UMC Utrecht
Julius Center

Individual Participant Data (IPD) Meta-analysis of prediction modelling studies

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

Cochrane

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 presentation

Prediction models: dynamic world

- Waves of new biomarkers and prediction models
- Increasing pressure for their evaluation
- Recognition of the importance of external validation
- Performance of models is likely to be variable
- Individual patient data: insight why models vary in performance or to build more robust models
- Improvements in methodology

Illustration

https://www.youtube.com/watch?v=OM_X_Czujrs&feature=player_detailpage

Workshop objectives

Provide guidance to conduct individual participant data (IPD) meta-analysis in prediction research

- To explain key concepts in prediction research
- To describe potential benefits of IPD
- To identify challenges for IPD reviews
- To provide examples of IPD meta-analyses
- To illustrate basic and novel methods

Prediction

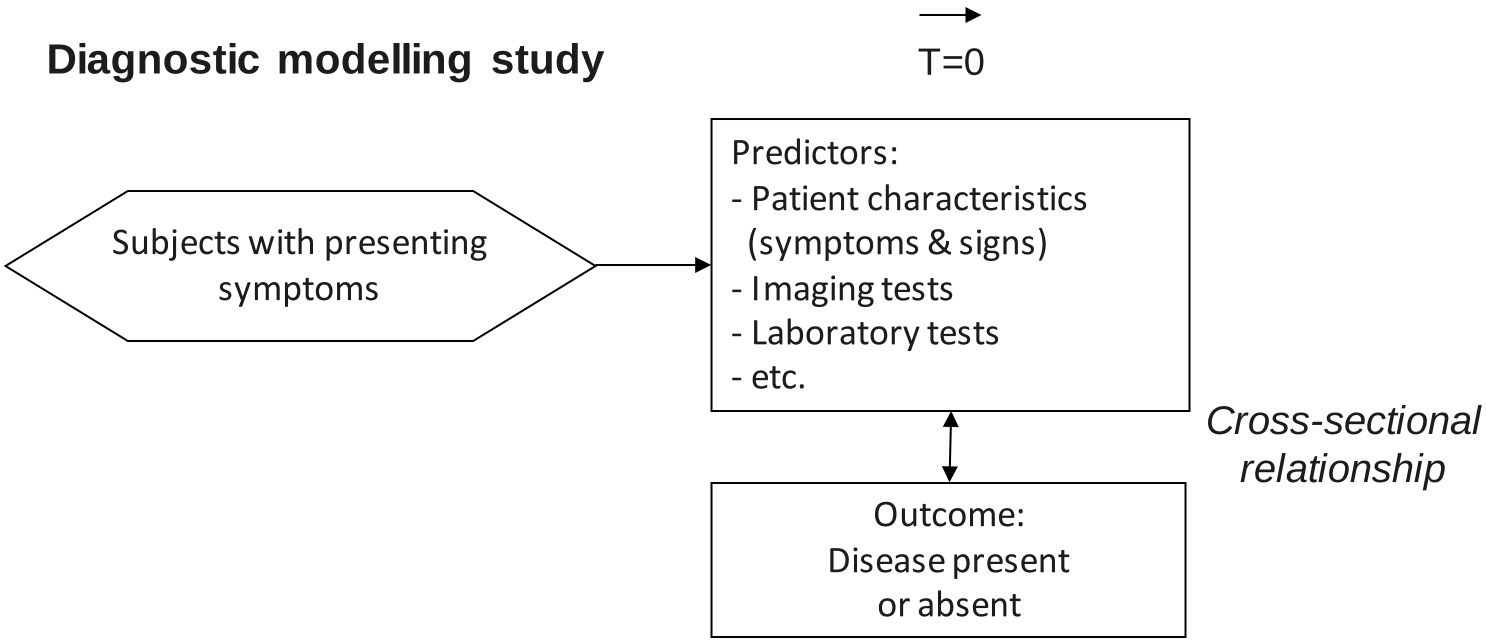
- Risk prediction = foreseeing / foretelling
  … (probability) of something that is yet unknown

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

Multivariable prediction models

- To calculate absolute risk based on individual profile
- Predict outcome from demographic, patient and disease characteristics (predictors, covariates, risk factors, X variables)
- Use of regression models, two main types:
  - Logistic regression
  - Time-to-event analysis (Kaplan-Meier, Cox)
- Statistical modelling: (1) overlap in information from different predictors; (2) acknowledge strength of each predictor

Diagnostic modelling study



T=0

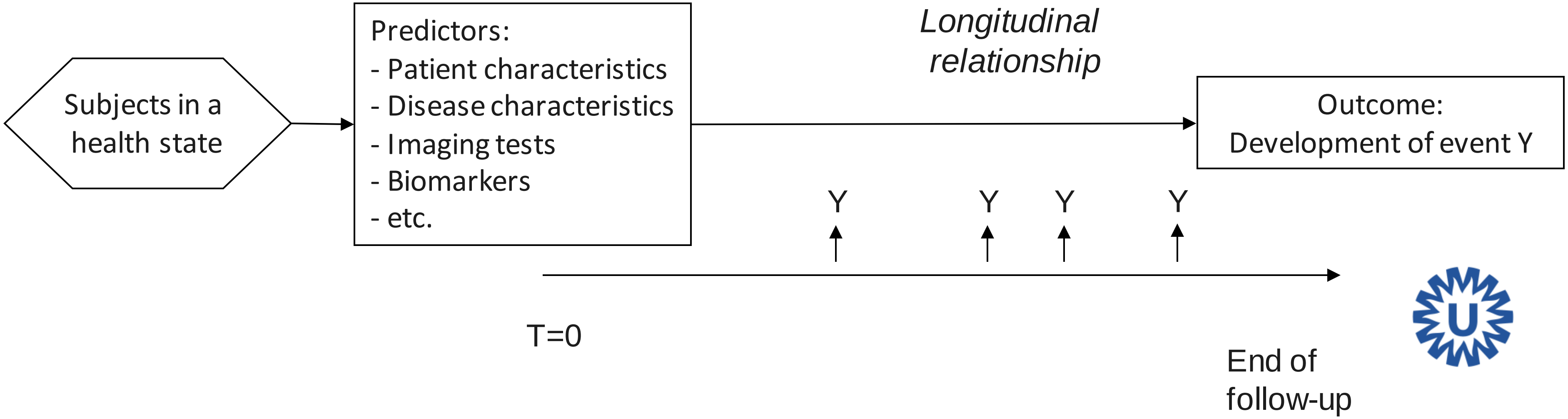
Subjects with presenting symptoms

Predictors:
- Patient characteristics (symptoms & signs)
- Imaging tests
- Laboratory tests
- etc.

Cross-sectional relationship

Outcome:
Disease present or absent

Prognostic modelling study



Subjects in a health state

Predictors:
- Patient characteristics
- Disease characteristics
- Imaging tests
- Biomarkers
- etc.

Longitudinal relationship

Outcome:
Development of event Y

T=0

End of follow-up

Prediction models

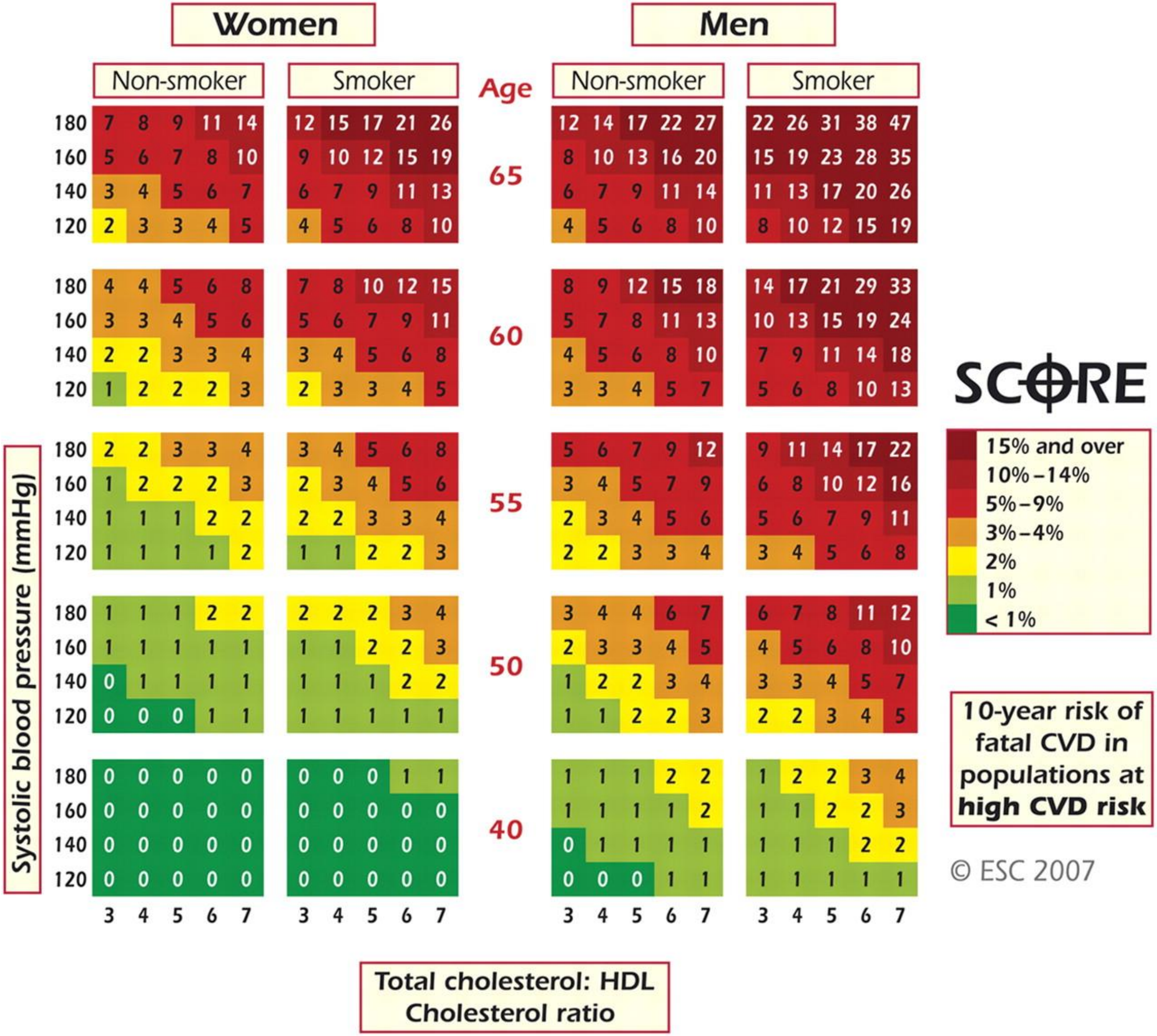
Predictors (in both diagnostic & prognostic models) are from:

- history taking
- physical examination
- tests (imaging, ECG, biomarkers, genetic ‘markers’)
- disease severity
- therapies received

Prediction models

Presented as:

- Mathematical formula requiring computer
- Simple scoring rules
- Score charts / Nomograms



SCORE

10-year risk of fatal CVD in populations at high CVD risk

Age	Systolic blood pressure (mmHg)	Women										Men									
		Non-smoker					Smoker					Non-smoker					Smoker				
		3	4	5	6	7	3	4	5	6	7	3	4	5	6	7	3	4	5	6	7
65	180	7	8	9	11	14	12	15	17	21	26	12	14	17	22	27	22	26	31	38	47
	160	5	6	7	8	10	9	10	12	15	19	8	10	13	16	20	15	19	23	28	35
	140	3	4	5	6	7	6	7	9	11	13	6	7	9	11	14	11	13	17	20	26
	120	2	3	3	4	5	4	5	6	8	10	4	5	6	8	10	8	10	12	15	19
60	180	4	4	5	6	8	7	8	10	12	15	8	9	12	15	18	14	17	21	29	33
	160	3	3	4	5	6	5	6	7	9	11	5	7	8	11	13	10	13	15	19	24
	140	2	2	3	3	4	3	4	5	6	8	4	5	6	8	10	7	9	11	14	18
	120	1	2	2	2	3	2	3	3	4	5	3	3	4	5	7	5	6	8	10	13
55	180	2	2	3	3	4	3	4	5	6	8	5	6	7	9	12	9	11	14	17	22
	160	1	2	2	2	3	2	3	4	5	6	3	4	5	7	9	6	8	10	12	16
	140	1	1	1	2	2	2	2	3	3	4	2	3	4	5	6	5	6	7	9	11
	120	1	1	1	1	2	1	1	2	2	3	2	2	3	3	4	3	4	5	6	8
50	180	1	1	1	2	2	2	2	2	3	4	3	4	4	6	7	6	7	8	11	12
	160	1	1	1	1	1	1	1	2	2	3	2	3	3	4	5	4	5	6	8	10
	140	0	1	1	1	1	1	1	1	2	2	1	2	2	3	4	3	3	4	5	7
	120	0	0	0	1	1	1	1	1	1	1	1	1	2	2	3	2	2	3	4	5
40	180	0	0	0	0	0	0	0	0	1	1	1	1	1	2	2	1	2	2	3	4
	160	0	0	0	0	0	0	0	0	0	0	1	1	1	1	2	1	1	2	2	3
	140	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	2	2
	120	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1

Total cholesterol: HDL Cholesterol ratio

- 15% and over
- 10% – 14%
- 5% – 9%
- 3% – 4%
- 2%
- 1%
- < 1%

© ESC 2007

Predicting bacterial cause in conjunctivitis

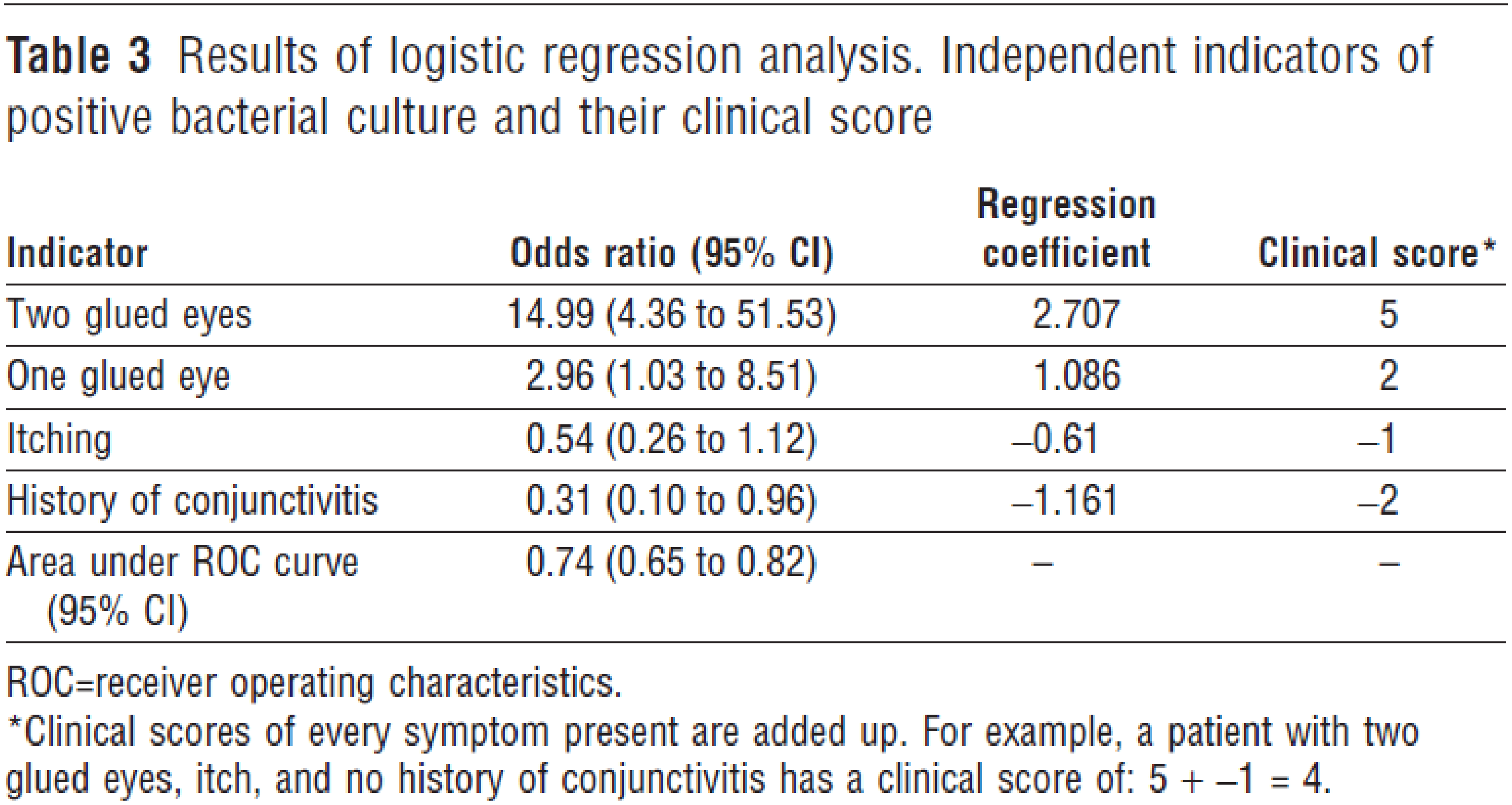


Table 3 Results of logistic regression analysis. Independent indicators of positive bacterial culture and their clinical score

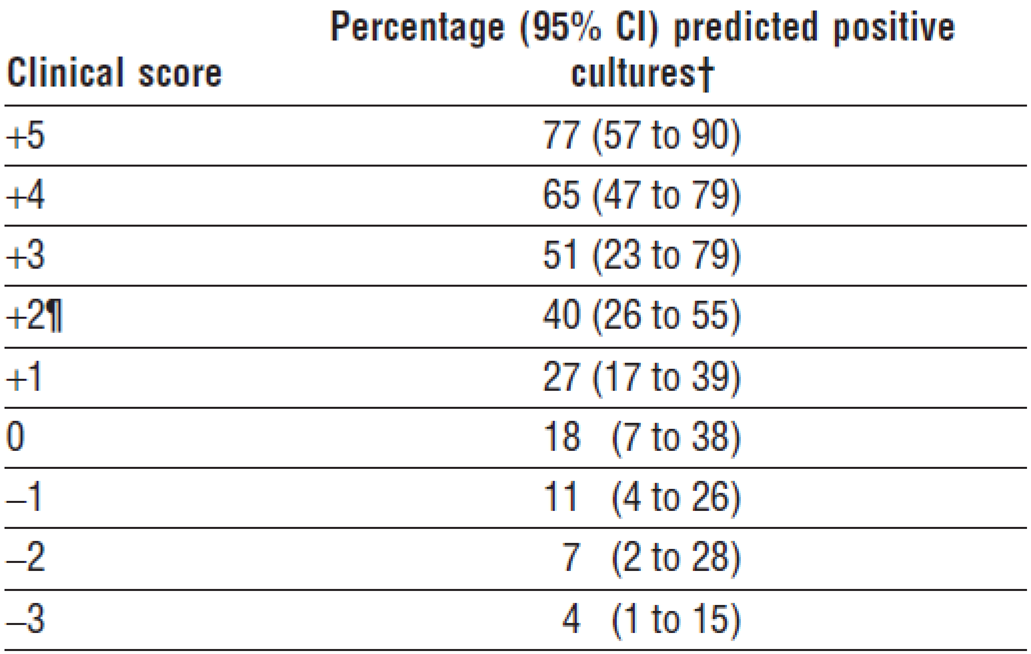
Indicator	Odds ratio (95% CI)	Regression coefficient	Clinical score*
Two glued eyes	14.99 (4.36 to 51.53)	2.707	5
One glued eye	2.96 (1.03 to 8.51)	1.086	2
Itching	0.54 (0.26 to 1.12)	–0.61	–1
History of conjunctivitis	0.31 (0.10 to 0.96)	–1.161	–2
Area under ROC curve (95% CI)	0.74 (0.65 to 0.82)	–	–

ROC=receiver operating characteristics.

*Clinical scores of every symptom present are added up. For example, a patient with two glued eyes, itch, and no history of conjunctivitis has a clinical score of: 5 + –1 = 4.

Rietveld et al. BMJ 2004;329:206

Predicting bacterial cause in conjunctivitis



Clinical score	Percentage (95% CI) predicted positive cultures†
+5	77 (57 to 90)
+4	65 (47 to 79)
+3	51 (23 to 79)
+2¶	40 (26 to 55)
+1	27 (17 to 39)
0	18 (7 to 38)
–1	11 (4 to 26)
–2	7 (2 to 28)
–3	4 (1 to 15)

Pitfalls of prediction research

- The **quality** of much prognosis research is poor (incomplete reporting, poor data sharing, incomplete registrations, absent study protocols)
- Development dataset often **too small or too local**
- Most prediction models are never validated in independent data (**external validation**)
- **Heterogeneity** across studies and settings, requiring local adjustments
- Many prediction models **generalize poorly** across different but related study populations, and tend to perform more poorly than anticipated when applied in routine care

Meta-analysis of individual participant data

Opportunities

- Increase total sample size
  - Reduce risk of overfitting
  - Ability to investigate more complex associations
- Increase available case-mix variability -> enhances the model's potential generalisability
- Ability to standardize analysis methods across IPD sets
- Ability to evaluate generalisability and usability of prediction models across different situations

Meta-analysis of individual participant data

IPD – are we realistic?

- Researchers **protective** over their own data
- Worried about Data Protection Act (**ethics**) – however, no need to include patient ID numbers
- **Cost, time** – when does it become worthwhile?

To conduct better prognostic & diagnostic research we need:

- To be prepared to **collaborate** and share data to make IPD available – in paper, on Web, on request
- To be involved in **prospectively planned** pooled analyses

Meta-analysis of individual participant data

Why do we need specific guidance?

Evidence synthesis currently gold standard for summarizing relative treatment effects – many methods available!

However,

- Meta-analysis models cannot *mutate mutandis* be applied to prediction modeling studies
- Researchers often simply combine all IPD, and produce a prediction model averaged across all study populations
- There are major differences in the aims, design and analysis of primary studies between prediction modeling and intervention studies!

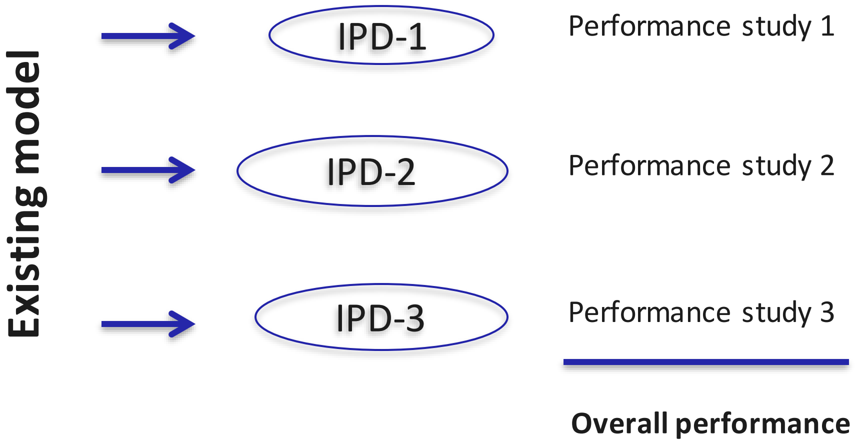
What are the main differences between prediction and intervention research?

Intervention research	Prediction research
Aim(s) - Estimation of therapeutic effect of a specific treatment - Study treatment effect in subgroups	**Aim(s)** - Estimation of absolute risk probabilities for distinct individuals across different populations or subgroups - Evaluate accuracy of model predictions across subgroups
Association measures: relative risk estimates	**Association measures**: absolute probability of risk estimates
Study design: Randomized studies	**Study design**: observational research
Evaluation: bias and precision of estimated comparative treatment effects	**Evaluation**: model discrimination and calibration

Types of IPD-MA of prediction modeling studies

1. Validation and comparison of existing model(s)
2. Improving upon existing model(s)
   - Updating
   - Added value of novel marker
3. Development of new model(s)

Validation and comparison of existing model(s)



Existing model

IPD-1

IPD-2

IPD-3

Performance study 1

Performance study 2

Performance study 3

Overall performance

Apply meta-analysis to:

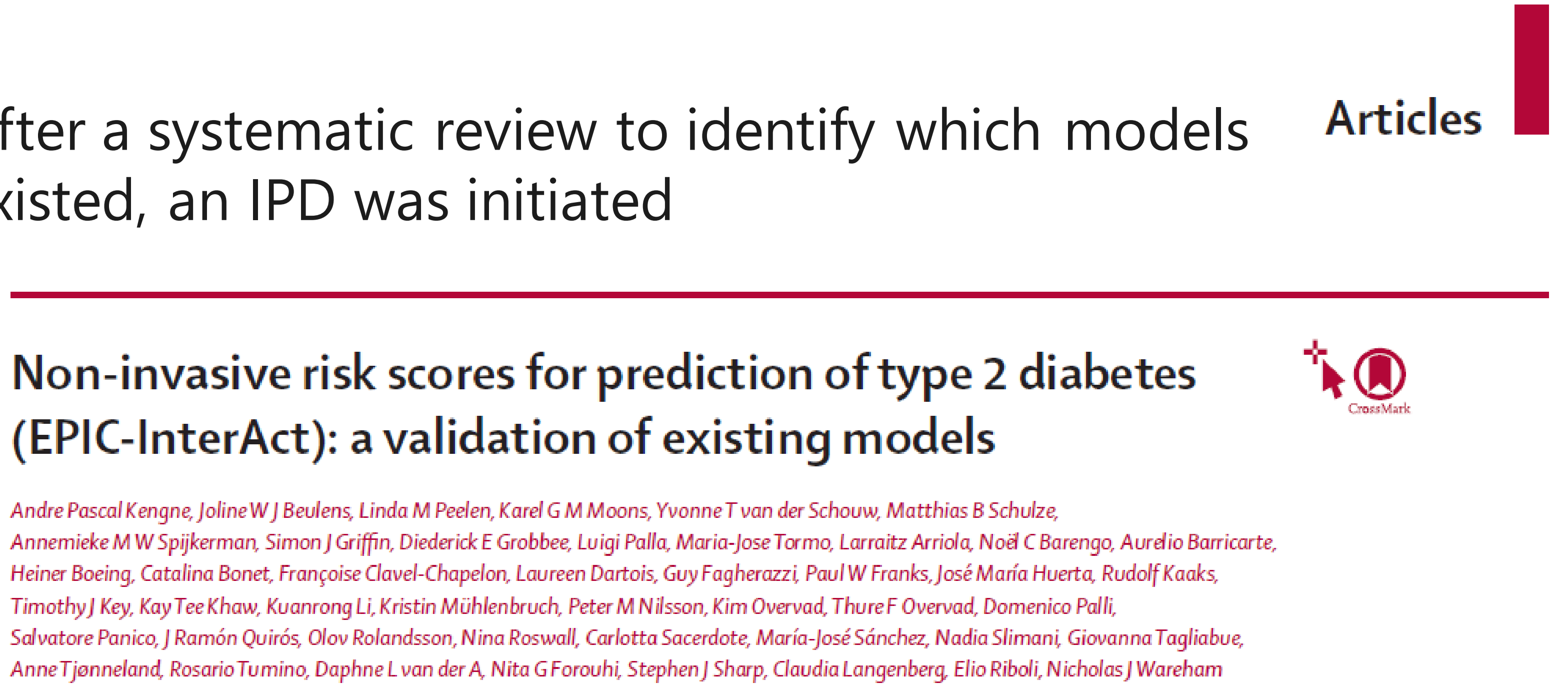
- Summarize estimates of model discrimination and calibration

Use IPD to:

- Investigate sources of heterogeneity in model performance
- Identify which models perform best in what (sub)population, setting or country

Validation and comparison of existing model(s)

After a systematic review to identify which models existed, an IPD was initiated



Articles

Non-invasive risk scores for prediction of type 2 diabetes (EPIC-InterAct): a validation of existing models

Andre Pascal Kengne, Joline W J Beulens, Linda M Peelen, Karel G M Moons, Yvonne T van der Schouw, Matthias B Schulze, Annemieke M W Spijkerman, Simon J Griffin, Diederick E Grobbee, Luigi Palla, Maria-Jose Tormo, Larraitz Arriola, Noël C Barengo, Aurelio Barricarte, Heiner Boeing, Catalina Bonet, Françoise Clavel-Chapelon, Laureen Dartois, Guy Fagherazzi, Paul W Franks, José María Huerta, Rudolf Kaaks, Timothy J Key, Kay Tee Khaw, Kuanrong Li, Kristin Mühlenbruch, Peter M Nilsson, Kim Overvad, Thure F Overvad, Domenico Palli, Salvatore Panico, J Ramón Quirós, Olov Rolandsson, Nina Roswall, Carlotta Sacerdote, María-José Sánchez, Nadia Slimani, Giovanna Tagliabue, Anne Tjønneland, Rosario Tumino, Daphne L van der A, Nita G Forouhi, Stephen J Sharp, Claudia Langenberg, Elio Riboli, Nicholas J Wareham

The Lancet, Diabetes & Endocrinology (2014)

Validation and comparison of existing model(s)

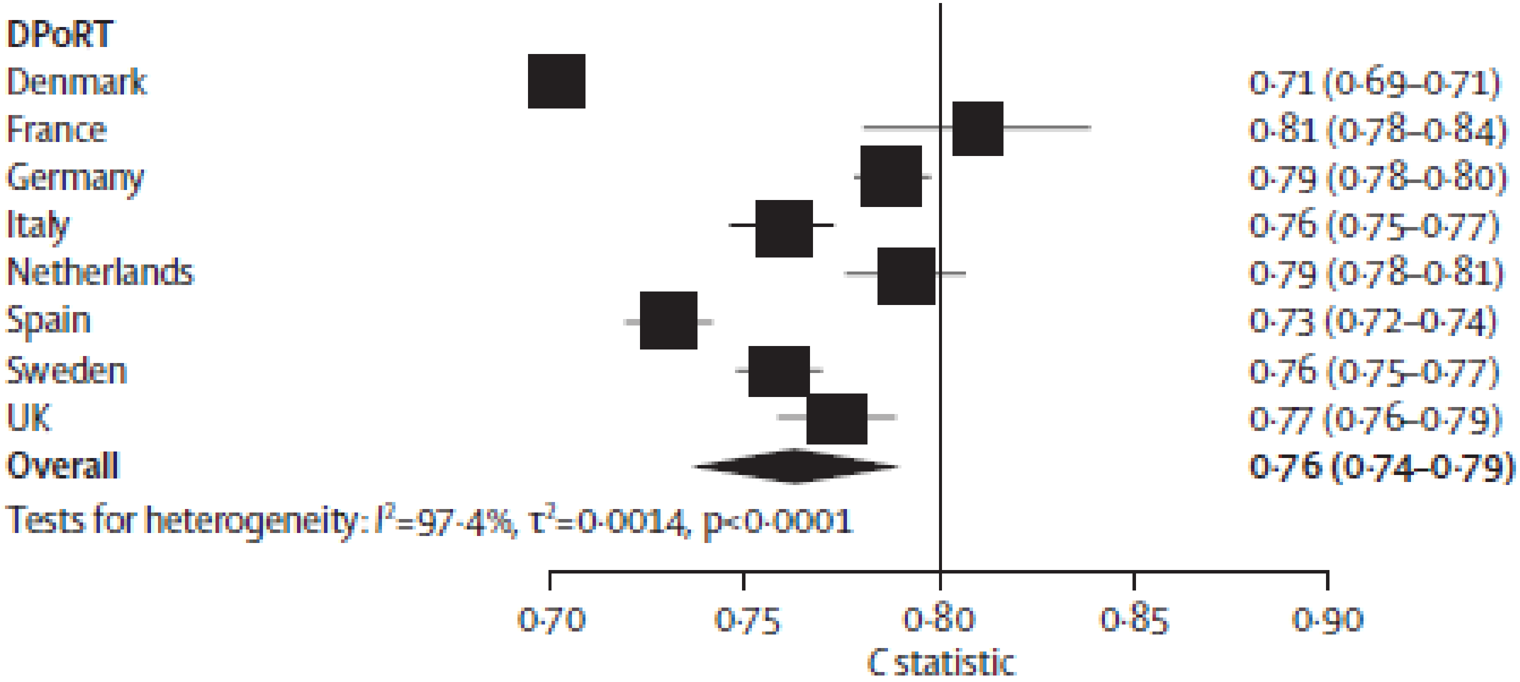
IPD meta-analysis

- EPIC-InterAct international study
  - 27,779 participants of whom 12,403 with incident diabetes
  - 8 countries
- External validation of 12 literature models (with non-laboratory based variables)
  - Discrimination: c-statistic
  - Calibration: calibration plot, ratio expected versus observed
  - Other performance measures: Yates slope, Brier score

Validation and comparison of existing model(s)

Discrimination of model "DPoRT"

(overall and by country)



Prediction of incident type 2 diabetes at 10 years of follow-up

Improving upon existing model(s)

Different types of improvements

- Adjusting baseline risk (e.g. intercept term)
- Adjusting common slope
- Updating individual predictor effects
- Adding new predictors or (bio)markers
- Removing exiting predictors

Aim: Tailor the model(s) to specific (sub)populations, settings or countries

Improving upon existing model(s)

Example: Majed and colleagues evaluated whether the calibration of the Framingham risk equation for coronary heart disease and stroke improved by applying local adjustments.

	E:O ratio			C statistic		
	O	**R**	**L**	**O**	**R**	**L**
PRIME-total	1.94	0.98	1.00	0.68	0.68	0.68
PRIME-France	2.23	0.99	1.00	0.67	0.67	0.68
PRIME-Ireland	1.42	0.99	1.00	0.67	0.67	0.67

Ref: Majed *et al. Preventive Medicine* 2008 **57**.

Improving upon existing model(s)

Apply meta-analysis to:

- Summarize estimates of added value
  - Adjusted predictor effects
  - Improvement in model calibration
  - Improvement in model discrimination
  - Improvement in model reclassification

Use IPD to:

- Investigate sources of heterogeneity in added value
- Identify relevant subgroups that yield different added value

Improving upon existing model(s)

Example: The clinical usefulness of carotid intima-media thickness measurements (CIMT) in cardiovascular risk prediction

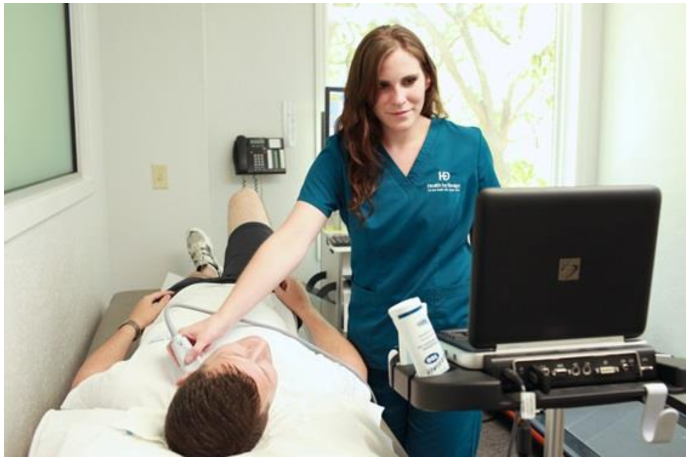
Background: problems with Framingham risk score in predicting CVD risk

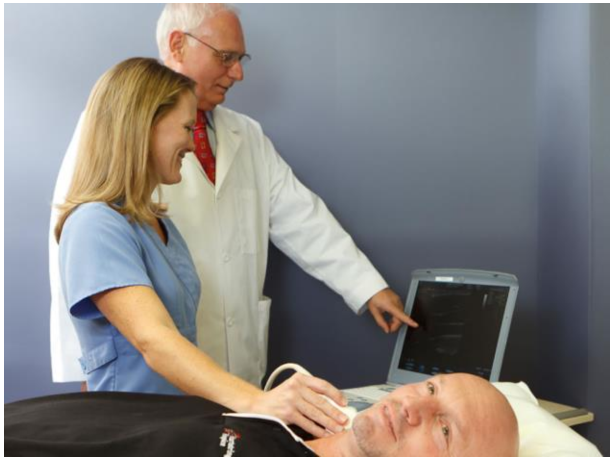
- No events despite high risk
- Many events in low risk categories

(Hester den Ruijter, Department of experimental cardiology, Julius Center for Health Sciences and Primary Care)

Improving upon existing model(s)

B-mode ultrasound measurement of the Carotid Intima Media Thickness (CIMT)





https://www.youtube.com/watch?v=OM_X_Czujrs&feature=player_detailpage

Improving upon existing model(s)

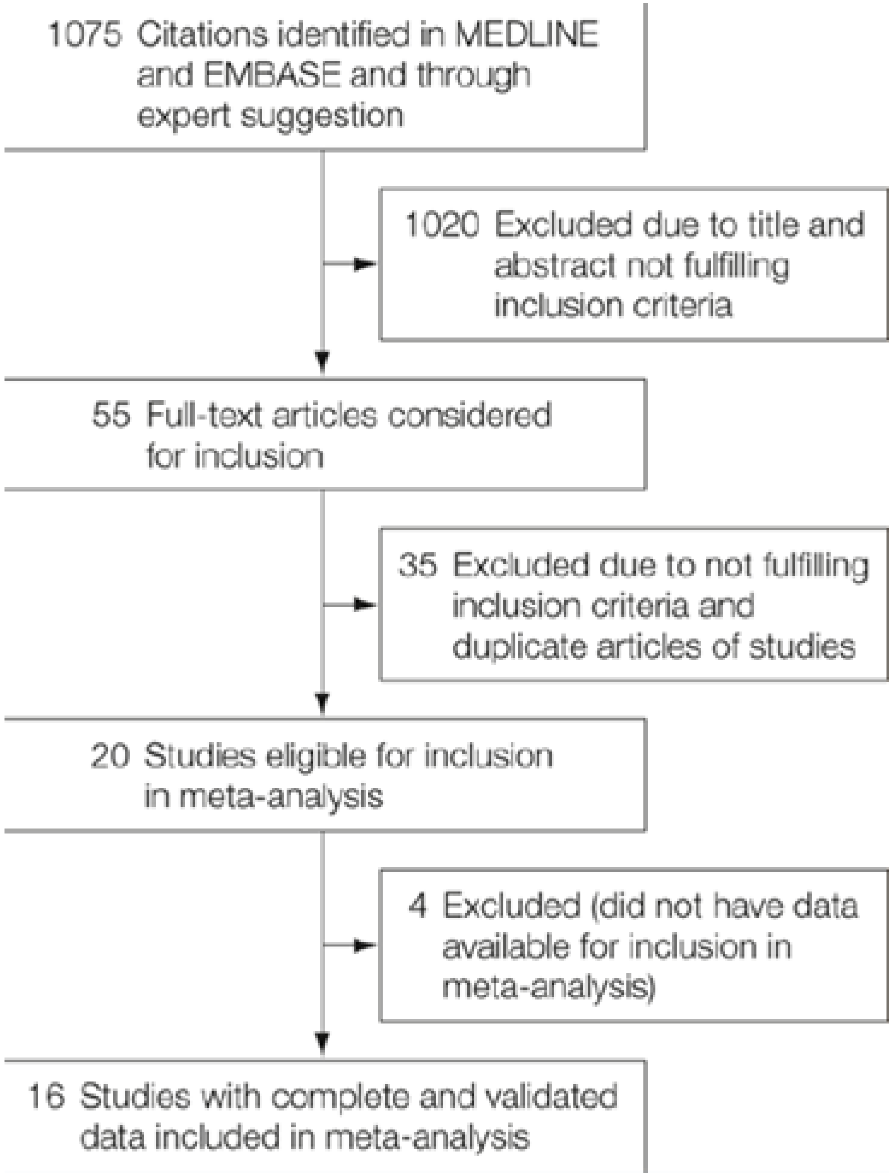
Improvement in CVD risk prediction: incorporation of non-invasive measurement of **atherosclerosis** by means of CIMT measurements

- Reflects long-term exposure to risk factor levels
- Predicts future cardiovascular events
- Modifiable by treatment
- Intermediate between risk factors and events

Improving upon existing model(s)

USE-IMT collaboration

- Ongoing individual participant data meta-analysis of general population
- Studies were invited to participate when they had data on Framingham risk score, CIMT measurements and follow-up to CVD



1075 Citations identified in MEDLINE and EMBASE and through expert suggestion

1020 Excluded due to title and abstract not fulfilling inclusion criteria

55 Full-text articles considered for inclusion

35 Excluded due to not fulfilling inclusion criteria and duplicate articles of studies

20 Studies eligible for inclusion in meta-analysis

4 Excluded (did not have data available for inclusion in meta-analysis)

16 Studies with complete and validated data included in meta-analysis

Improving upon existing model(s)

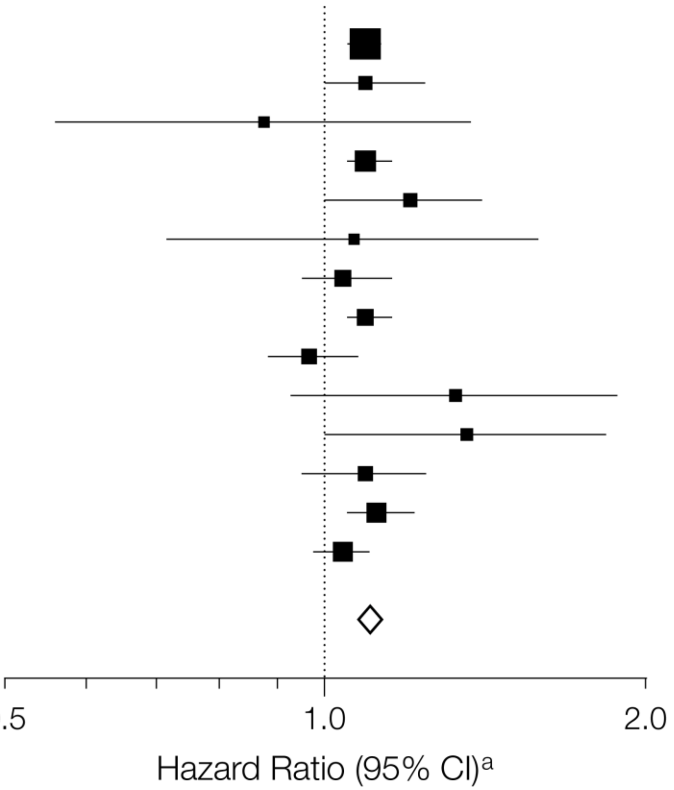
- Two Cox proportional hazards models with stroke and MI
  - FRS (refit age, gender, cholesterol, blood pressure, smoking, blood pressure medication)
  - FRS (refit age, gender, cholesterol, blood pressure, smoking, blood pressure medication) + ***CIMT***

- Do these two models reclassify patients differently?

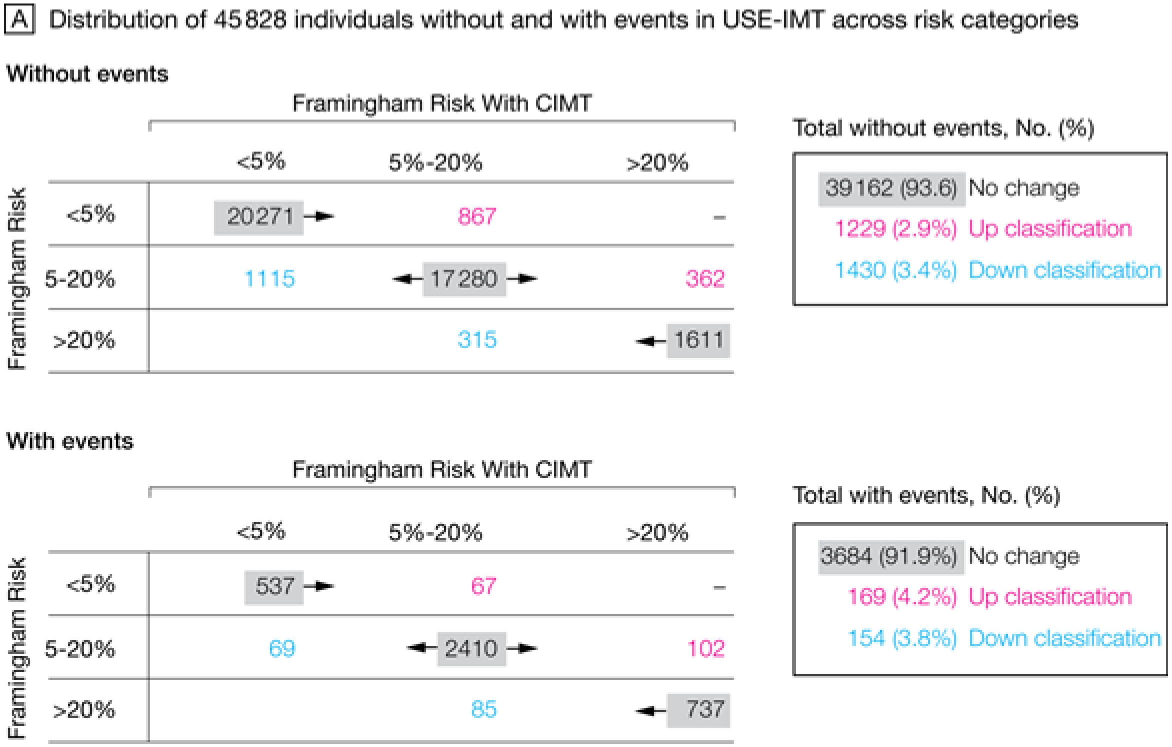
FRS = Framingham Risk Score

Improving upon existing model(s)

Source	Contribution to Total USE-IMT Population, % of Total	Hazard Ratio (95% CI)[a]
ARIC,[25] 1994	31	1.11 (1.08-1.14)
CAPS,[26] 2006	8	1.10 (0.99-1.23)
Charlottesville,[27] 2006	1	0.88 (0.56-1.36)
CHS,[28] 2007	7	1.11 (1.06-1.16)
FATE,[8] 2011	3	1.20 (1.01-1.42)
Hoorn Study,[29] 2003	1	1.07 (0.72-1.59)
KIHD,[30] 1991	2	1.05 (0.96-1.16)
Malmo,[31] 2000	10	1.10 (1.04-1.17)
MESA,[32] 2007	13	0.98 (0.89-1.08)
Nijmegen Study,[33] 2009	3	1.34 (0.94-1.90)
NOMAS,[34] 2007	2	1.36 (0.99-1.85)
OSACA2 Study,[35] 2007	1	1.09 (0.96-1.24)
Rotterdam Study,[36] 1997	8	1.13 (1.06-1.20)
Tromsø Study,[37] 2000	9	1.04 (0.98-1.10)
$I^2 = 12.30\%$; Q test for heterogeneity, $P = .24$		1.09 (1.07-1.12)



Improving upon existing model(s)



A Distribution of 45 828 individuals without and with events in USE-IMT across risk categories

Without events

Framingham Risk	Framingham Risk With CIMT		
	<5%	5%-20%	>20%
<5%	20271	867	–
5-20%	1115	17280	362
>20%		315	1611

Total without events, No. (%)

39162 (93.6)	No change
1229 (2.9%)	Up classification
1430 (3.4%)	Down classification

With events

Framingham Risk	Framingham Risk With CIMT		
	<5%	5%-20%	>20%
<5%	537	67	–
5-20%	69	2410	102
>20%		85	737

Total with events, No. (%)

3684 (91.9%)	No change
169 (4.2%)	Up classification
154 (3.8%)	Down classification

Improving upon existing model(s)

Conclusion

The **added value of common CIMT** in 10-year risk prediction of cardiovascular events, in addition to the Framingham risk score, **is small and unlikely to be of clinical importance**

Den Ruijter et al. , JAMA 2012

Developing a new prediction model

Main opportunities

- Increase total sample size
  - Avoid overfitting
  - Investigate more complex associations
- Increase available case-mix variability
  - Improve generalizability of risk predictions
  - Assess model performance across different settings and populations

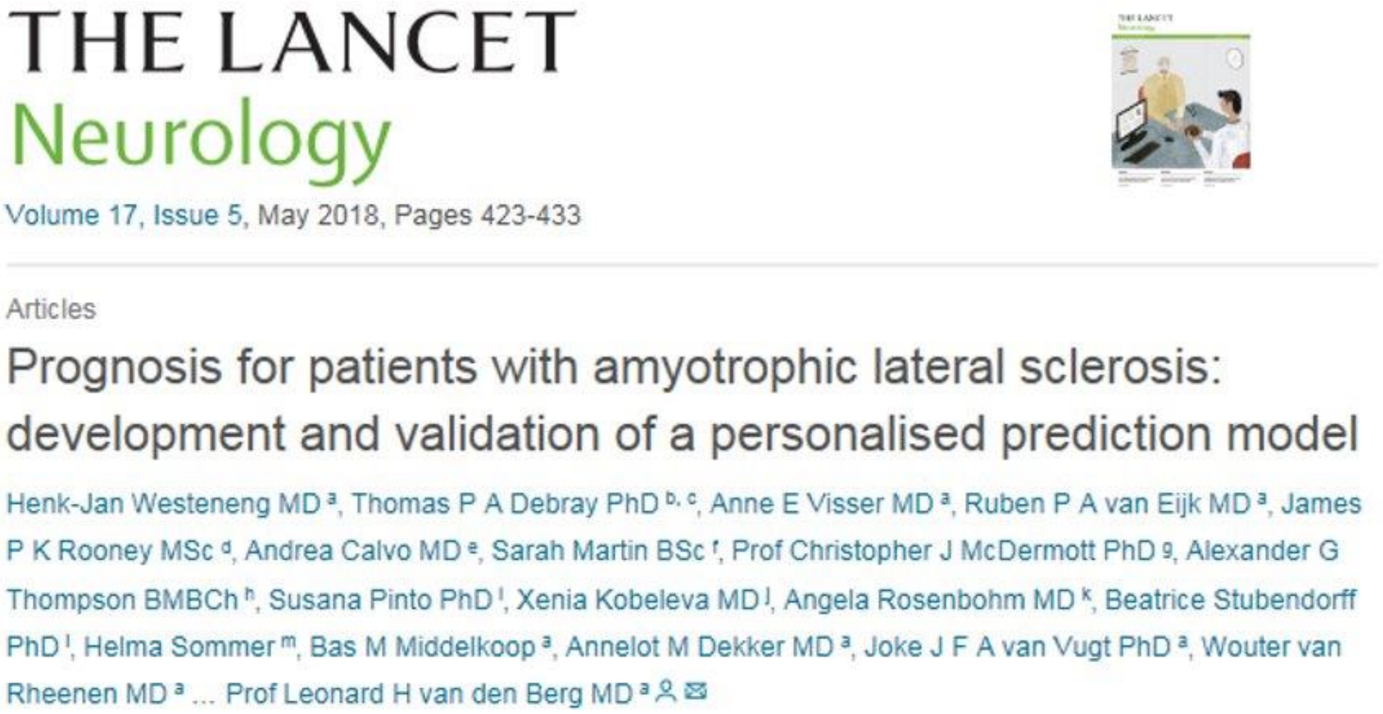
Developing a new prediction model

Prognosis of amyotrophic lateral disease

- IPD-MA
  - – 14 cohort studies (specialized ALS centres)
- Sample size
  - – 190 to 1,936 per study (total N = 11,475)
- Composite endpoint
  - – Non-invasive ventilation for more than 23h/day, or death
  - – Total number of events E = 8,819
- Median follow-up: 97.5 months

Development of the NCALS model

Developing a new prediction model



THE LANCET
Neurology

Volume 17, Issue 5, May 2018, Pages 423-433

Articles

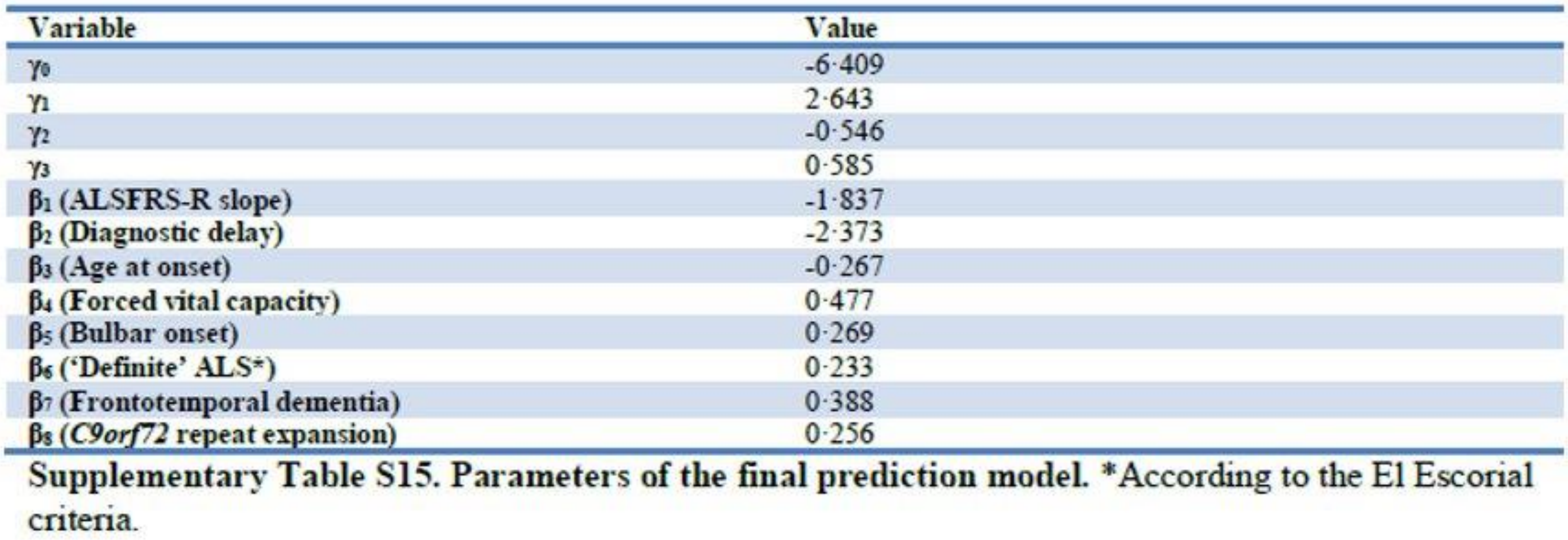
Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model

Henk-Jan Westeneng MD [a], Thomas P A Debray PhD [b, c], Anne E Visser MD [a], Ruben P A van Eijk MD [a], James P K Rooney MSc [d], Andrea Calvo MD [e], Sarah Martin BSc [f], Prof Christopher J McDermott PhD [g], Alexander G Thompson BMBCh [h], Susana Pinto PhD [i], Xenia Kobeleva MD [j], Angela Rosenbohm MD [k], Beatrice Stubendorff PhD [l], Helma Sommer [m], Bas M Middelkoop [a], Annelot M Dekker MD [a], Joke J F A van Vugt PhD [a], Wouter van Rheenen MD [a] ... Prof Leonard H van den Berg MD [a] ✉

Developing a new prediction model

Prognosis of amyotrophic lateral disease

- Royston-Parmar survival model with country-specific (but proportional) baseline hazard



Variable	Value
γ_0	-6·409
γ_1	2·643
γ_2	-0·546
γ_3	0·585
β_1 (ALSFRS-R slope)	-1·837
β_2 (Diagnostic delay)	-2·373
β_3 (Age at onset)	-0·267
β_4 (Forced vital capacity)	0·477
β_5 (Bulbar onset)	0·269
β_6 ('Definite' ALS*)	0·233
β_7 (Frontotemporal dementia)	0·388
β_8 (*C9orf72* repeat expansion)	0·256

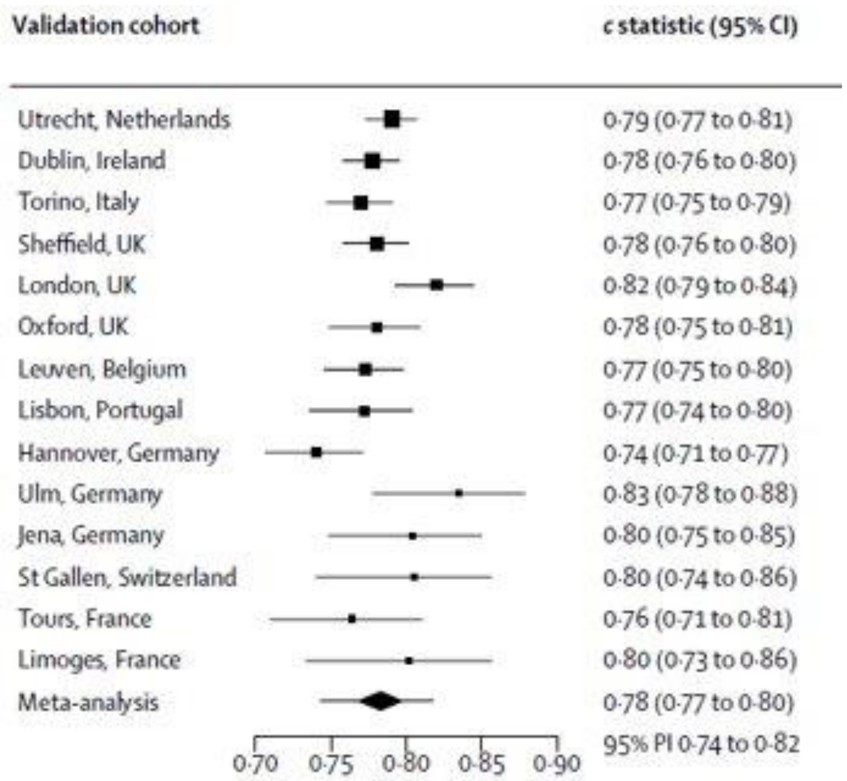
Supplementary Table S15. Parameters of the final prediction model. *According to the El Escorial criteria.

Developing a new prediction model

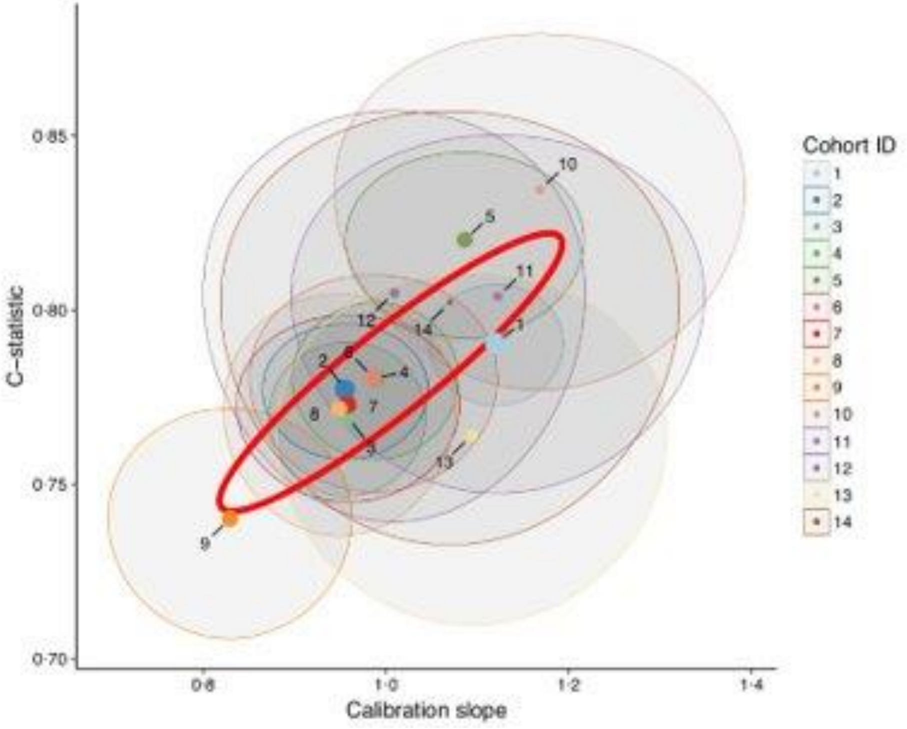
Iteratively develop pre-defined model in 13 studies, and externally validate in remaining study (**Internal-external cross-validation**)

- Meta-analysis of concordance statistic
  - Summary estimate: **0.78** (0.77 to 0.80)
  - 95% PI: 0.74 to 0.82
- Meta-analysis of calibration-in-the-large
  - Summary: **-0.12** (-0.33 to 0.08)
  - 95% PI: -0.88 to 0.63
- Meta-analysis of calibration slope
  - Summary: **1.01** (0.95 to 1.07)
  - 95% PI: 0.83 to 1.18

Developing a new prediction model



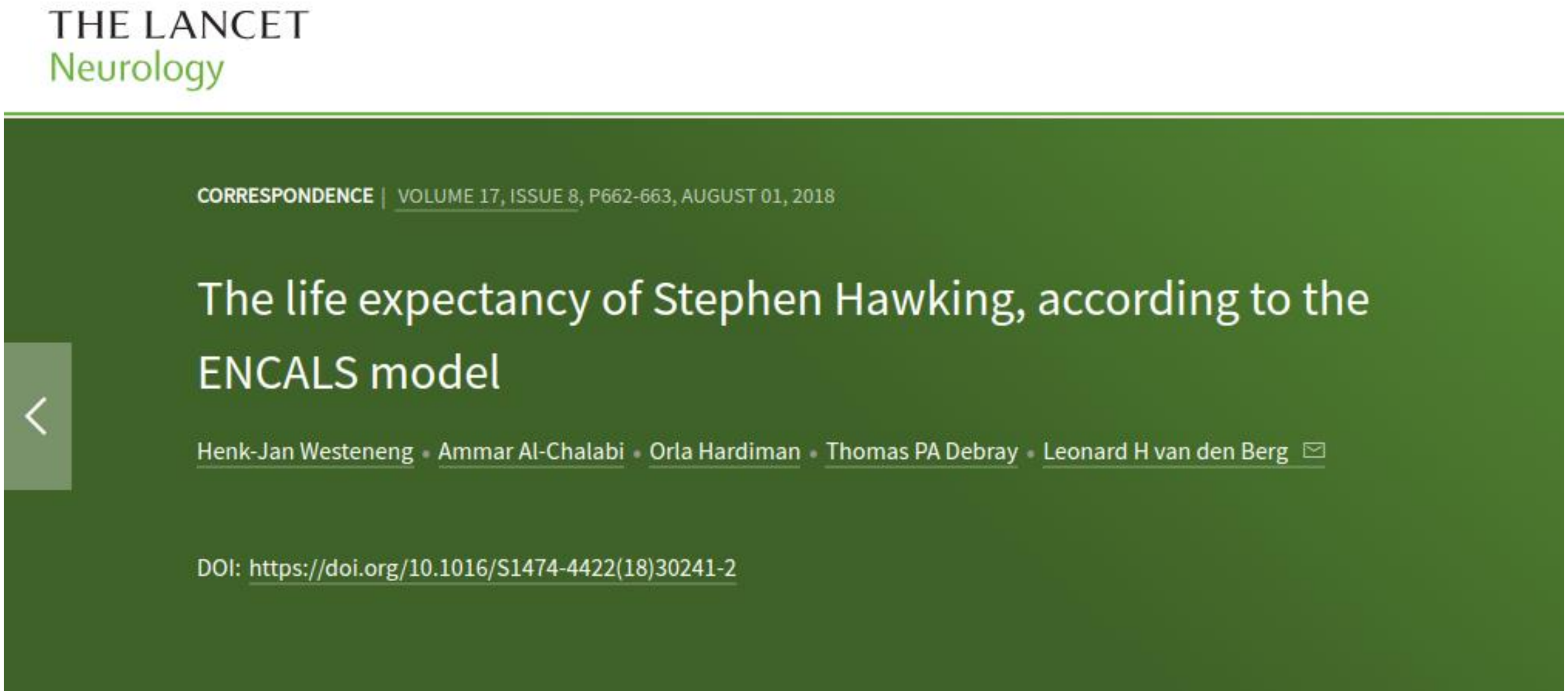
Validation cohort	c statistic (95% CI)
Utrecht, Netherlands	0·79 (0·77 to 0·81)
Dublin, Ireland	0·78 (0·76 to 0·80)
Torino, Italy	0·77 (0·75 to 0·79)
Sheffield, UK	0·78 (0·76 to 0·80)
London, UK	0·82 (0·79 to 0·84)
Oxford, UK	0·78 (0·75 to 0·81)
Leuven, Belgium	0·77 (0·75 to 0·80)
Lisbon, Portugal	0·77 (0·74 to 0·80)
Hannover, Germany	0·74 (0·71 to 0·77)
Ulm, Germany	0·83 (0·78 to 0·88)
Jena, Germany	0·80 (0·75 to 0·85)
St Gallen, Switzerland	0·80 (0·74 to 0·86)
Tours, France	0·76 (0·71 to 0·81)
Limoges, France	0·80 (0·73 to 0·86)
Meta-analysis	0·78 (0·77 to 0·80)
	95% PI 0·74 to 0·82



Developing a new prediction model

Measure	Criteria	Prob. of “good”performance	Joint probability
Concordance statistic	> 0.70	100%	98.3%
Calibration slope	0.80 to1.20	97.1%	
Calibration-in-the-large	-0.587 to 0.587	85.5%	

Developing a new prediction model



THE LANCET
Neurology

CORRESPONDENCE | VOLUME 17, ISSUE 8, P662-663, AUGUST 01, 2018

The life expectancy of Stephen Hawking, according to the ENCALS model

Henk-Jan Westeneng • Ammar Al-Chalabi • Orla Hardiman • Thomas PA Debray • Leonard H van den Berg ✉

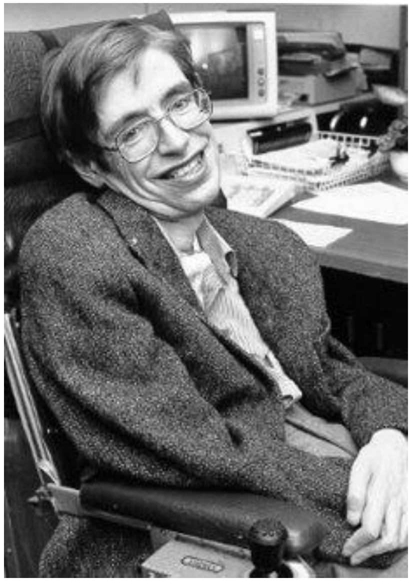
DOI: https://doi.org/10.1016/S1474-4422(18)30241-2

Developing a new prediction model

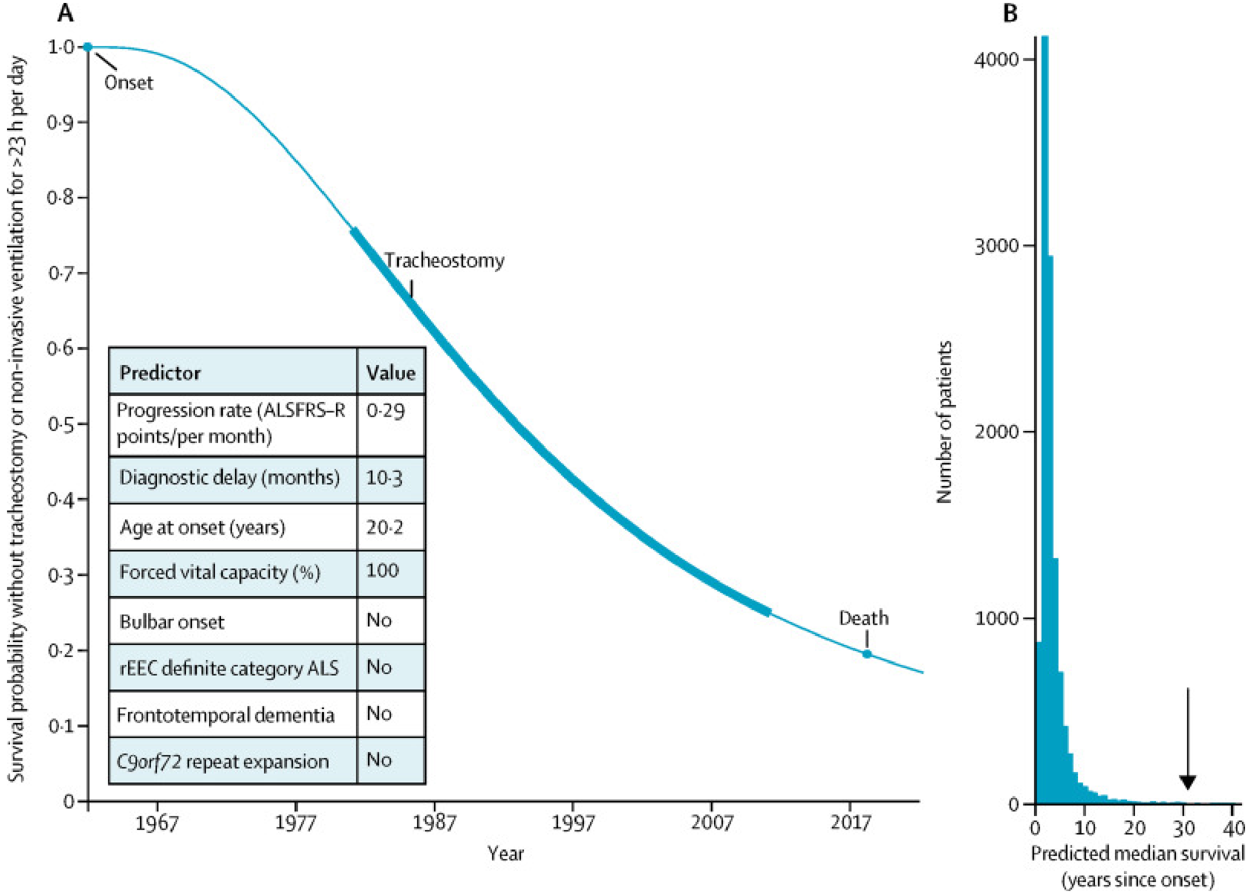
The life expectancy of Stephen Hawking, according to the ENCALS model

"Using publicly available data, we examined whether Professor Hawking's survival was as rare as his intellectual performance, or could be predicted solely based on his disease characteristics at diagnosis in 1963."

- Predicted 10-year survival probability: 94%
- The IQR for his predicted survival lay between 1981 and 2011
- Young age of onset was the most important factor for his long survival



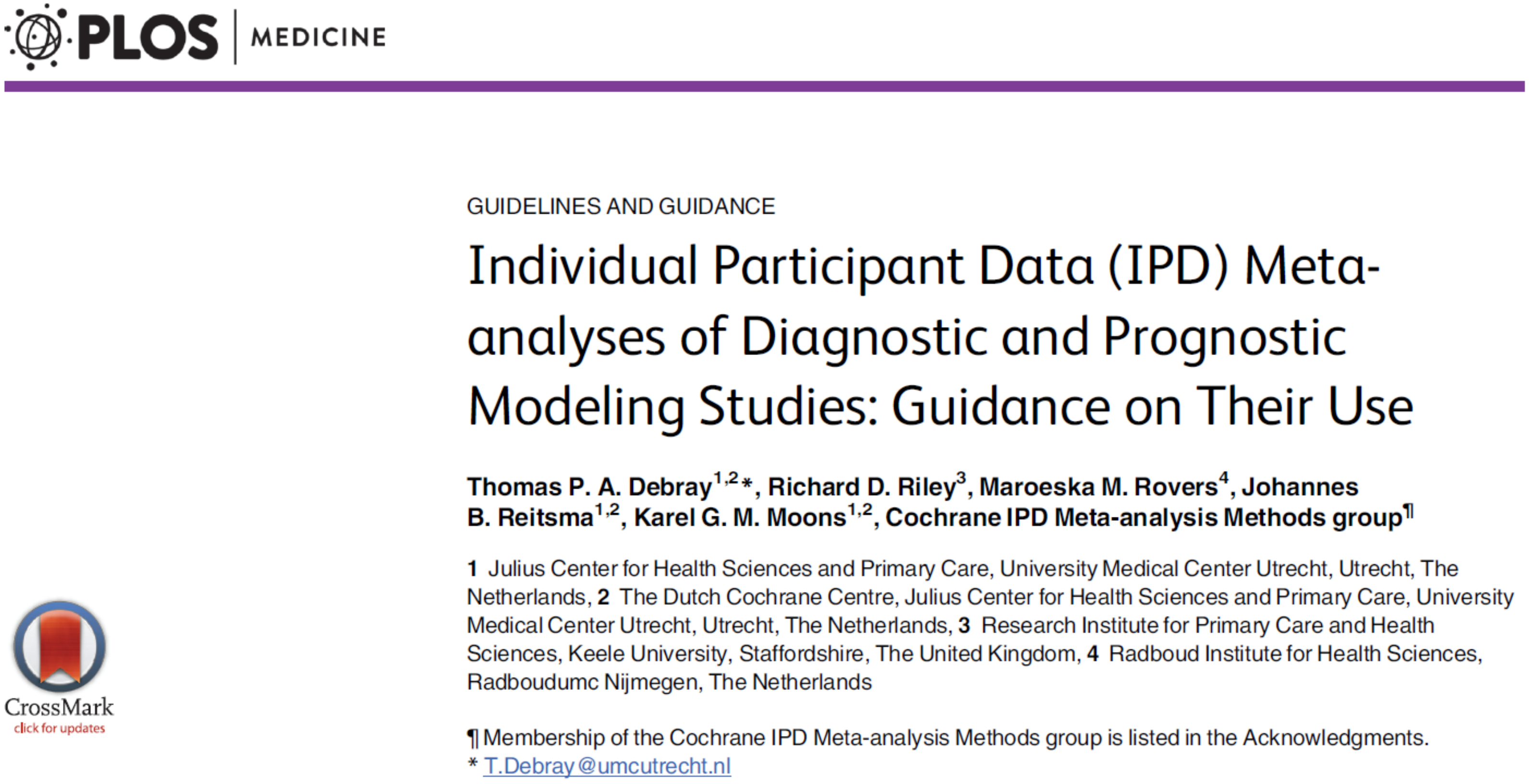
Developing a new prediction model



Predictor	Value
Progression rate (ALSFRS–R points/per month)	0·29
Diagnostic delay (months)	10·3
Age at onset (years)	20·2
Forced vital capacity (%)	100
Bulbar onset	No
rEEC definite category ALS	No
Frontotemporal dementia	No
C9orf72 repeat expansion	No

Personalised survival curve for Stephen Hawking (A) and comparison with other patients with ALS (B)

Statistical Methods



PLOS | MEDICINE

GUIDELINES AND GUIDANCE

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

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

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

¶ Membership of the Cochrane IPD Meta-analysis Methods group is listed in the Acknowledgments.

* T.Debray@umcutrecht.nl

CrossMark
click for updates

Take home messages

Major advantages IPD-MA

- Better insight in performance of prediction model(s) within and across different settings and populations
  - Quantify heterogeneity
  - Notably calibration

- Improving the performance of prediction model(s)
  - Tailoring of model(s)

- Integrate development and external validation when developing a new model

Take home messages

Remaining challenges in IPD meta-analysis

- IPD-MA no panacea against poorly designed primary studies
  - Prospective multi-center studies remain important

- Addressing heterogeneity in prediction model performance
  - One model fits all?
  - Role of received intervention(s)
  - Updating continuous process?

New methods are on their way!

Take home messages

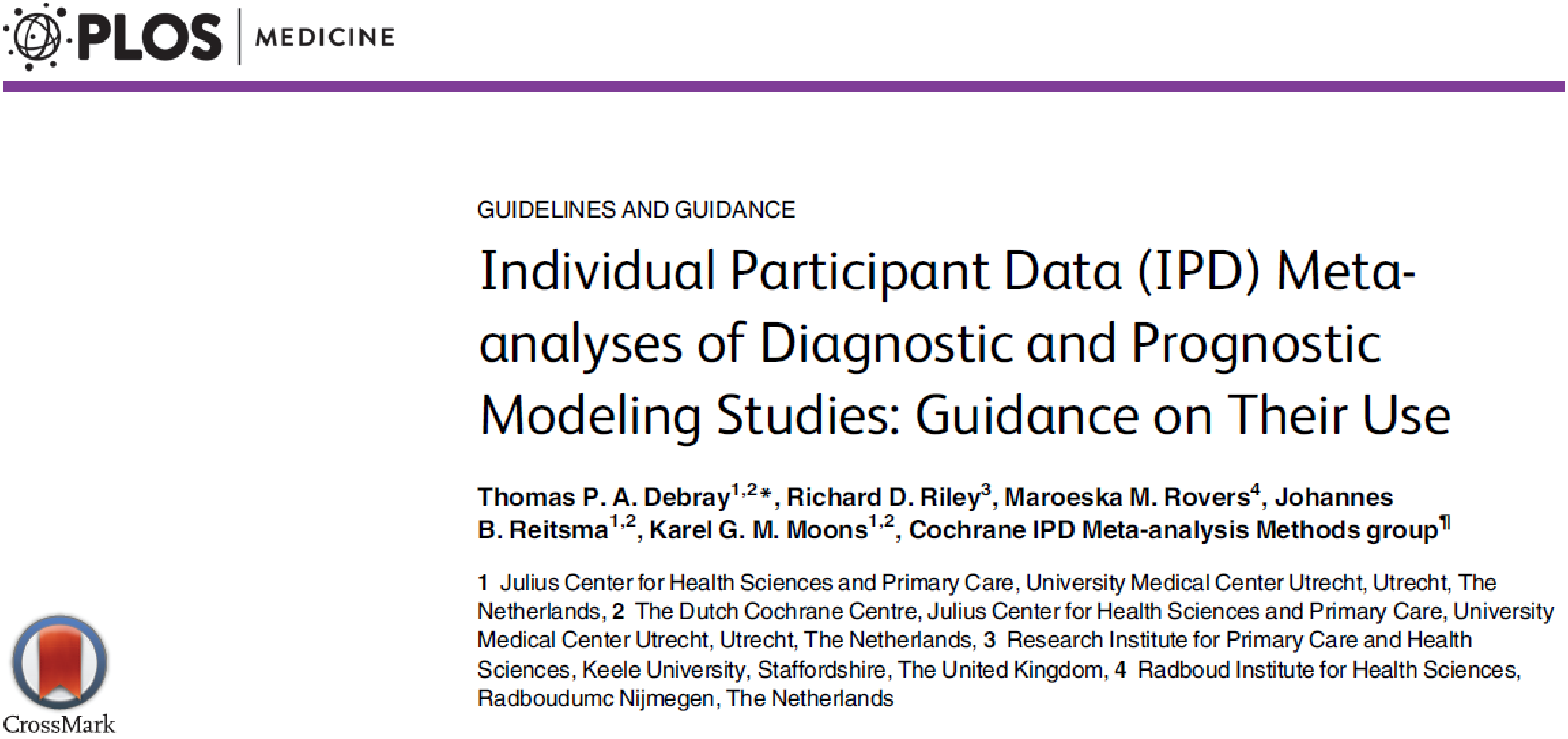
Reasons to be optimistic

Cochrane Prognosis Methods Group

- Aims to facilitate evidence-based prognosis research
- Improve design, quality & reporting of primary studies
- Facilitate systematic reviews & meta-analysis in long-run
- Bring together prognosis researchers, and guide Cochrane reviewers facing prognostic information
- Developing guidance

Take home messages

Reasons to be optimistic



PLOS | MEDICINE

GUIDELINES AND GUIDANCE

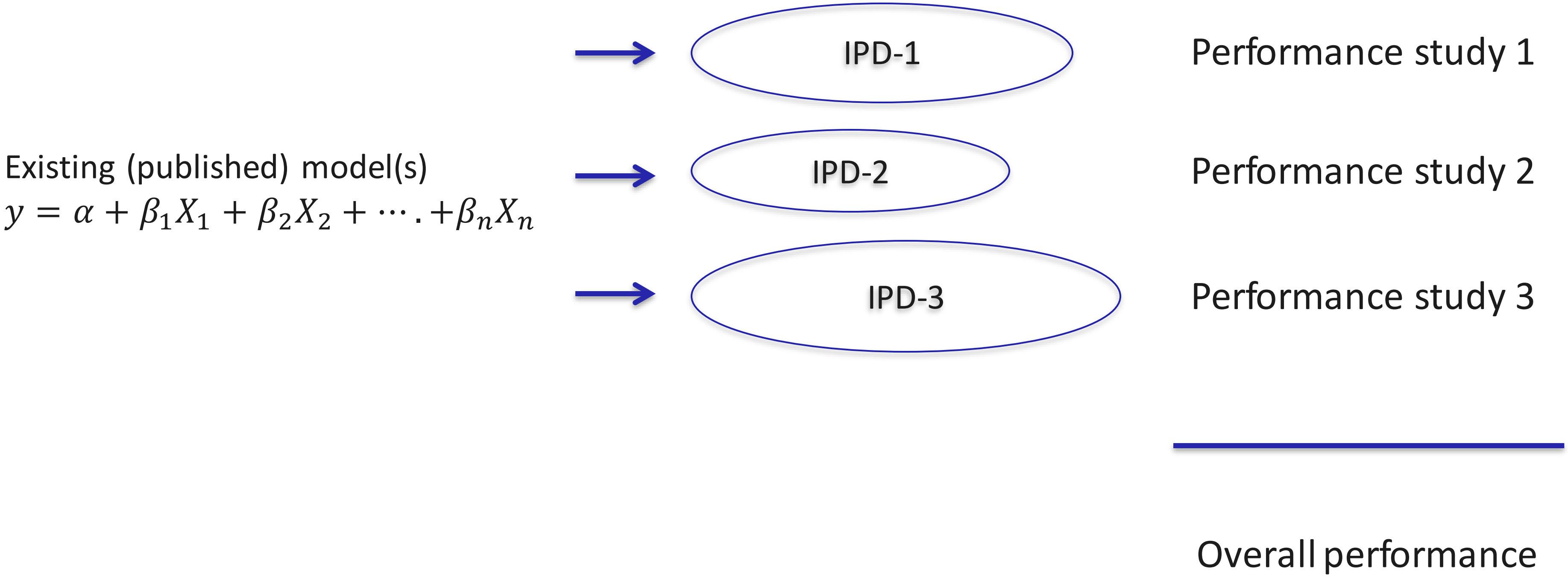
Individual Participant Data (IPD) Meta-analyses of Diagnostic and Prognostic Modeling Studies: Guidance on Their Use

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

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

CrossMark

Type I: Validation of existing model(s)



Existing (published) model(s)

$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots . + \beta_n X_n$

IPD-1

IPD-2

IPD-3

Performance study 1

Performance study 2

Performance study 3

Overall performance

Output:

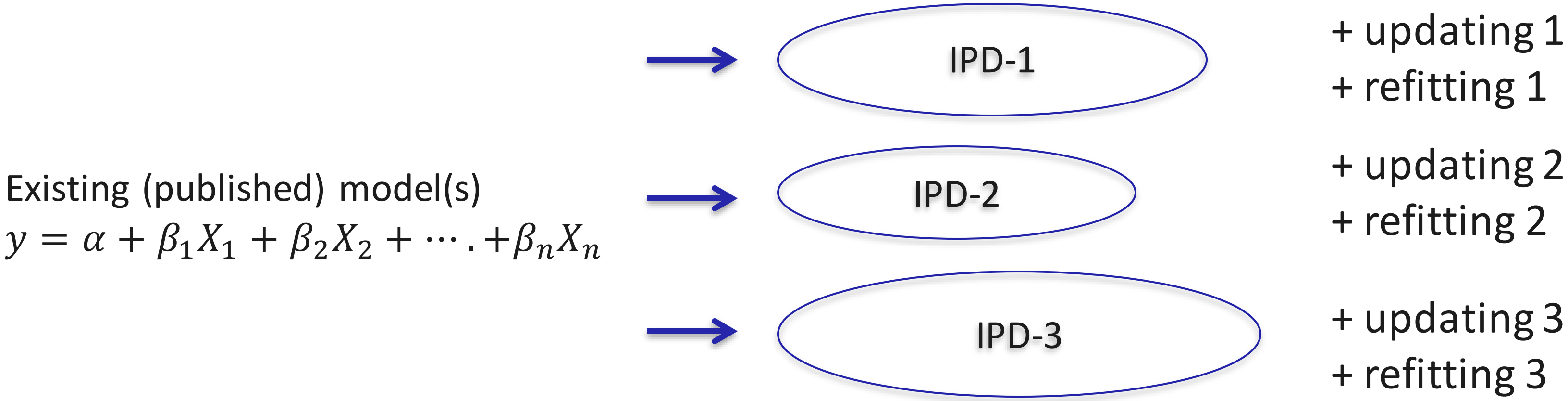
What is the overall performance?

How large is the heterogeneity?

What are drivers of heterogeneity?

Competing models: difference in performance?

TYPE II: TAILORING EXISTING MODEL



Existing (published) model(s)

$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots . + \beta_n X_n$

IPD-1

IPD-2

IPD-3

+ updating 1
+ refitting 1

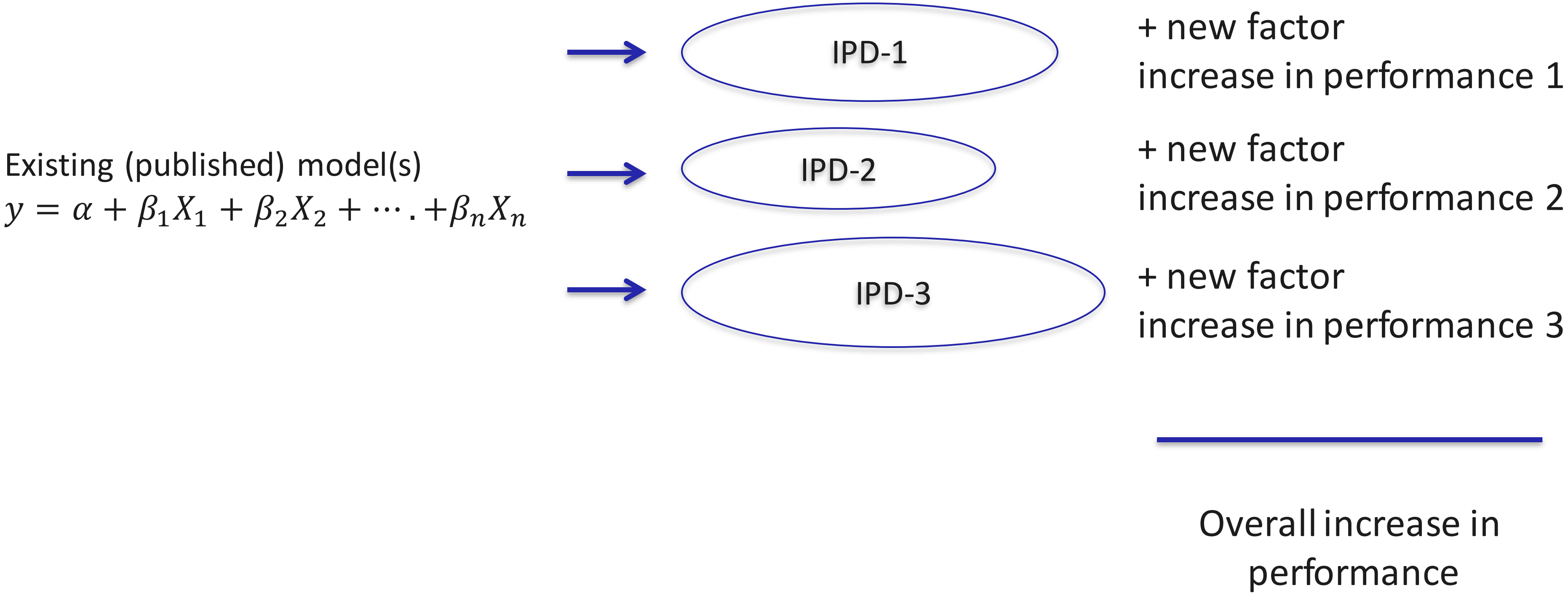
+ updating 2
+ refitting 2

+ updating 3
+ refitting 3

Updating needed?
Refitting needed?

Output:
Updating needed?
For which setting / populations
Updated model(s)

TYPE III: EXAMINING ADDED VALUE



Existing (published) model(s)

$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots . + \beta_n X_n$

IPD-1

+ new factor
increase in performance 1

IPD-2

+ new factor
increase in performance 2

IPD-3

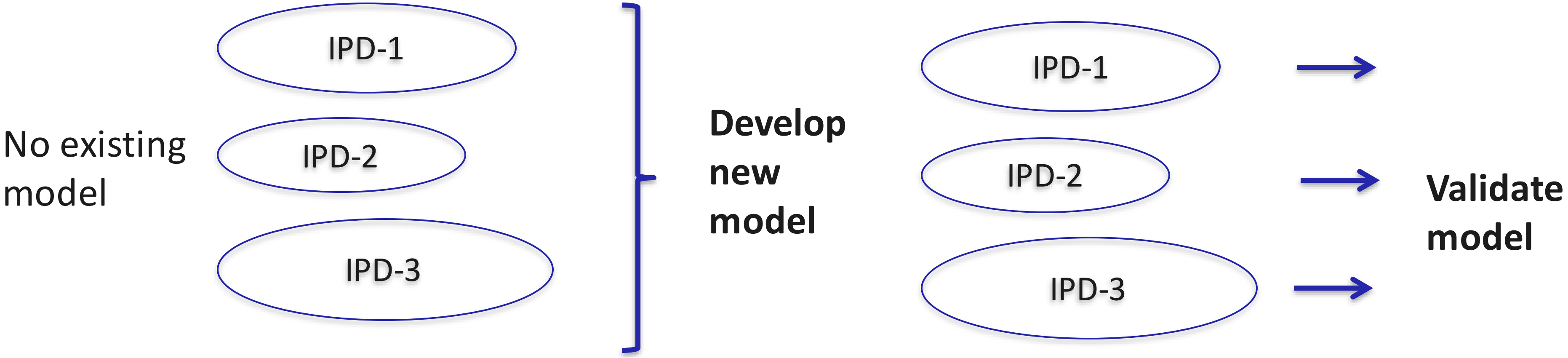
+ new factor
increase in performance 3

Overall increase in performance

Output:

What is the overall added value?
Heterogeneity in added value?
Drivers of heterogeneity?
What is the updated model?

TYPE IV: DEVELOPMENT NEW MODEL AND VALIDATION

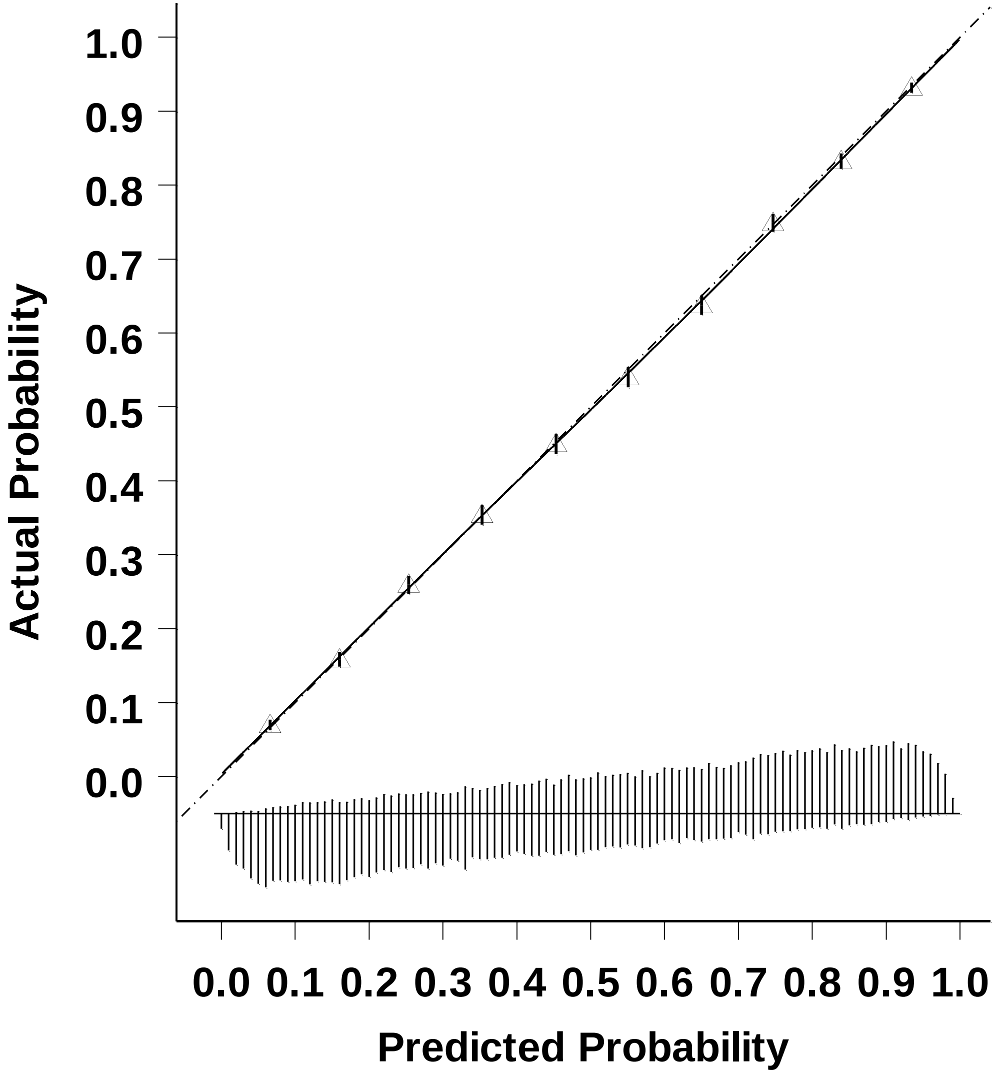


Output:
New model / tailored models

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
  - Assess the potential effect on patient-level outcomes
  - Comparative test accuracy studies
  - Examples: Net Reclassifiation Index, Net Benefit, …

Calibration plot

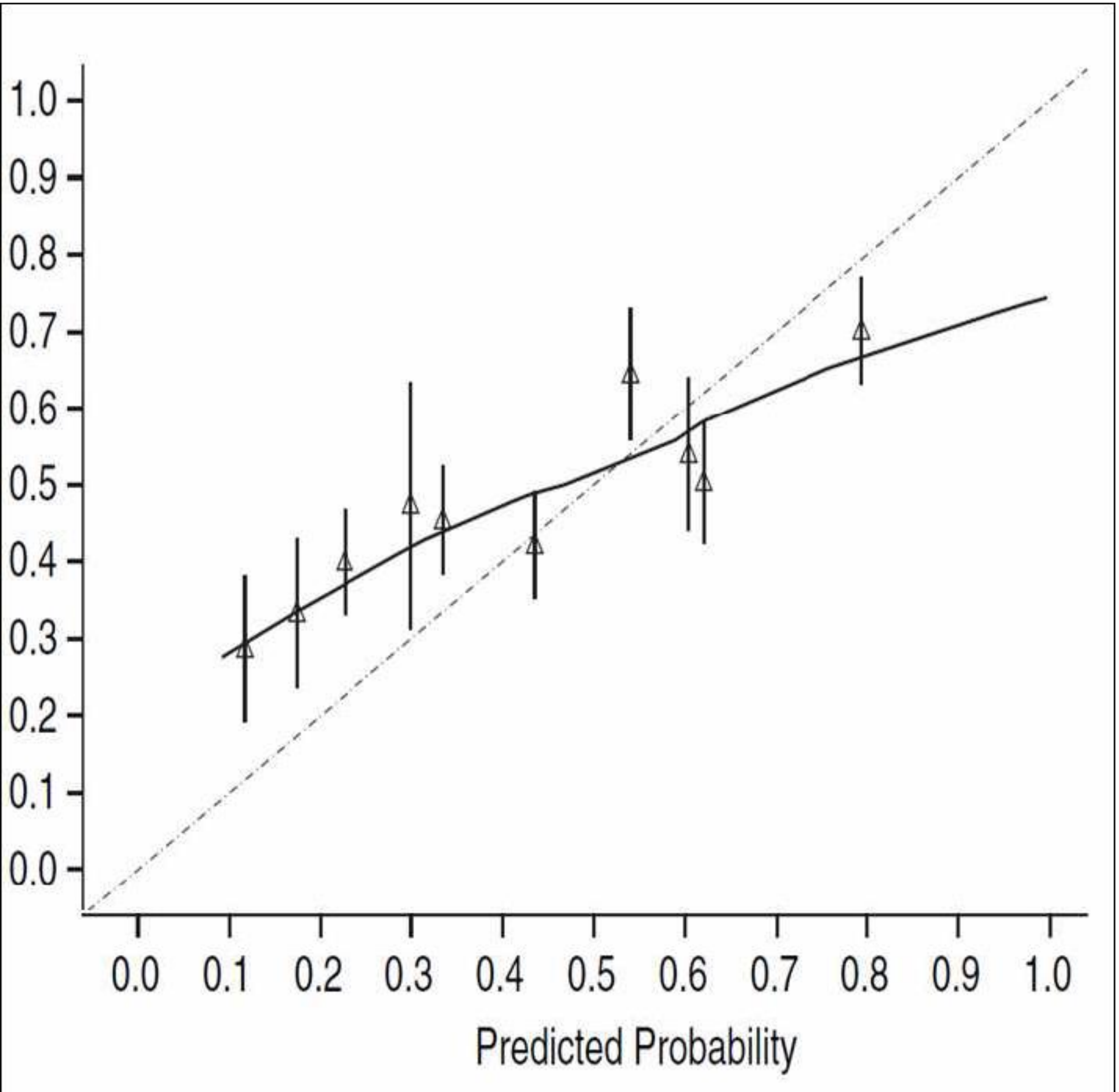


Ideal calibration

Observed versus expected risk (O/E) = 1

Slope = 1

External validation: typical result



- Slope plot < 1.0
  - Low prob too low
  - High prob too high
    - Overfitted

- AUC= 0.63 (was 0.75)