A framework for developing, implementing and evaluating clinical prediction models across multiple studies with binary outcomes

Thomas Debray

Moons KGM, Ahmed I, Koffijberg H, Riley RD

Supported by Netherlands Organisation for Scientific Research (TOP 9120.8004, 918.10.615 and 916.11.126)
and the MRC Midlands Hub for Trials Methodology Research (Medical Research Council Grant G0800808)

University Medical Center
Utrecht

Julius Center

for Health Sciences and Primary Care

Prediction modeling and IPD meta-analysis

- Opportunities
  - Increase effective sample size
  - Improve generalizability
- Challenges
  - **Heterogeneity** of IPD populations (e.g. baseline risk)
  - **Validation** of aggregated model
  - **Implementation** of aggregated model in new individuals
- Assumptions
  - Logistic regression models
  - Homogeneity of predictor-outcome associations
- Illustrative example
  - Diagnosis of Deep Venous Thrombosis
  - IPD from 12 studies ($N = 153 - 1768$)

Step 1: Estimation of predictor-outcome associations

What β terms will be used in the final model?

- Stacking

$$y_i \sim \text{Bernoulli}\left(\pi_i\right)$$
$$\text{logit}\left(\pi_i\right) = \alpha + \boldsymbol{\beta}'\mathbf{X}_i$$

- Random effects modeling of the intercept

$$\text{logit}\left(\pi_{ij}\right) = \alpha_j + \boldsymbol{\beta}'\mathbf{X}_{ij} \quad \text{with} \quad \alpha_j \sim \mathcal{N}\left(\alpha, \tau_\alpha^2\right)$$

- Stratified estimation of the intercept

$$\text{logit}\left(\pi_{ij}\right) = \sum_{m=1}^{M}\left(\alpha_m I_{m=j}\right) + \boldsymbol{\beta}'\mathbf{X}_{ij}$$

Step 2: Choosing an appropriate model intercept

What α term will be used in the final model?

- Average intercept
  - Stacking
  - Random effects
- Intercept from an included study
  - Random effects
  - Stratified estimation
  - Select intercept by similarity in outcome frequency
- New intercept
  - Estimate from outcome prevalence
    (requires mean-centering of predictor variables)
  - Estimate from new IPD

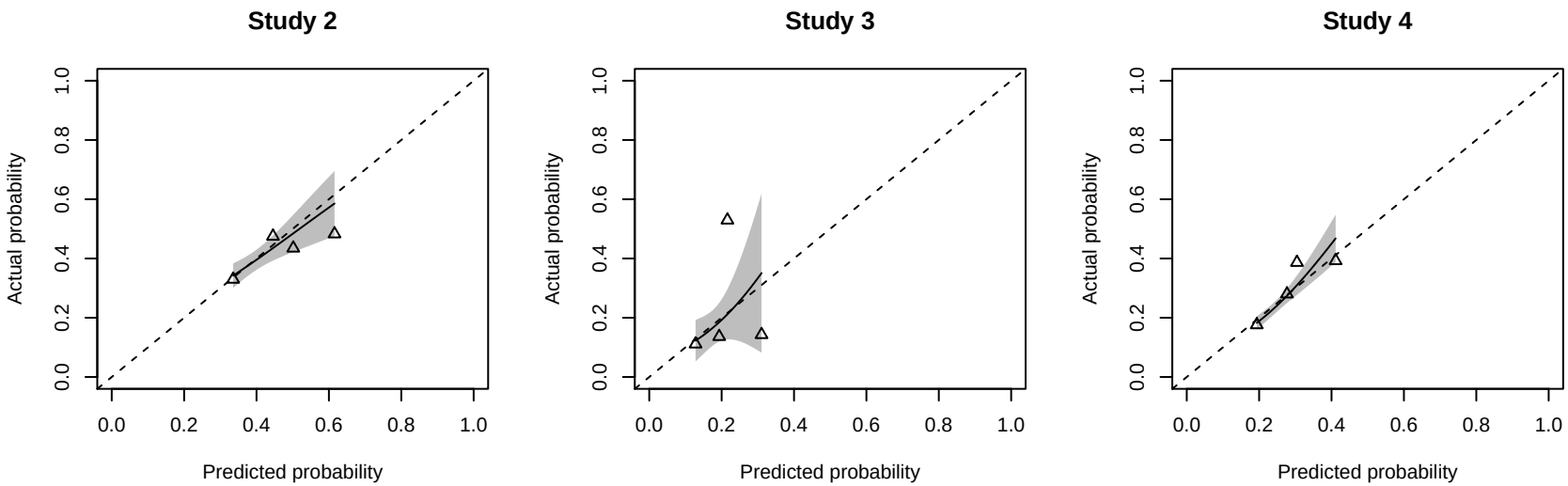
Step 3: Model evaluation

Evaluate entire strategy of **model development** and **intercept choice**

- Internal-external cross-validation (IECV, by Royston *et al.*)
- Iteratively use M-1 studies for derivation and the remaining study for validation
- Distinguish between **discrimination** and **calibration**
- Interpret model performance across M validation rotations
- Develop final model

Illustrative example: DVT (stratified estimation)

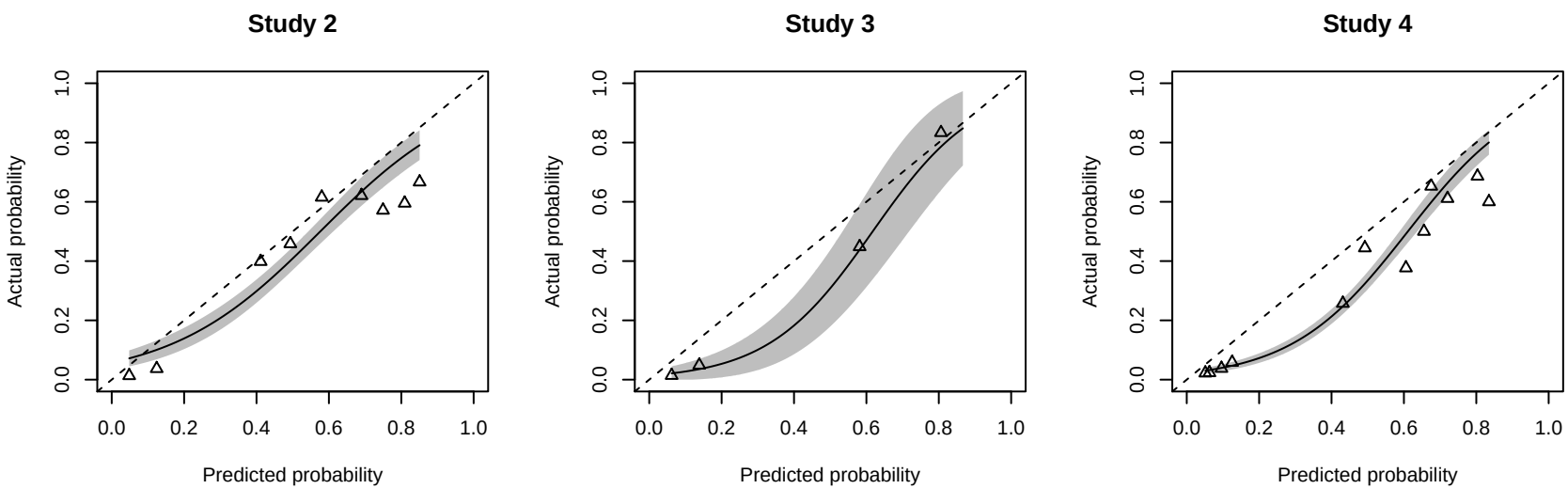
- (Nearly) homogeneous predictor-outcome associations
  - $\hat{\alpha} = -1.80$ ($\hat{\tau} = 0.47$)
  - $\hat{\beta}_{\text{sex}} = 0.47$ ($\hat{\tau} = 0.03$)
  - $\hat{\beta}_{\text{surg}} = 0.67$ ($\hat{\tau} = 0.05$)
- AUC between **0.55** and **0.65** in the IECV



Results for stratified estimation of the intercept (mean-centering of predictor variables). The intercept is estimated from the outcome frequency in the validation population.

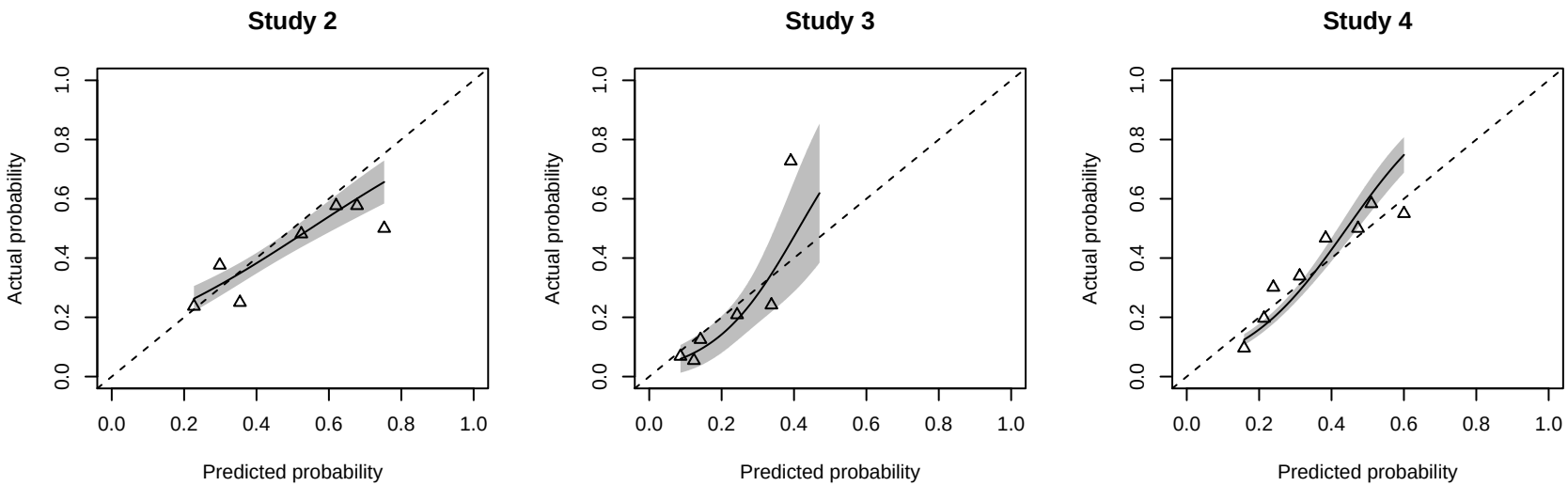
Illustrative example: DVT (stratified estimation)

- Heterogeneous predictor-outcome associations
  - $\hat{\alpha} = -3.98$ ($\hat{\tau} = 0.31$)
  - $\hat{\beta}_{\text{malign}} = 0.38$ ($\hat{\tau} = 0.35$)
  - $\hat{\beta}_{\text{calfdif3}} = 1.05$ ($\hat{\tau} = 0.16$)
  - $\hat{\beta}_{\text{surg}} = 0.25$ ($\hat{\tau} = 0.09$)
  - $\hat{\beta}_{\text{ddimdich}} = 2.76$ ($\hat{\tau} = 0.41$)
- AUC between **0.73** and **0.92** in the IECV



Illustrative example: DVT (stratified estimation)

- Weakly heterogeneous predictor-outcome associations
  - $\hat{\alpha} = -2.25$ ($\hat{\tau} = 0.47$)
  - $\hat{\beta}_{\text{sex}} = 0.37$ ($\hat{\tau} = 0.06$)
  - $\hat{\beta}_{\text{surg}} = 0.56$ ($\hat{\tau} = 0.15$)
  - $\hat{\beta}_{\text{calfdif3}} = 1.28$ ($\hat{\tau} = 0.19$)
- AUC between **0.64** and **0.76** in the IECV



Discussion

- Stratified estimation helps to improve generalizability
  - Final intercept estimated from outcome frequency
  - Final intercept selected based on outcome frequency
  - Average final intercept
  - Requires reporting of estimated intercepts!
- Internal-external cross-validation
  - Appraise model fit and its predictive ability
  - Identify heterogeneous populations
  - Ascertain the best strategy for choosing an intercept
- Avoid heterogeneity
  - Focus on (nearly) homogeneous predictor-outcome associations
  - Investigate non-linear or interaction terms
  - Discard heterogeneous studies from the meta-analysis