UMC Utrecht
Julius Center

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

Thomas Debray, PhD

Julius Center for Health Sciences and Primary Care
Utrecht, The Netherlands

Cochrane

March 7, 2016

University Medical Center Utrecht

UMC Utrecht

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

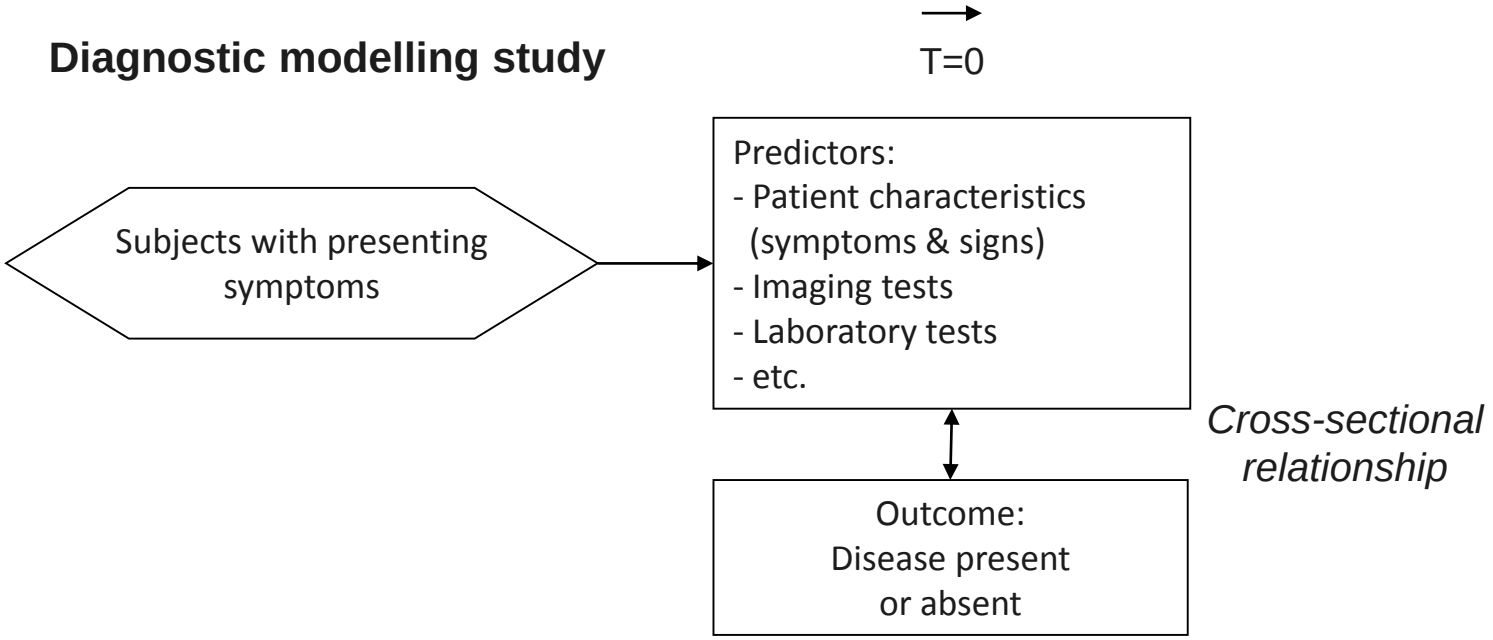
Prediction

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

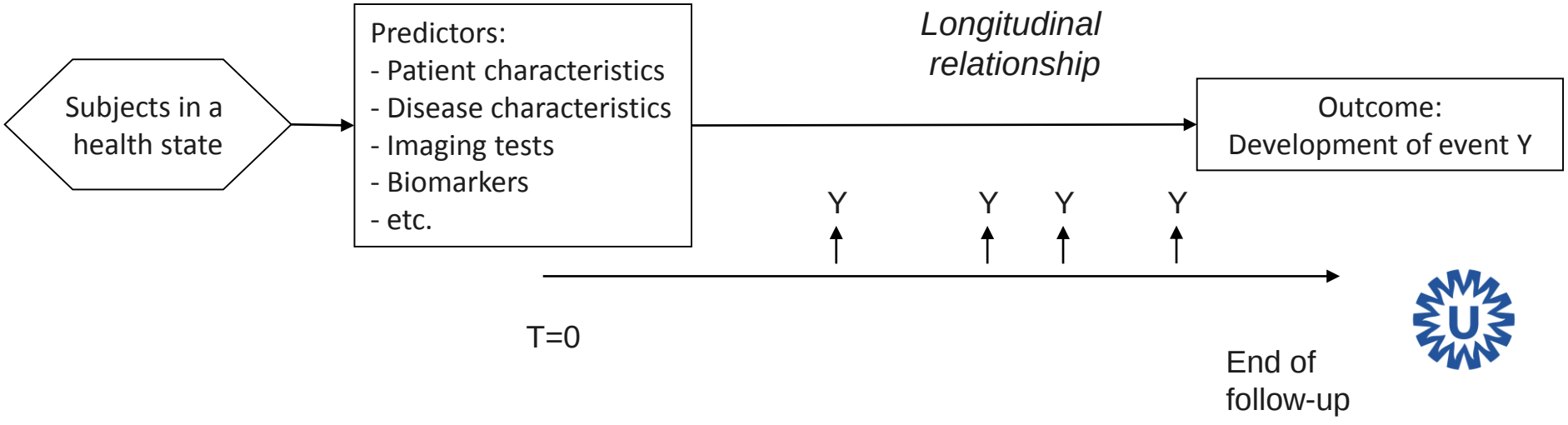
- Turn available information (predictors) into a statement about the probability:
  … of having a particular disease -> **diagnosis**
  … of developing a particular event -> **prognosis**

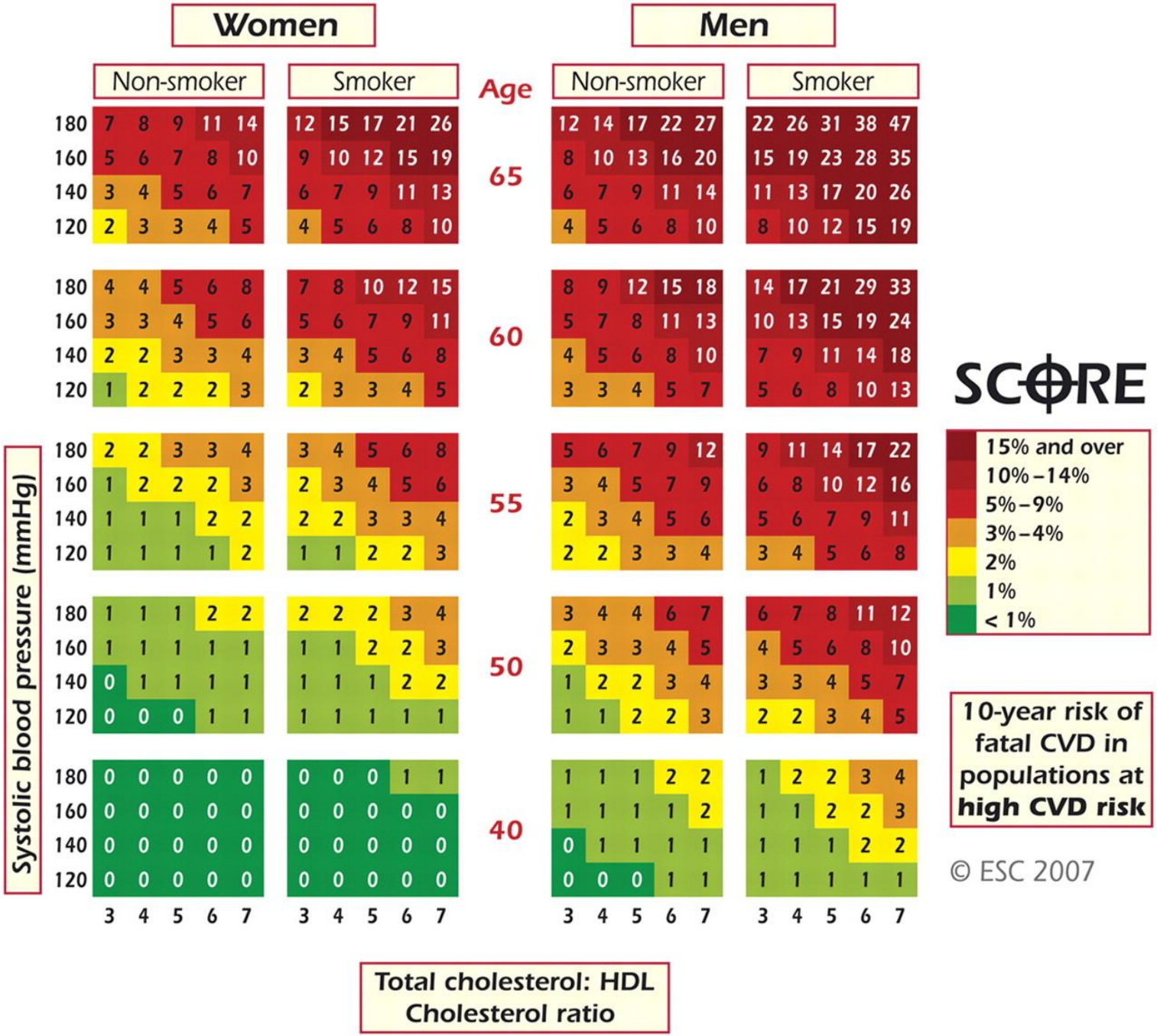
- Use of prognostic information:
  - to inform patients and their families
  - to guide treatment and other clinical decisions
  - to create risk groups
  - …

Diagnostic modelling study



Prognostic modelling study




Women
Men
Non-smoker
Smoker
Non-smoker
Smoker
Age
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
3 4 5 6 7 | 3 4 5 6 7 | 3 4 5 6 7 | 3 4 5 6 7
Systolic blood pressure (mmHg)
Total cholesterol: HDL Cholesterol ratio
SCORE
15% and over
10% – 14%
5% – 9%
3% – 4%
2%
1%
< 1%
10-year risk of fatal CVD in populations at high CVD risk
© ESC 2007

Survival curves / Kaplan Meier

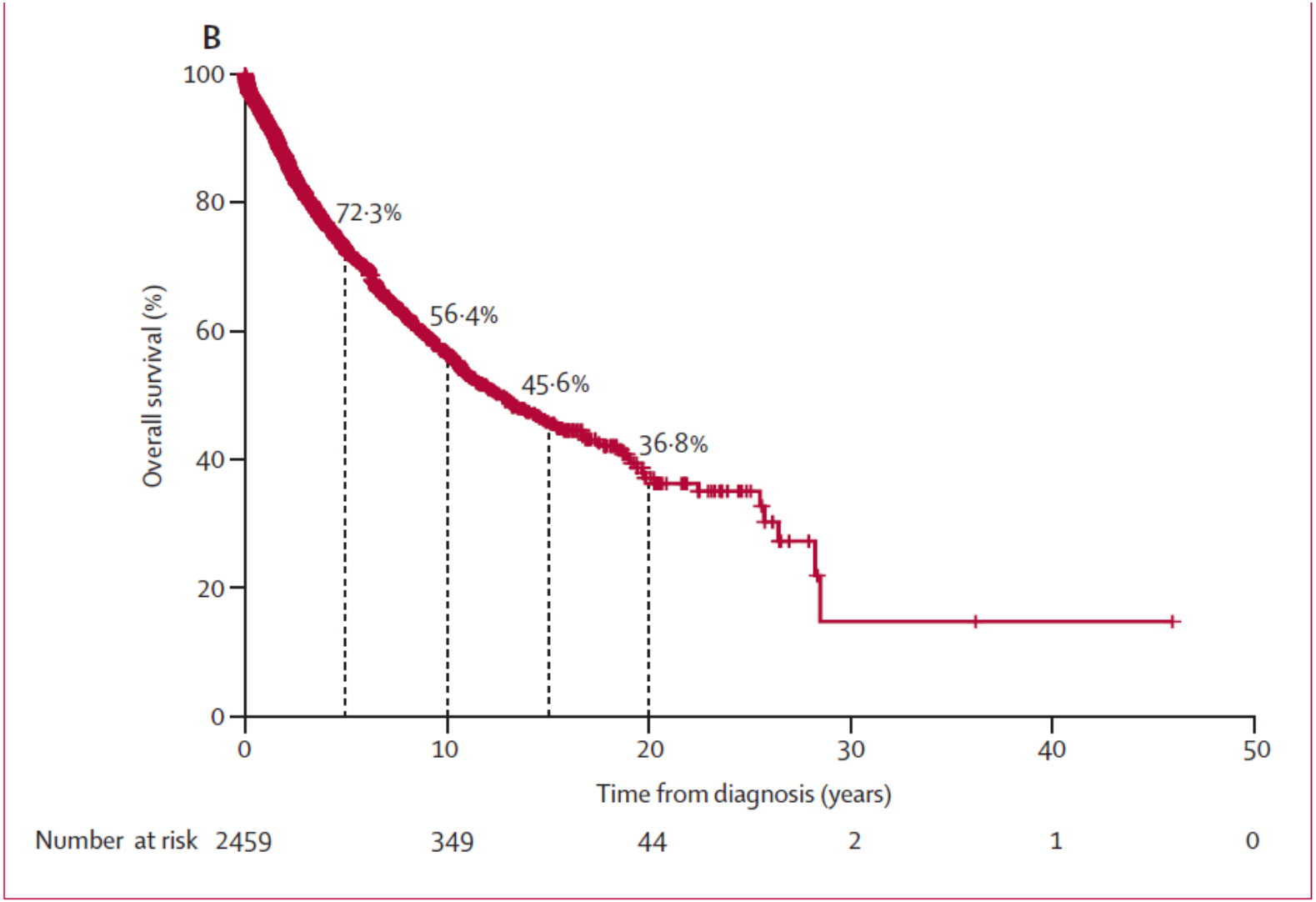
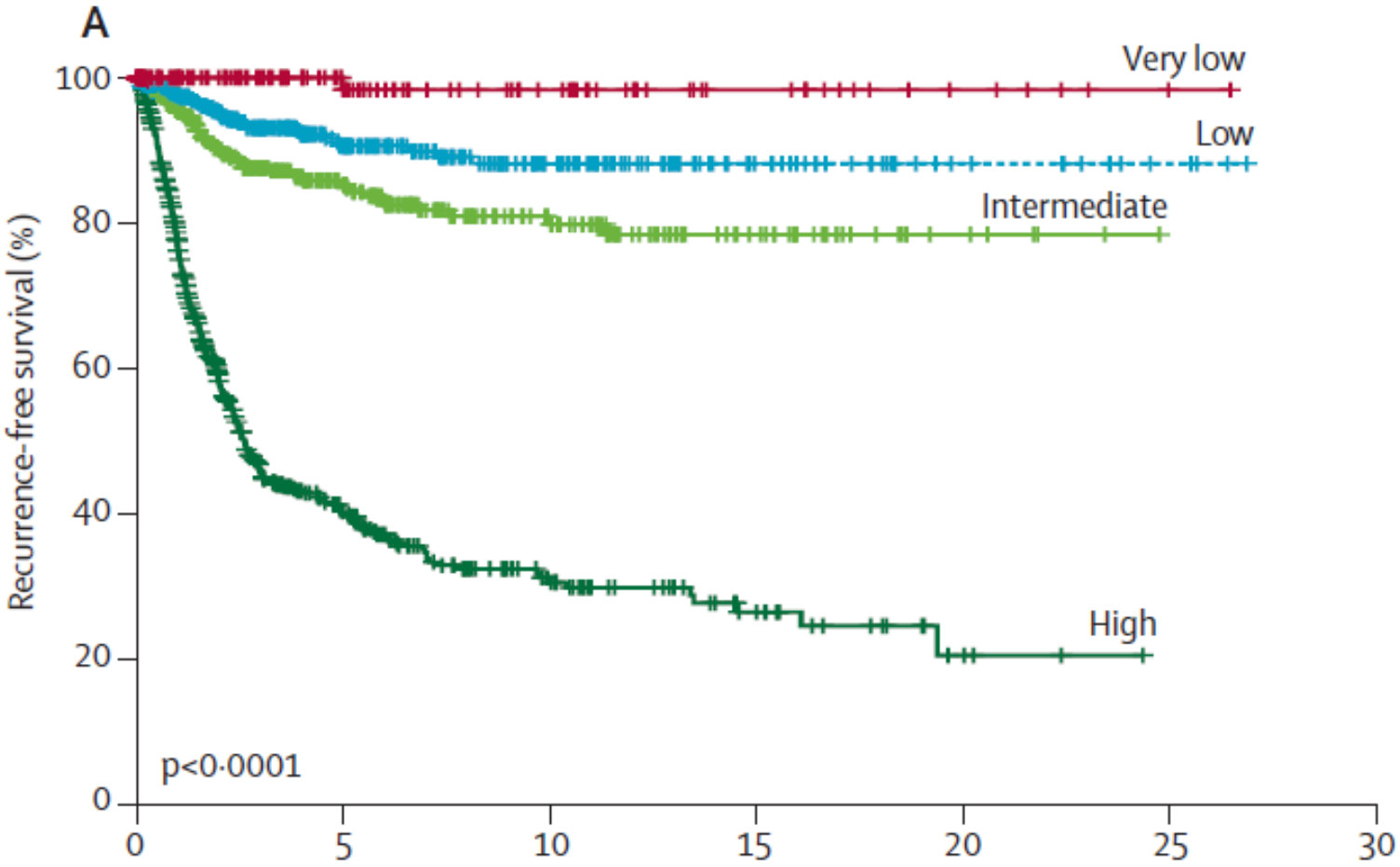


Figure 1: Recurrence-free survival (A) and overall survival (B) in the pooled series



Number at risk

Very low risk	130	59	33	16	6	1	0
Low risk	429	164	74	32	12	4	0
Intermediate risk	395	161	71	29	6	0	0
High risk	554	122	47	20	4	0	0

Prediction

What is a good model?

- Generates accurate predictions in individuals from potential population(s) for clinical use
- Ability to discriminate between different risk groups
- Improves patient outcomes by informing treatment decisions

Prediction problems

Most models are not as good as we think

- Quality of many prognostic modelling studies is poor
  - Limited sample size
  - Incomplete registrations & reporting
  - Absent study protocols
- Transportability of many models is limited
  - Case-mix variation across populations
  - Differences in measurement methods
  - Time-varying predictor effects
  - Changes in standards of care and treatment strategies
- Lack of external validation

The rise of big data

What is 'big data'?

- Meta-analysis of individual participant data (IPD) from multiple studies
- Analyses of databases and registry data containing e-health records

Data for thousands or even millions of patients from multiple practices, hospitals, or countries.

Example: QRISK2 was developed using e-health data from the QRESEARCH database using over 1.5 million patients (with over 95000 new cardiovascular events) from 355 randomly selected general practices

Prediction research using big data

Why do we need 'big data'?

- Development of better prediction models
  - Reduced risk of overfitting
  - Ability to address wider spectrum of patients
  - Ability to investigate more complex associations

- More extensive testing of model performance
  - Ability to externally validate across multiple settings (also upon model development)
  - Ability to investigate sources of poor or inconsistent model performance
  - Ability to assess usability of prediction models across different situations

Prediction research using big data

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

Prediction research using big data

Main challenges

- Missing data
  - Partially missing data within studies
  - Systematically missing data within studies
  - Entire study missing (e.g. non-publication)
- Between-study heterogeneity
  - Outcome occurrence
  - Predictor effects
  - (Change in) model performance
- Combination of IPD and AD
  - Published prediction models
  - Published predictor effects
  - Published estimates of (increased) model performance

Prediction research using big data

What is heterogeneity?

Differences in outcome occurrence, predictor effects and/or model performance across studies, settings, …

- Case-mix variation (spectrum effect)
- Missed interactions and non-linear trends of predictors
- Biomarkers: different measurement method, recording time point or cut-off across settings
- Different standards of care and treatment strategies
- Different startpoints (e.g. due to screening)

Typically, heterogeneity is explored using meta-analysis methods with mixed or random effects

Model development & validation

Dealing with heterogeneity

UMC Utrecht

Model development using big data

Problem: random effects summaries are of limited value

- Predictor effects and/or baseline risk may take different values for each included study
- Which parameters to use when validating/implementing the model in new individuals or study populations?
- When do study populations differ too much to combine?

Need for a framework that can identify the extent to which aggregation of IPD is justifiable, and provide the optimal approach to achieve this.

Model development using big data

Recommendations from Debray *et al* & Ahmed *et al*.

- **Allow for different baseline risks in each of the IPD studies**
  - Account for differences in outcome prevalence (or incidence) across studies
  - Examine between-study heterogeneity in predictor effects and prioritize inclusion of (weakly) homogeneous predictors
  - Appropriate intercept for a new study can be selected using information on outcome prevalence (or incidence)
- **Implement a framework that uses internal-external cross-validation**

Model development using big data

Step 1: Different choices to combine IPD

- Merge all data into one big dataset and ignore heterogeneity
- Allow heterogeneous baseline risk across studies
  - assume random effects distribution for the intercept terms
  - estimate study-specific intercept terms
- Advanced modeling of predictor effects is also possible
  - Nonlinear effects
  - Interaction terms

Model development using big data

Step 2: Choosing an appropriate model intercept when implementing the model to new individuals

- Average intercept term
  (e.g. pooled estimate)
- Updating of intercept term
  (requires patient-level data)
- Use intercept of included study
  (e.g. based on outcome occurrence)

Propose which intercept term to use in new populations

!! More difficult in case of heterogeneous predictor effects

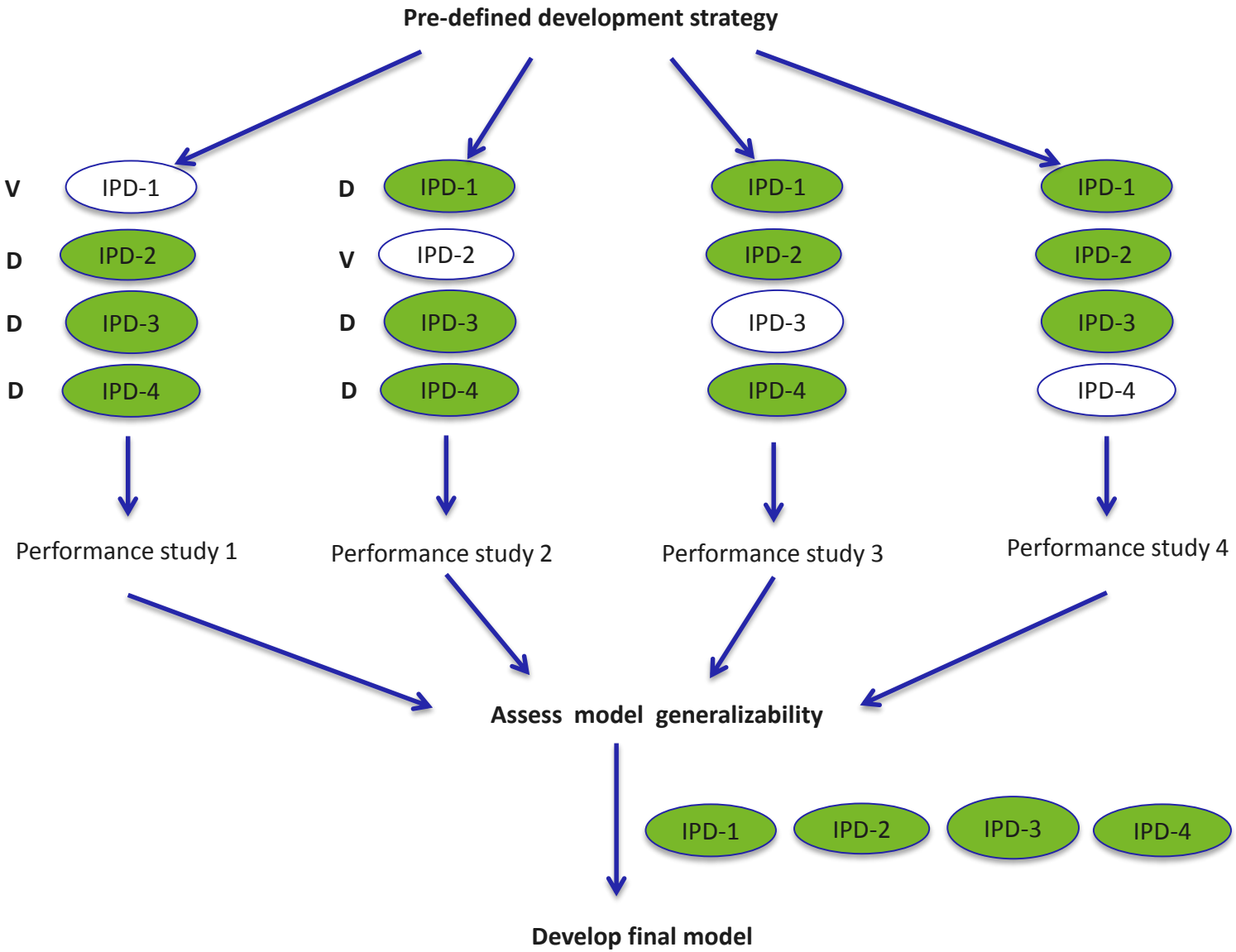
Model development using big data

Step 3: Model evaluation to check whether…

- Strategy for estimating predictors and intercept is adequate
- Strategy for choosing intercept term (and predictor effects) in new study population is adequate
- Model performance is consistently well across studies
  - Discrimination
  - Calibration

=> Use of internal-external cross-validation

Internal-external cross-validation



Model development using big data

Example 1 (diagnosis of deep vein thrombosis; N=12)

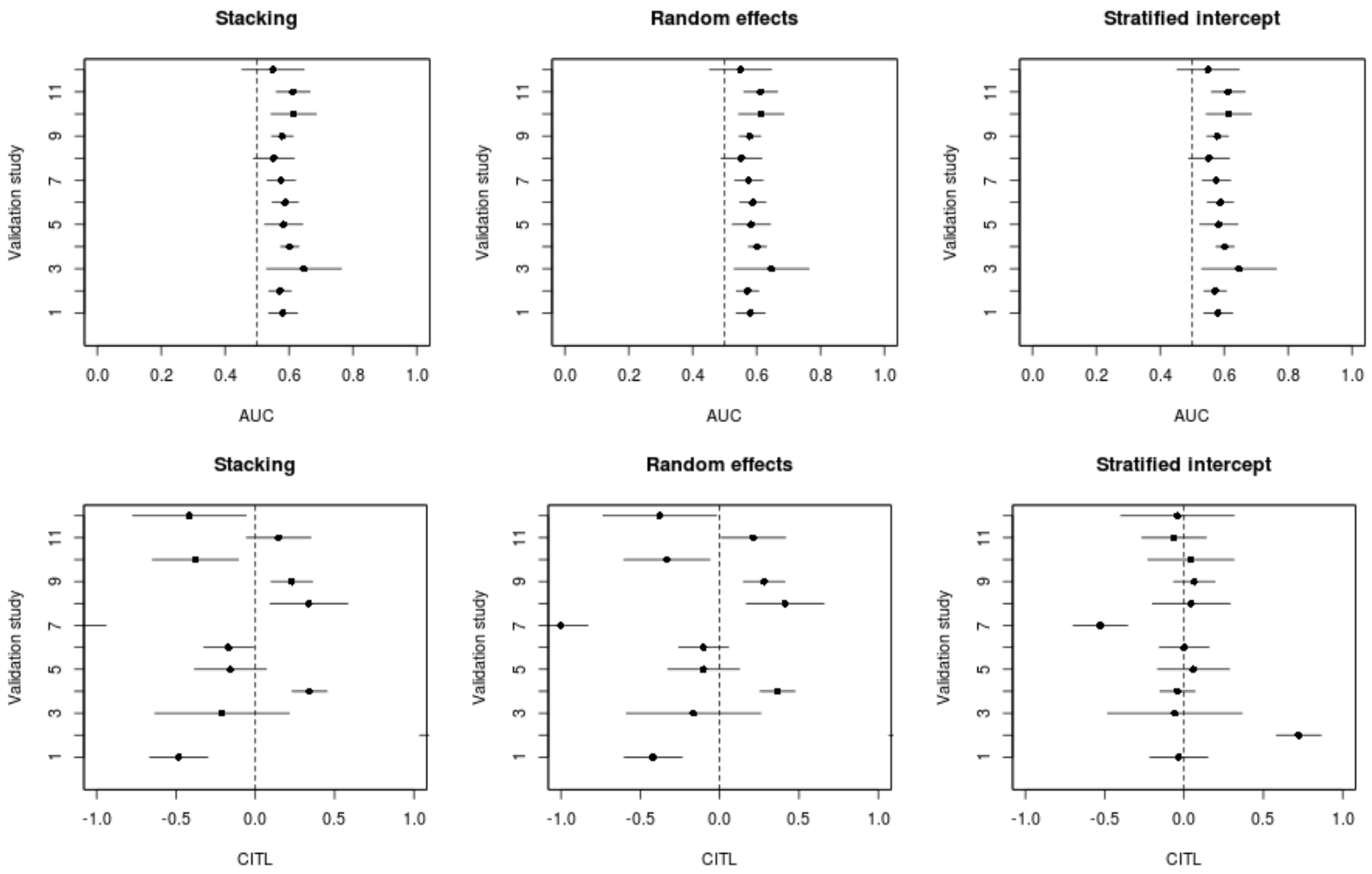
Strategies evaluated:

- Inclusion of 2 predictors (gender & recent surgery)
- Modelling of intercept term
  - Ignore heterogeneity (“stacking”)
  - Meta-analysis (“random effects”)
  - Stratify intercept term across studies
- Model implementation
  - Average intercept (stacking; random effects)
  - Select estimated intercept term based on outcome occurrence

Assessment of AUC and calibration-in-the-large (CITL)

Model development using big data

Example 1 (diagnosis of deep vein thrombosis; N=12)

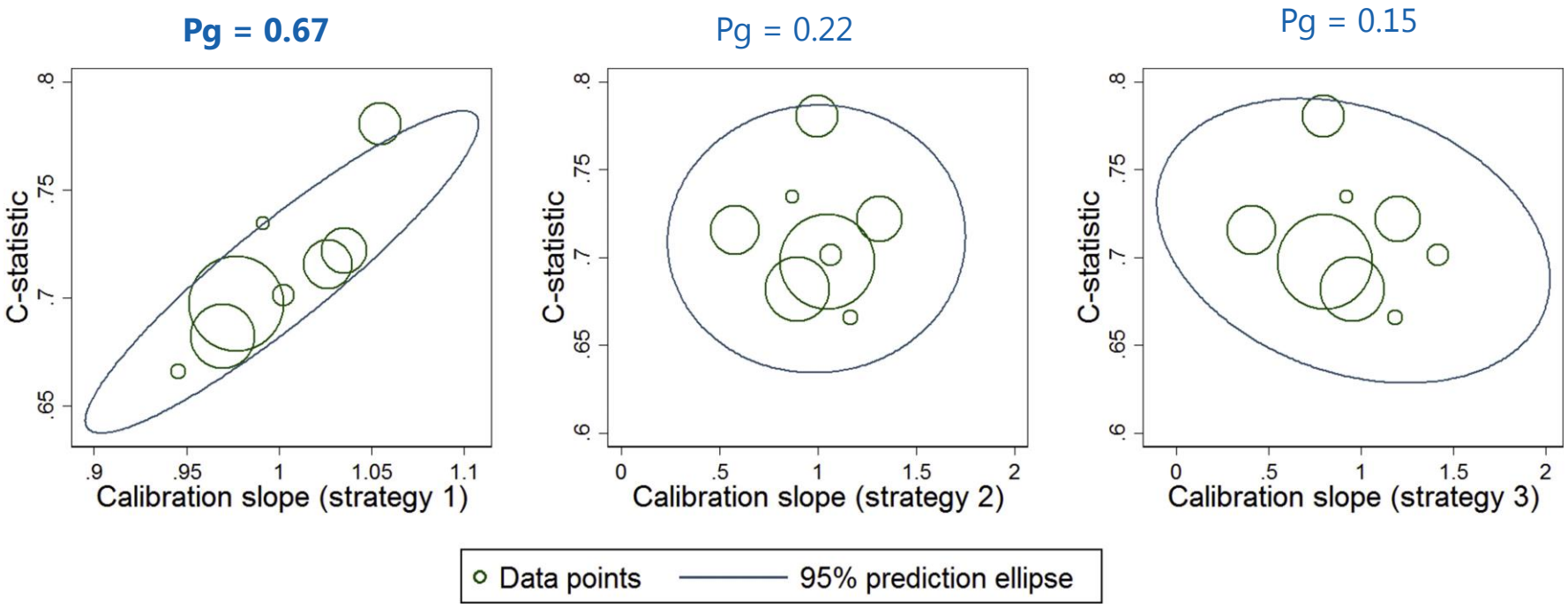


Model development using big data

Example 2 (prognosis of breast cancer; N=8)

- Strategy 1: Develop using Royston-Parmar and implement with baseline hazard estimated in validation study
- Strategy 2: Develop using Royston-Parmar and implement with average baseline hazard from developed model
- Strategy 3: Develop using Royston-Parmar and implement with the estimated baseline hazard from the closest geographical country

Model development using big data



Pg = Joint probability of "good performance"
(C>= 0.7 and calibration slope between 0.9 and 1.1)

=> **Updating of baseline hazard recommended!**

Model validation using big data

Validation of existing model(s)

- Evaluate model discrimination and calibration separately within each IPD set
- Investigate whether model performance is adequate and consistent across populations/subgroups/settings
- Heterogeneity in performance can be expected!
  (not necessarily related to misfit of model coefficients)
  - Need to adjust for case-mix differences
  - Need to inspect relatedness between included populations
  - Random effects meta-analysis recommended
- If possible, make head-to-head comparisons with existing models
- Evaluate need for updating and required updating strategy

Missing data

Why is relevant in ‘big data’ and what can we do about it?

UMC Utrecht

Dealing with missing data

The problem of heterogeneity

- Traditional imputation methods do not (properly) accommodate for differences in predictive associations
- As a result, imputation may mask the actual extent of between-study heterogeneity

=> Subsequent analyses (e.g. model development and validation) may lead to over-optimistic results!

Dealing with missing data

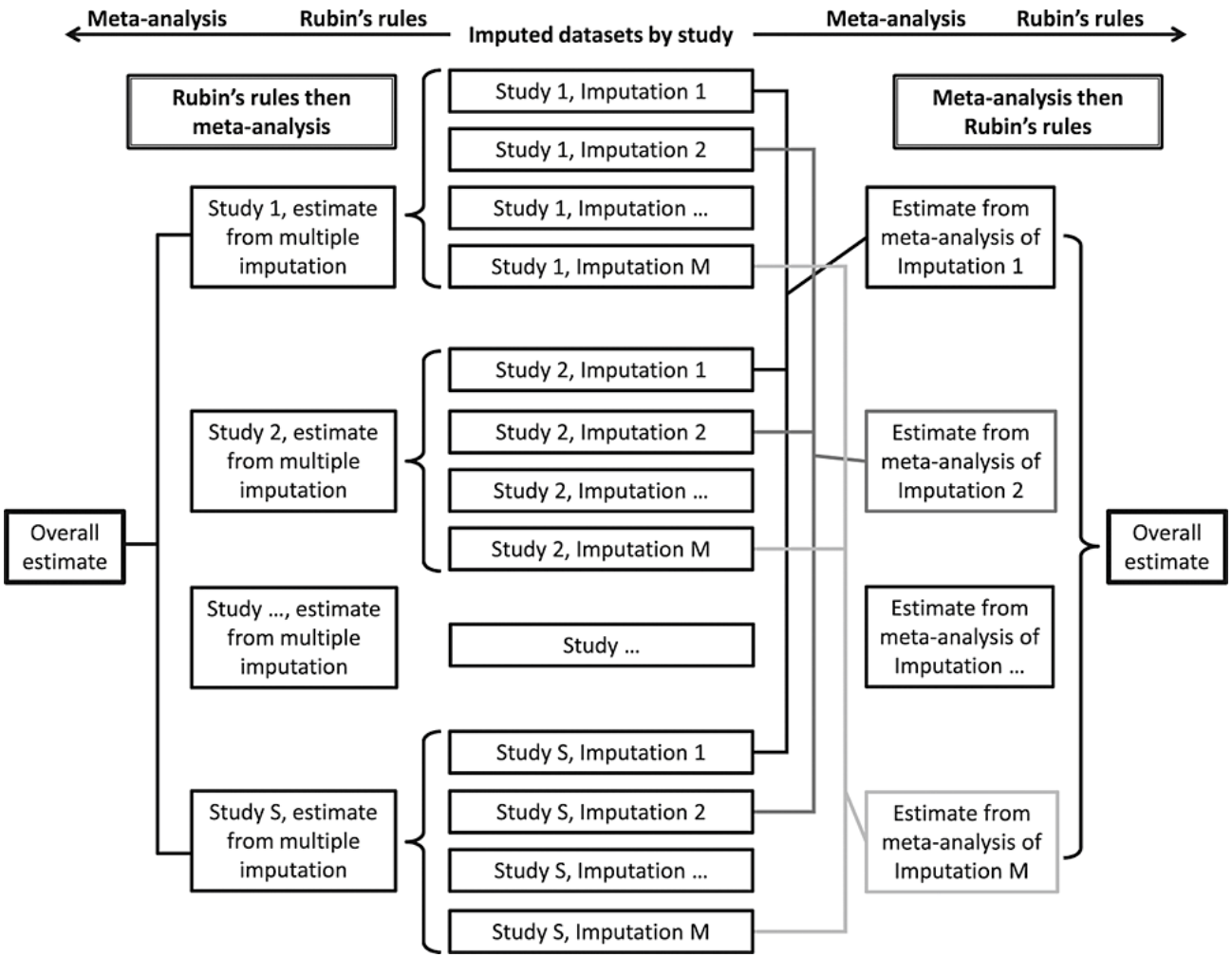
Partially missing data (missing at random)

- Imputation
  - Two-stage (stratified per study)
    - Fully stratified vs. Stratification of intercept term only
  - One-stage (hierarchical approach)
    - Homoscedastic vs. Heteroscedastic error variance
    - Fully Bayesian vs. Large sample approximations
- Analysis
  - Two-stage (stratified per study)
  - One-stage (hierarchical approach)

Note that several combinations are possible!

Dealing with missing data

Partially missing data



Dealing with missing data

Partially missing data

- Acknowledging heterogeneity during imputation is paramount if subsequent analysis aims to explore its presence
  - Inappropriate to ignore clustering of subjects
  - Inappropriate to include a study dummy variable in the imputation model
- One-stage approach for imputation and analysis most powerful, but computationally very complex
- Two-stage imputation performs relatively well, and can be implemented fairly straightforward

Dealing with missing data

Systematically missing data

Two-stage imputation and traditional one-stage imputation no longer feasible (as within-study variance is unidentifiable)

⇒ Need for more advanced one-stage imputation methods

⇒ Implementation of generalized linear mixed effect model

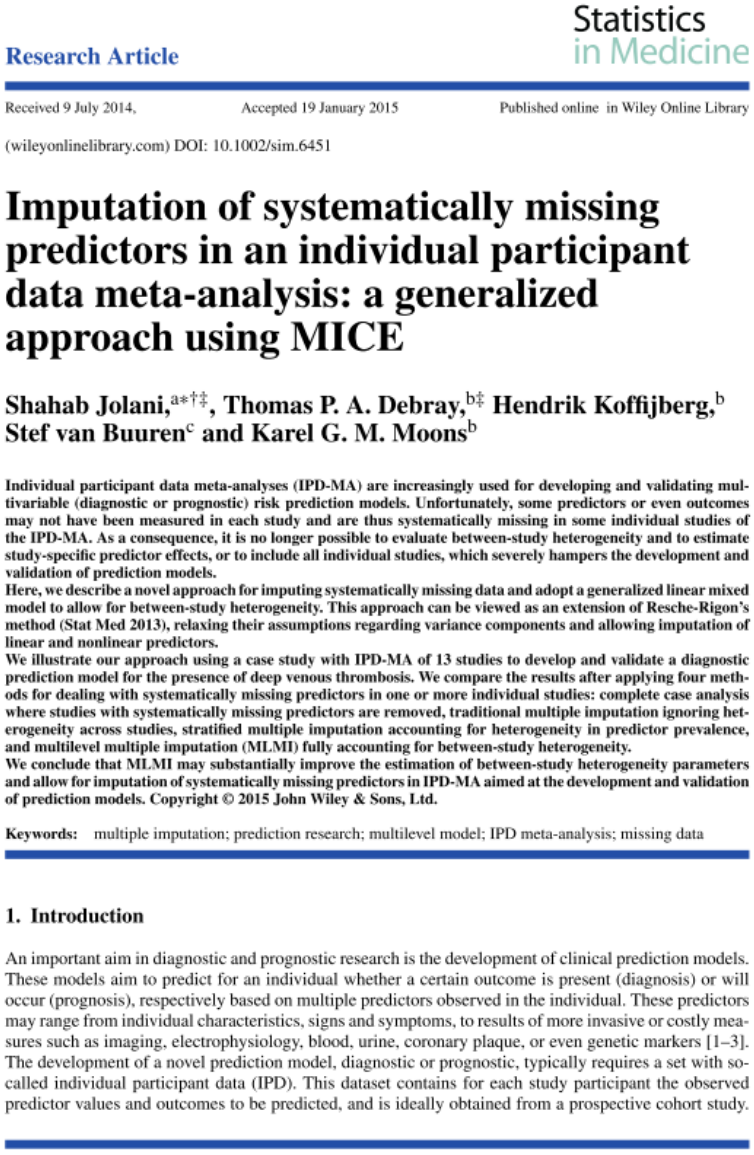
- Allow for random effects
  (modeled by multivariate normal distribution)
- Allow for between-study covariance
  (modeled by an inverse Wishart distribution)
- Implement diffuse prior distributions

- Alternative strategy: joint modeling

Dealing with missing data

Systematically missing data

- Jolani, S., Debray, T. P. A., Koffijberg, H., van Buuren, S., & Moons, K. G. M. (2015). Imputation of systematically missing predictors in an individual participant data meta-analysis: a generalized approach using MICE. *Statistics in Medicine*, *34*(11), 1841–1863. doi:10.1002/sim.6451
- Quartagno, M., & Carpenter, J. R. (2015). Multiple imputation for IPD meta-analysis: allowing for heterogeneity and studies with missing covariates. *Statistics in Medicine*, (November), n/a–n/a. doi:10.1002/sim.6837
- Resche-Rigon, M., White, I. R., Bartlett, J. W., Peters, S. A. E., Thompson, S. G., & on behalf of the PROG-IMT Study Group. (2013). Multiple imputation for handling systematically missing confounders in meta-analysis of individual participant data. *Statistics in Medicine*, *32*(28), 4890–4905. doi:10.1002/sim.5894



Research Article

Statistics in Medicine

Received 9 July 2014, Accepted 19 January 2015 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6451

Imputation of systematically missing predictors in an individual participant data meta-analysis: a generalized approach using MICE

Shahab Jolani,[a*†‡], Thomas P. A. Debray,[b‡] Hendrik Koffijberg,[b] Stef van Buuren[c] and Karel G. M. Moons[b]

Individual participant data meta-analyses (IPD-MA) are increasingly used for developing and validating multivariable (diagnostic or prognostic) risk prediction models. Unfortunately, some predictors or even outcomes may not have been measured in each study and are thus systematically missing in some individual studies of the IPD-MA. As a consequence, it is no longer possible to evaluate between-study heterogeneity and to estimate study-specific predictor effects, or to include all individual studies, which severely hampers the development and validation of prediction models.

Here, we describe a novel approach for imputing systematically missing data and adopt a generalized linear mixed model to allow for between-study heterogeneity. This approach can be viewed as an extension of Resche-Rigon's method (Stat Med 2013), relaxing their assumptions regarding variance components and allowing imputation of linear and nonlinear predictors.

We illustrate our approach using a case study with IPD-MA of 13 studies to develop and validate a diagnostic prediction model for the presence of deep venous thrombosis. We compare the results after applying four methods for dealing with systematically missing predictors in one or more individual studies: complete case analysis where studies with systematically missing predictors are removed, traditional multiple imputation ignoring heterogeneity across studies, stratified multiple imputation accounting for heterogeneity in predictor prevalence, and multilevel multiple imputation (MLMI) fully accounting for between-study heterogeneity.

We conclude that MLMI may substantially improve the estimation of between-study heterogeneity parameters and allow for imputation of systematically missing predictors in IPD-MA aimed at the development and validation of prediction models.

Keywords: multiple imputation; prediction research; multilevel model; IPD meta-analysis; missing data

1. Introduction

An important aim in diagnostic and prognostic research is the development of clinical prediction models. These models aim to predict for an individual whether a certain outcome is present (diagnosis) or will occur (prognosis), respectively based on multiple predictors observed in the individual. These predictors may range from individual characteristics, signs and symptoms, to results of more invasive or costly measures such as imaging, electrophysiology, blood, urine, coronary plaque, or even genetic markers [1–3]. The development of a novel prediction model, diagnostic or prognostic, typically requires a set with so-called individual participant data (IPD). This dataset contains for each study participant the observed predictor values and outcomes to be predicted, and is ideally obtained from a prospective cohort study.

Take home messages

Major advantages IPD-MA

- Improving the performance of novel prediction models across different study populations
- Attain a better understanding of the generalizability of a prediction model
- Exploring heterogeneity in model performance and the added value of a novel (bio)marker

Unfortunately, most researchers analyze their IPD as if representing **a single dataset**!

Take home messages
Remaining challenges in IPD meta-analysis

- IPD-MA no panacea against poorly designed primary studies
  - Prospective multi-center studies remain important
- Synthesis strategies from intervention research cannot directly be applied in prediction research (due to focus on absolute risks)
- Adjustment to local circumstances often needed
  - One model fits all?
  - Methods for tailoring still underdeveloped

New methods are on their way!