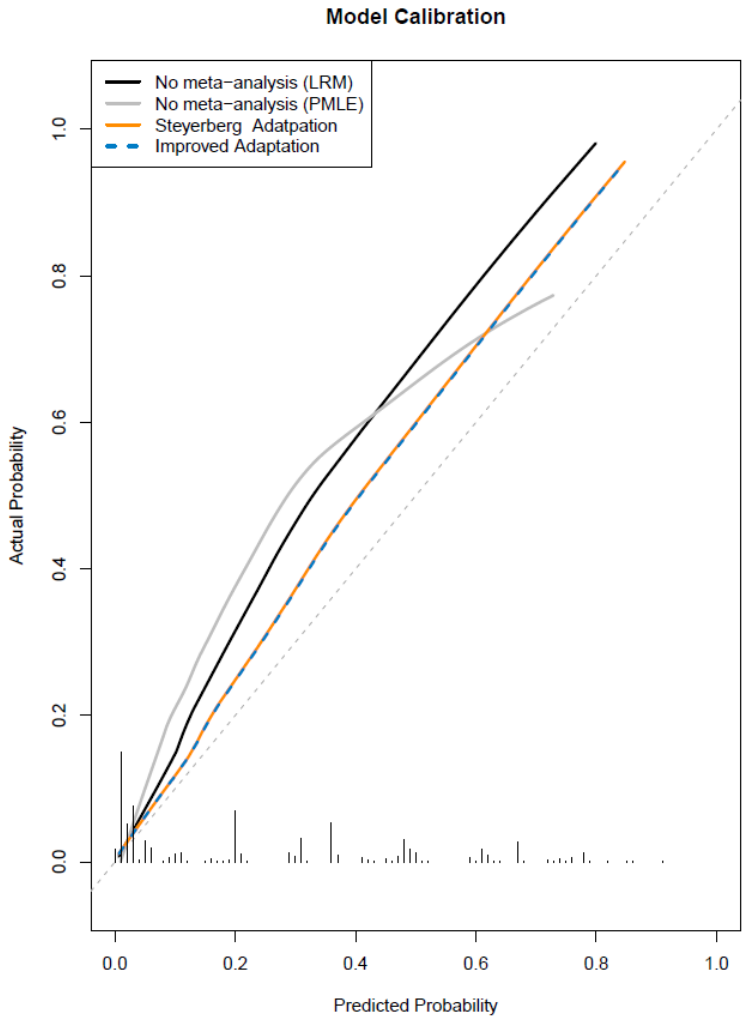
- Multivariable coefficient for D-dimer and 95% CI
- Model discrimination
- Model calibration

Case study

Diagnosis of Deep Vein Thrombosis







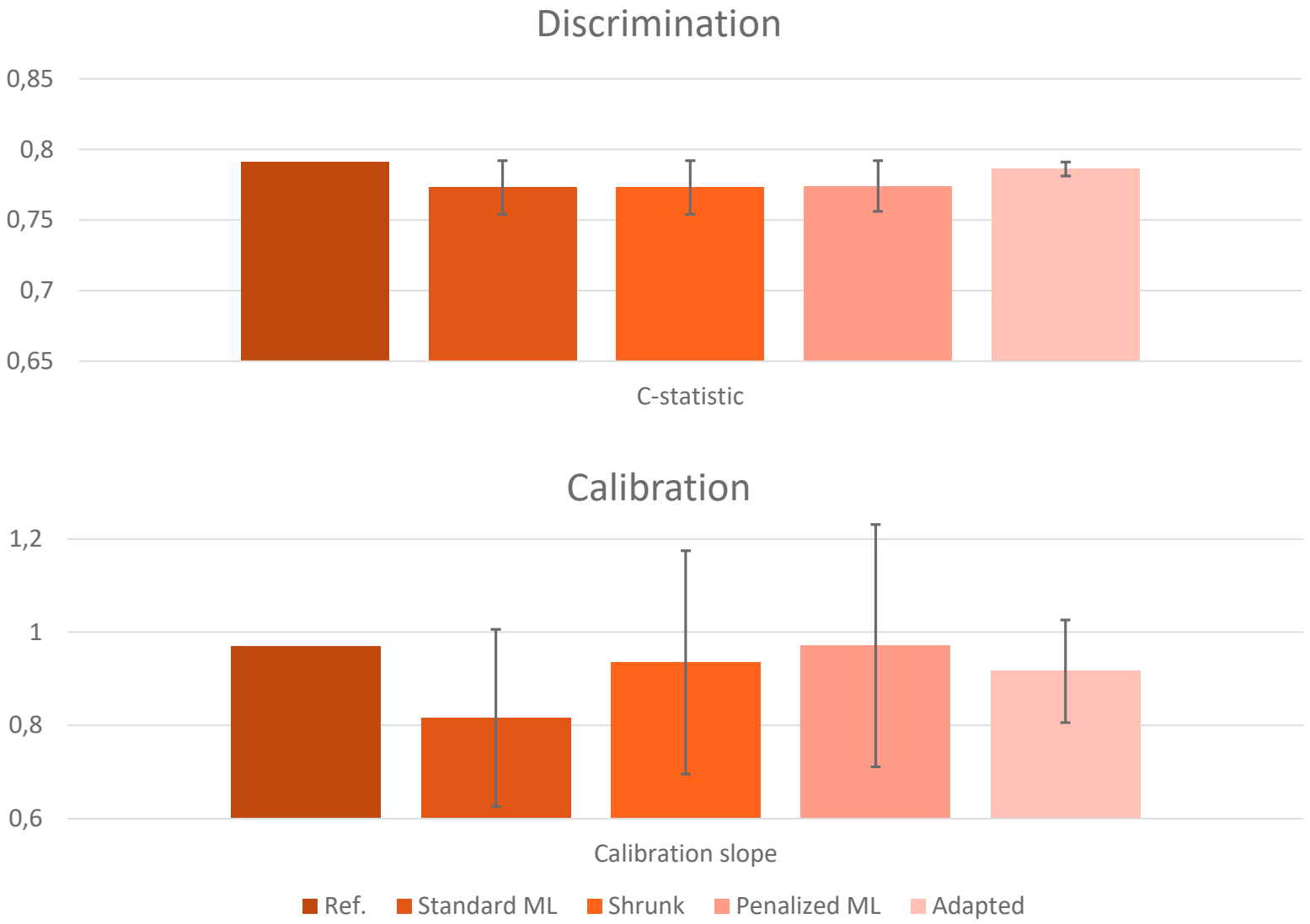
Case study

Prognosis of acute MI

- Outcome: 30-day mortality
- Population: patients with acute MI
- Predictors: Age, killip class, infarct location, ST elevation
- Literature coefficients from TIMI-II data set
- Model development datasets from GUSTO-I (*N=336 x 61*)
- Model validation dataset GUSTO-I (*N=40830*)

Case study

Prognosis of acute MI



Type 2

Incorporating adjusted predictor effects (models or rules with similar predictors)

UMC Utrecht

Background

Abundance of prediction models for the same clinical problem

- >150 models alike Framingham, SCOPE, Qrisk
- >100 models for brain trauma patients
- > 100 diabetes type 2 models
- > 60 models for breast cancer prognosis
- > 25 models for predicting long-term outcome in neurotrauma patients
- > 10 models to diagnose venous thromboembolism

Background

Prognosis of cardiovascular disease

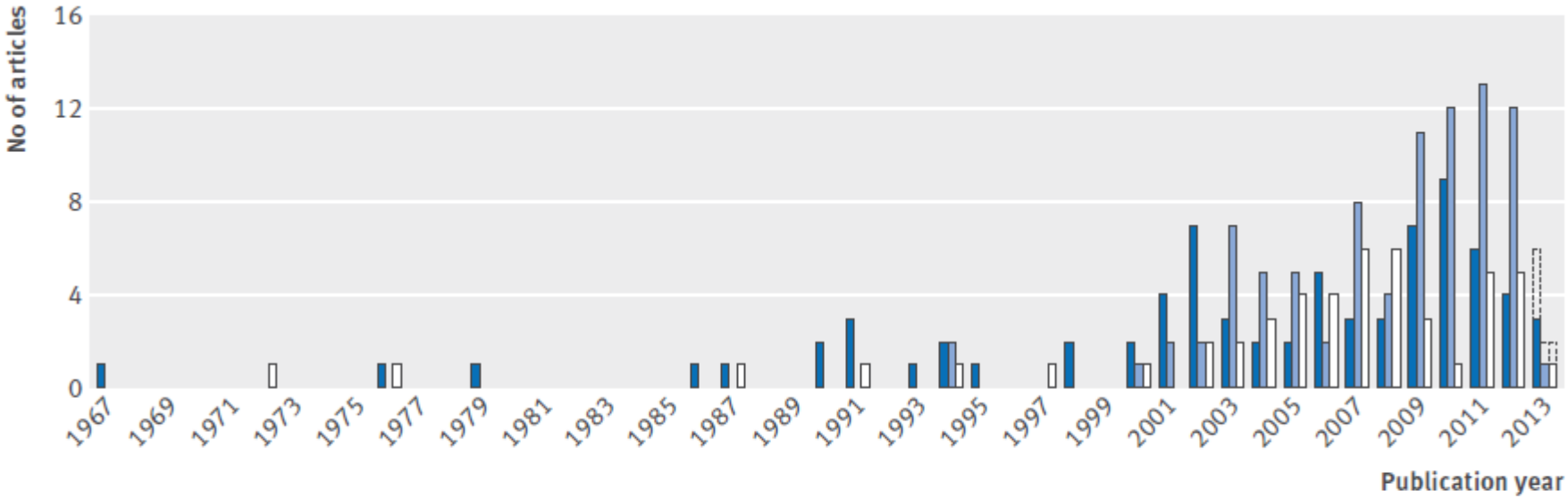


Fig 2 | Numbers of articles in which only one or more models were developed (dark blue), only one or more models were externally validated (light blue), or one or more models were developed and externally validated (white), ordered by publication year (up to June 2013). Predictions of the total numbers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

Ref: Damen JAAG et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ. 2016.

Background

Diagnosis of Deep Vein Thrombosis

Clinical feature	Score
Active cancer (treatment ongoing or within previous 6 months or palliative)	1
Paralysis, paresis, or recent plaster immobilisation of the lower extremities	1
Recently bedridden for more than 3 days or major surgery, within 4 weeks	1
Localised tenderness along the distribution of the deep venous system	1
Entire leg swollen	1
Calf swelling by more than 3 cm when compared with the asymptomatic leg (measured 10 cm below tibial tuberosity)	1
Pitting oedema (greater in the symptomatic leg)	1
Collateral superficial veins (non-varicose)	1
Alternative diagnosis as likely or greater than that of deep-vein thrombosis	−2

In patients with symptoms in both legs, the more symptomatic leg is used.

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)).

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

Background

Diagnosis of 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

Background

- How to take advantage of published predictor effects?
- How to take advantage of published weights?
- How to deal with between-study heterogeneity?

Methods

Research Article

Statistics in Medicine

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

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

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

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

Strategy 1

Meta-analysis of multivariable predictor effects

Summarize the multivariable regression coefficients and standard errors from literature + IPD at hand

- Univariate meta-analysis
- Multivariate meta-analysis

Similar to meta-analysis of IPD+AD of therapeutic trials

Result: a "pooled" prediction model applicable to the "average" population of development studies

Strategy 1

Meta-analysis of multivariable predictor effects

Summarize the multivariable regression coefficients and standard errors from literature + IPD at hand

- Univariate meta-analysis
- Multivariate meta-analysis

Similar to meta-analysis of IPD+AD of therapeutic trials

Result: a "pooled" prediction model applicable to the "average" population of development studies

Why may this not be desirable?

Strategy 2

Bayesian inference

Consider IPD at hand as the clinically relevant population

- Use evidence of existing models to inform estimation in the IPD at hand
- Bayesian estimation framework with informative prior distributions

Result: a “pooled” prediction model that is tailored to the current population

Restoring of missing information

Transforming weights to regression coefficients

- Re-estimate intercept + common slope in IPD at hand

Unknown regression coefficients

- Omit (e.g. through univariate meta-analysis)
- Impute (e.g. through multivariate meta-analysis)
- Re-estimate from IPD at hand

Unknown within-study correlations

- Assign a predefined value
- Borrow from IPD at hand

Case study

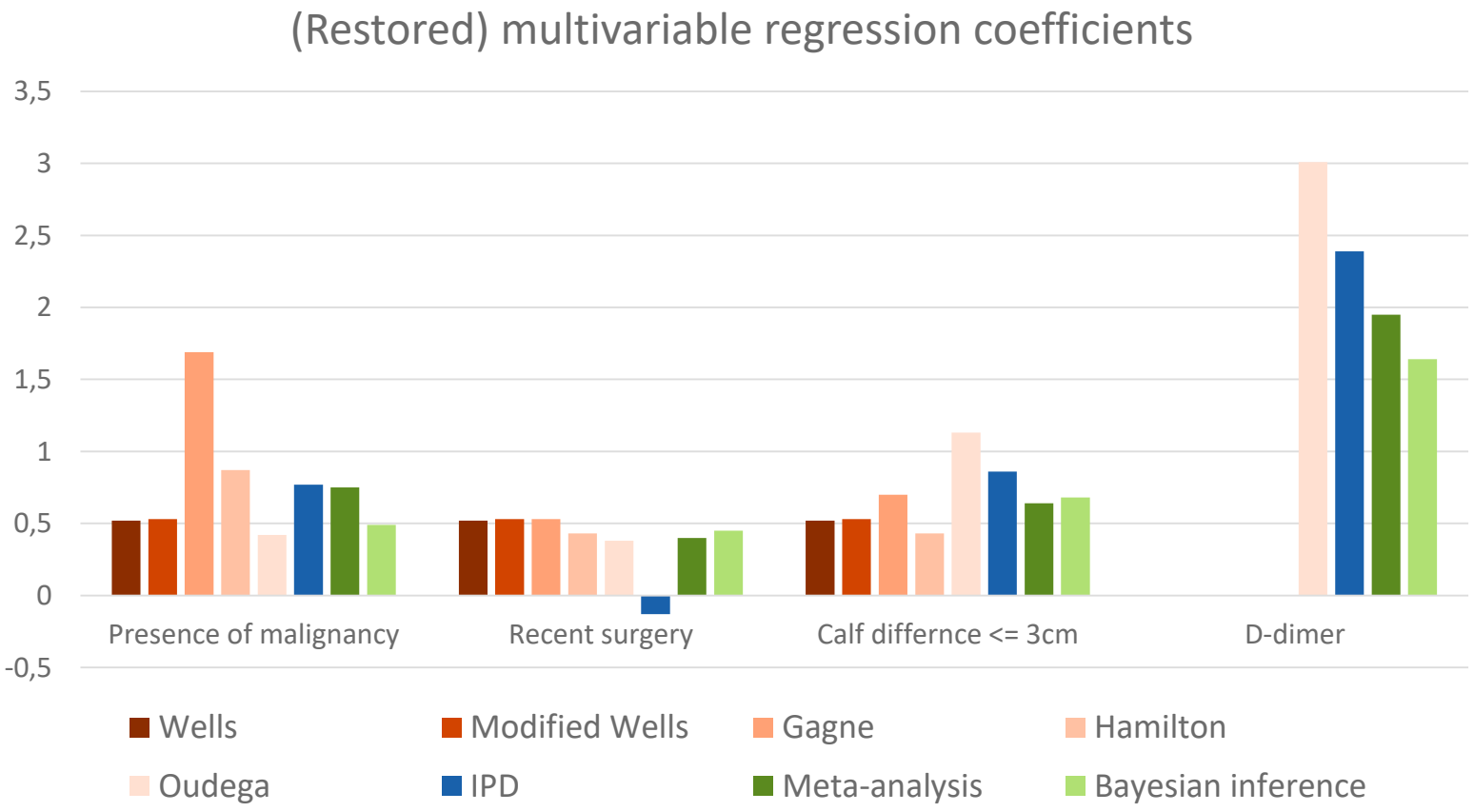
Diagnosis of Deep Vein Thrombosis

Illustration

- 5 previously published prediction models
  - Wells and Modified Wells
  - Gagne
  - Oudega
  - Hamilton
- Focus on 4 core predictors
- Model development dataset (*N = 1028*)
- External validation dataset (*N=791*)

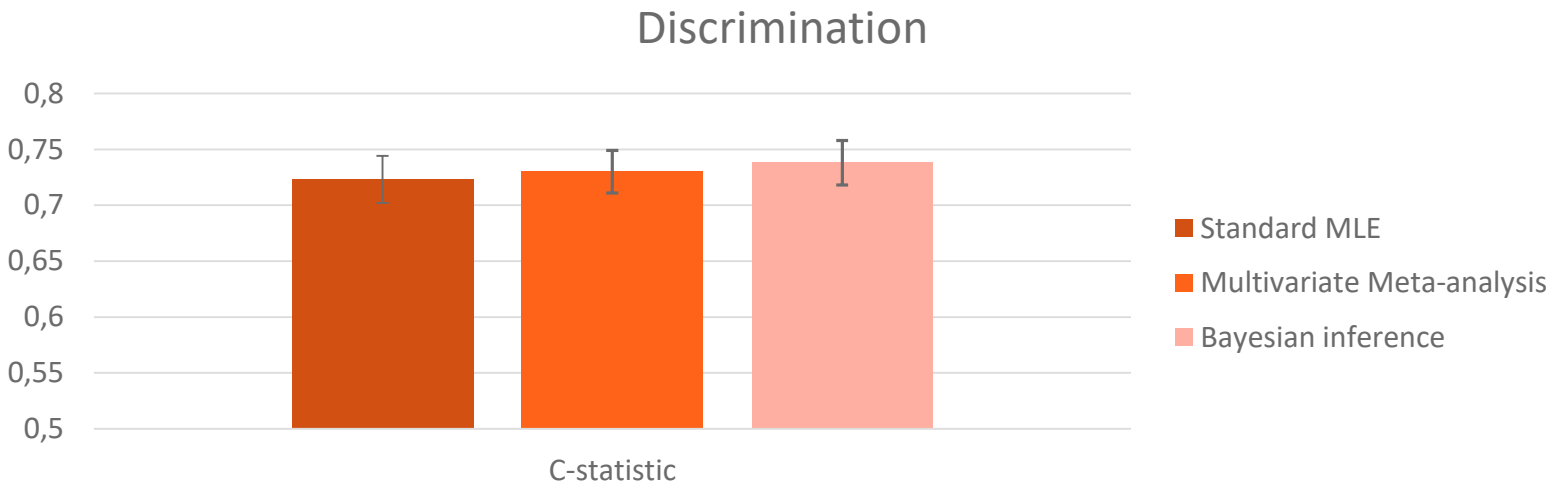
Case study

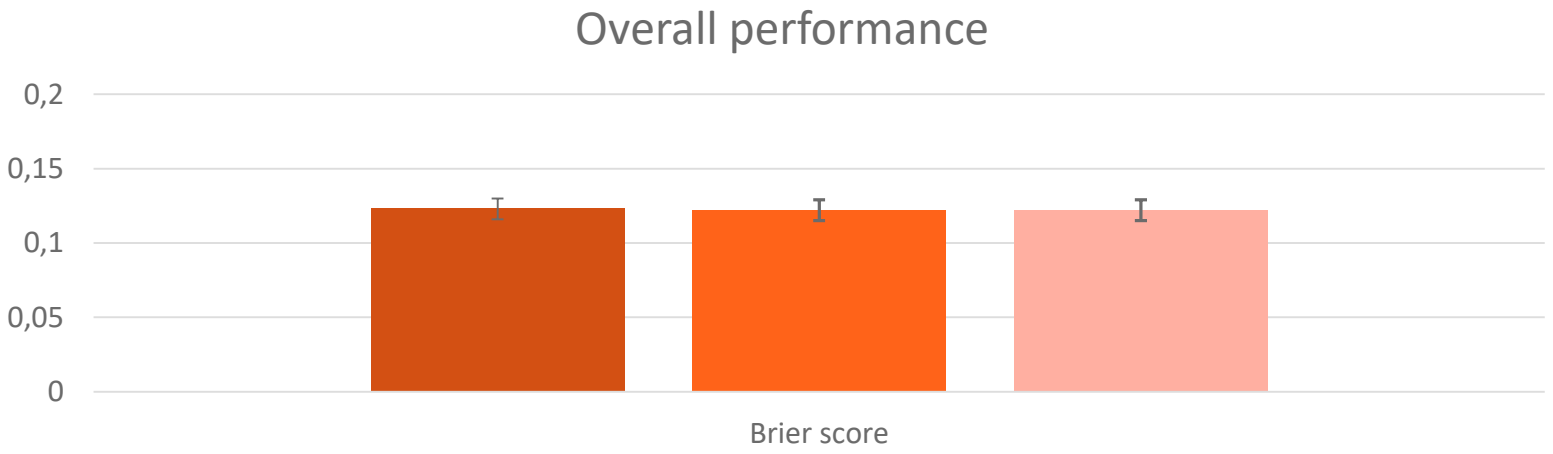
Diagnosis of Deep Vein Thrombosis



Case study

Diagnosis of Deep Vein Thrombosis





Summary points

Advantages

- Reduce danger of over-fitting in small data sets
- Feasible even when no IPD is at hand
- Acknowledgement of heterogeneity

Drawbacks

- Requirements usually not (fully) met
- Poor reporting of coefficients and standard errors
- Limited performance gain
- Limited adjustment for heterogeneity

Type 3

Incorporating adjusted predictor effects (models or rules with different predictors)

UMC Utrecht

Background

- Prediction models often include different predictors
- Inconsistent reporting of prediction models
- Incomplete reporting of regression analyses
- Heterogeneity between study populations

Background

Prognosis for recurrent venous thromboembolism

Table 5 Predictors included in final model

Model	HERDOO2	Vienna	DASH
Predictors included			
D-dimer	X	X	X
Age	X	–	X
Sex	–	X	X
BMI	X	–	–
Post-thrombotic signs	X	–	–
Site of index event	–	X	–
Hormone therapy	–	–	X

BMI, body mass index.

Ref: Ensor J et al. Systematic review of prognostic models for recurrent venous thromboembolism (VTE) post-treatment of first unprovoked VTE. BMJ Open.

Strengths and limitations of this study

- To our knowledge, this is the first systematic review identifying prognostic models for venous thromboembolism recurrence risk in the unprovoked population, using a robust systematic methodology.
- The study is also the first to assess the validity of the existing models in terms of risk of bias and applicability.
- We were unable to perform a quantitative analysis of the identified articles due to a lack of homogeneity in many areas, including the predictors used, model types and study populations.
- All models require further independent external validation, and as such the true performance of the models in the wider unprovoked population must be assessed in new research.

Methods

Research Article

Statistics in Medicine

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

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

Meta-analysis and aggregation of multiple published prediction models

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

Methods

Proposed solutions

- Prediction models often include different predictors
  - Generate “ensemble” of prediction models
- Inconsistent reporting of prediction models
  - Use “scores” rather than predicted probabilities
- Incomplete reporting of regression analyses
  - Avoid use of within-study (co)variance
- Heterogeneity between study populations
  - Integrate model validation, updating and meta-analysis

Strategy 1

Model averaging

Procedure

1. Validate and update literature models in “own” IPD
2. Use updated models to calculate prediction for each subject
3. Calculate model averaged prediction
   - Assign more weight to models with better fit in the IPD
   - Assign less weight to models that have been substantially revised

$$w_m = \frac{\exp(-0.5\,\mathrm{BIC}_m)}{\sum_{l=1}^{M} \exp(-0.5\,\mathrm{BIC}_l)}$$

4. Use the models’ averaged predictions as dependent variable to develop the meta-model

Strategy 1

Stacked regressions

Procedure

1. Treat the predictions or scores of each literature model as a predictor variable
2. Develop the meta-model by forming a linear combination of the model predictions
   - Estimation of a common intercept term
   - Estimation of a weight for each model
   - Allow omission of models with a “negative” contribution
3. Calculate regression coefficients of the meta-model by applying the estimated weights

Simultaneous updating, discovery and estimation of the best combination of literature models

Case study
Diagnosis of Deep Vein Thrombosis

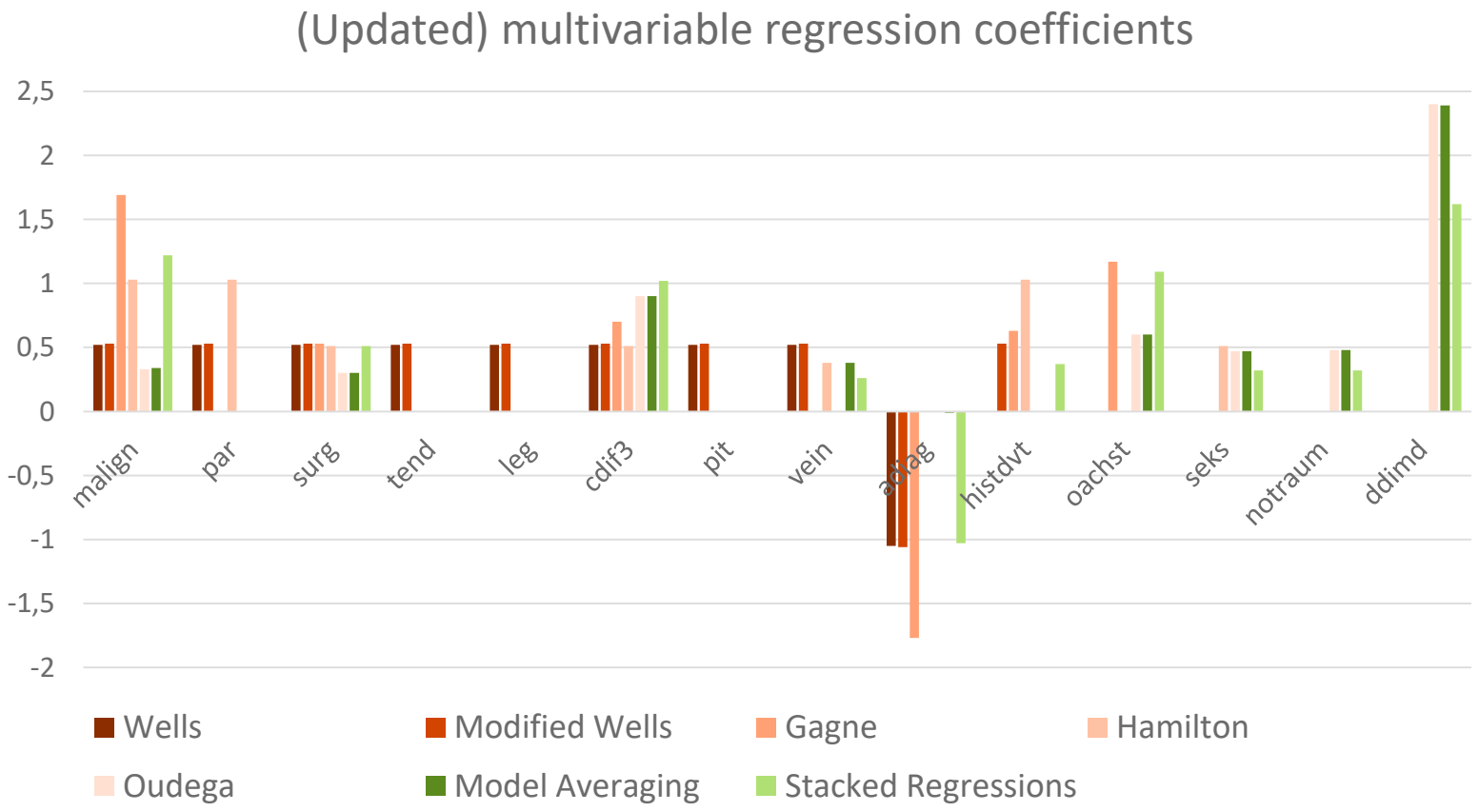
- 5 previously published prediction models
  - Primary care (Gagne, Oudega)
  - Secondary care (Wells, modified Wells, Hamilton)
- Model development dataset
  - Primary care (N=1028)
- 2 external validation datasets
  - Primary care (N=791)
  - Secondary care (N=1756)

Case study
Diagnosis of Deep Vein Thrombosis

Weights	Model Averaging	Stacked Regressions
Oudega	0,998	0,537
Gagne	0,002	0,497
Wells	0	0
Modified Wells	0	0
Hamilton	0	0
(Intercept)		1,01

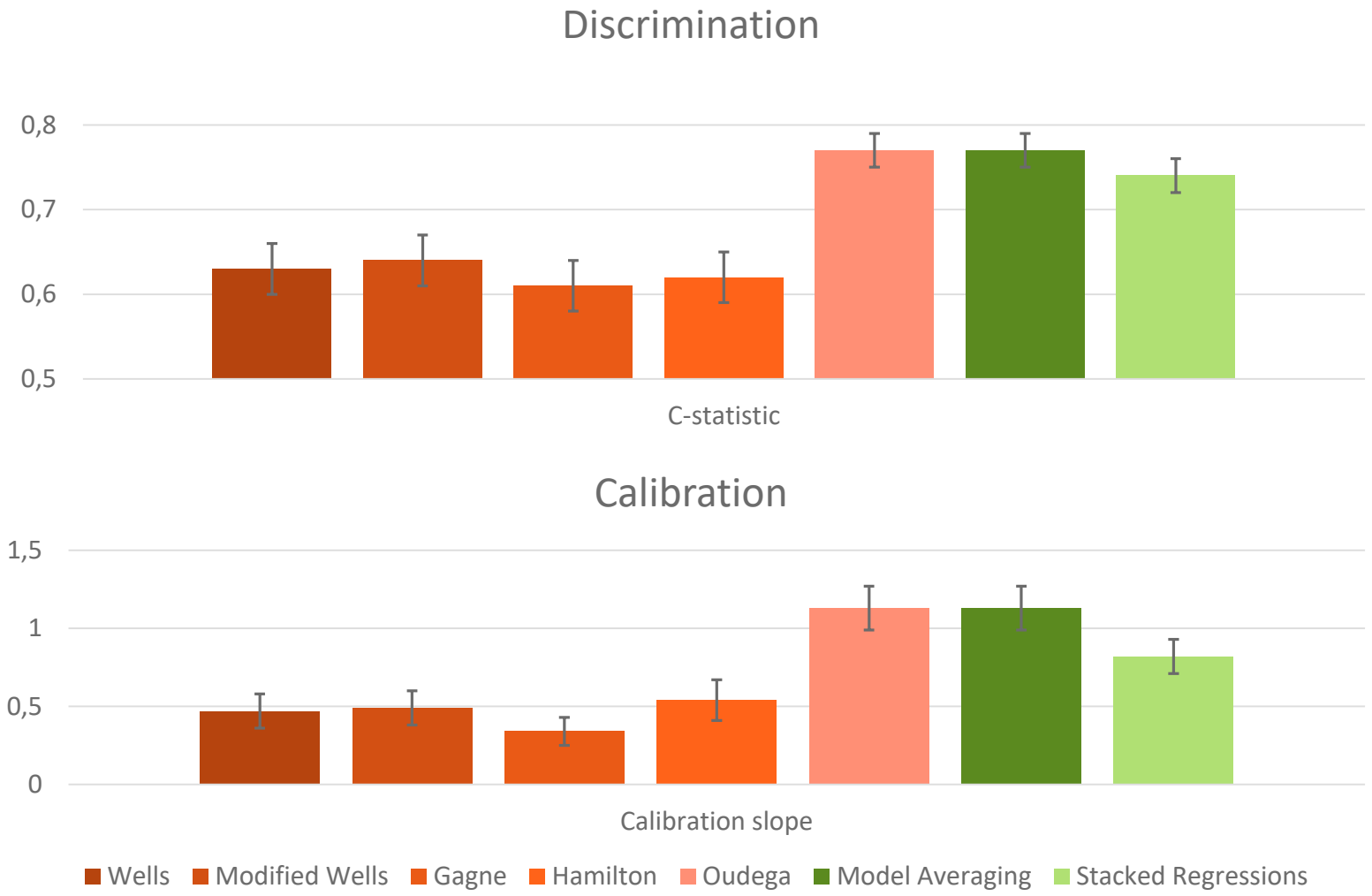
Case study

Diagnosis of Deep Vein Thrombosis



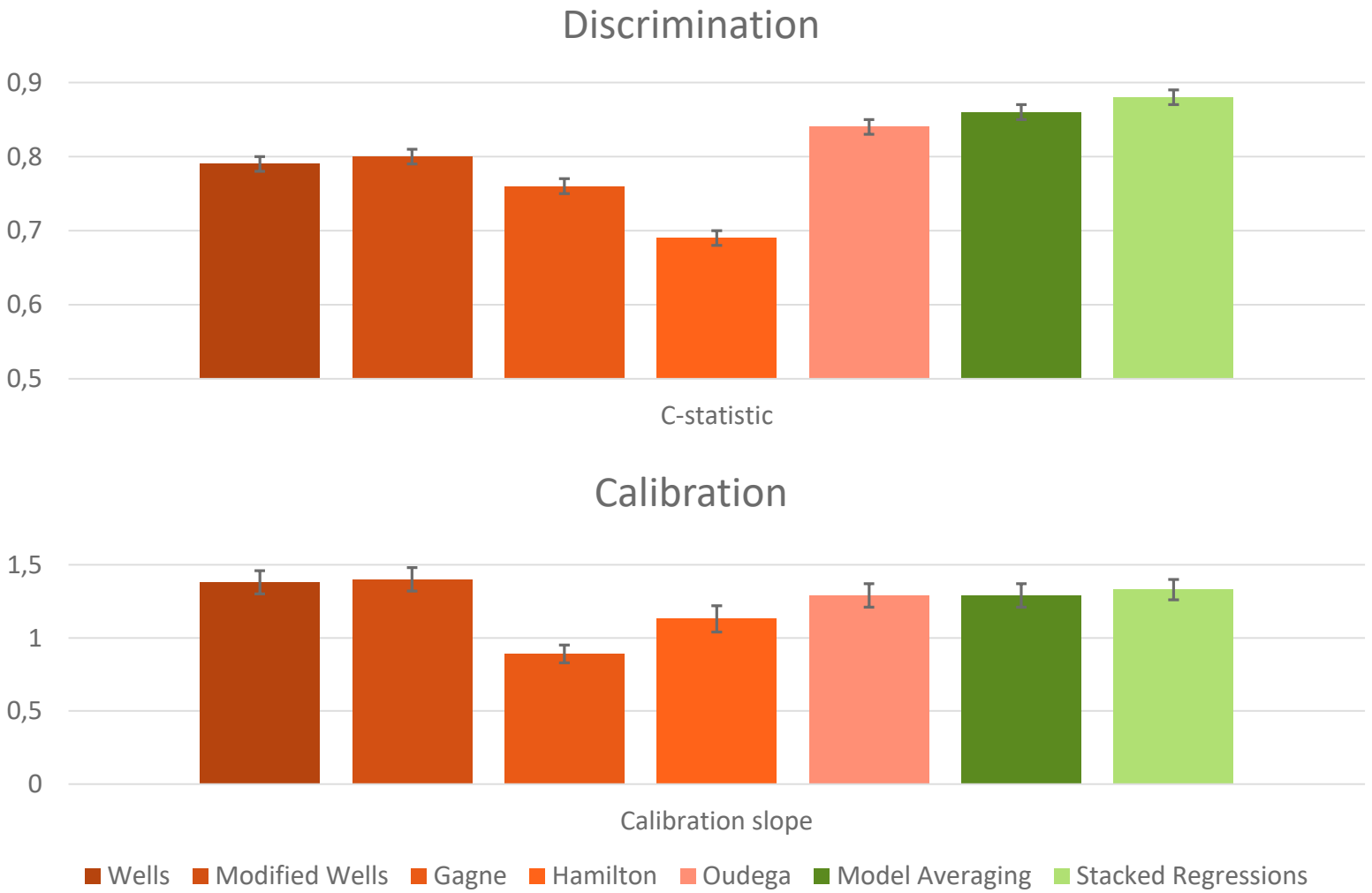
Case study

Diagnosis of Deep Vein Thrombosis (Primary care)



Case study

Diagnosis of Deep Vein Thrombosis (Secondary care)



Summary points

Advantages

- Natural extension of model updating
- Reduce danger of over-fitting in small data sets
- Acknowledgement & adjustment for heterogeneity

Summary points

UMC Utrecht

Strengths & weaknesses

Strengths

- Abundance of external information
- Straightforward implementation of methods
- Explicit aggregated models (no “black boxes”)
- Aggregation usually improves performance

Limitations

- Heterogeneity across studies
- Performance gain not always very large
- Additional efforts required during development

Courses

UMC Utrecht

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