Aggregating published prediction models with individual participant data

TPA Debray[1], H Koffijberg[1], Y Vergouwe[12], D Nieboer[2], EW Steyerberg[2], KGM Moons[1]

[1]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands
[2]Center for Medical Decision Sciences, Erasmus Medical Center Rotterdam, The Netherlands

University Medical Center Utrecht

Introduction

The development of clinical prediction models has substantially increased during the past decades and currently facilitates numerous diagnostic and prognostic challenges. Unfortunately, investigators usually develop their own prediction model from their data, even when various similar previous models are available. Thus, previous findings are ignored. Although several procedures have recently been proposed to update an existing model with new data [1,2], it remains unclear how multiple models can be combined when they include different predictors. Here, we propose two approaches to aggregate previously derived prediction models with newly collected data.

Model Averaging

Derive a summary model $\mathrm{logit}(\overline{p}) = \beta X$ that estimates a weighted average of predicted outcomes from the original models $\mathcal{M}_m$ (where $m = 1, \ldots, M$) in the individual participant data (IPD) [3].

- $\overline{p}_i = \sum_{m=1}^{M} w_m \mathcal{M}_m(X_i)$
- $w_m = 1 / \sum_{l=1}^{M} e^{\ell_l - \ell_m}$
- $\ell_m = \sum_{i=1}^{n} \left(y_i \log\left(\mathcal{M}_m(X_i)\right) + (1 - y_i) \log\left(1 - \mathcal{M}_m(X_i)\right)\right)$

Stacked Regressions

Use the IPD of size n to create a linear predictor with the (logit) predictions from the literature models [4-6].

- Weight the predictions from the literature models
- Minimize $\sum_{i=1}^{n} \left(y_i - \sum_{m=1}^{M} \alpha_m \mathcal{M}_m(X_i) \right)^2$
- Non-negative constraints on the unknown parameters α_m
- Explicit summary model

Extension

- Maximum Likelihood Estimation (SRM)
- Common weight term when literature models predict poorly

Case Study

Deep Venous Thrombosis (DVT) is the formation of a blood clot in a deep vein, usually in a calf or thigh muscle. It may lead to organ damage and death when undetected. Here, we aggregate 5 previously published prediction models for diagnosing DVT with new data ($N = 1\,028$) to derive a single updated model. Afterwards, we validate the original and updated models in an external validation dataset ($N = 791$).

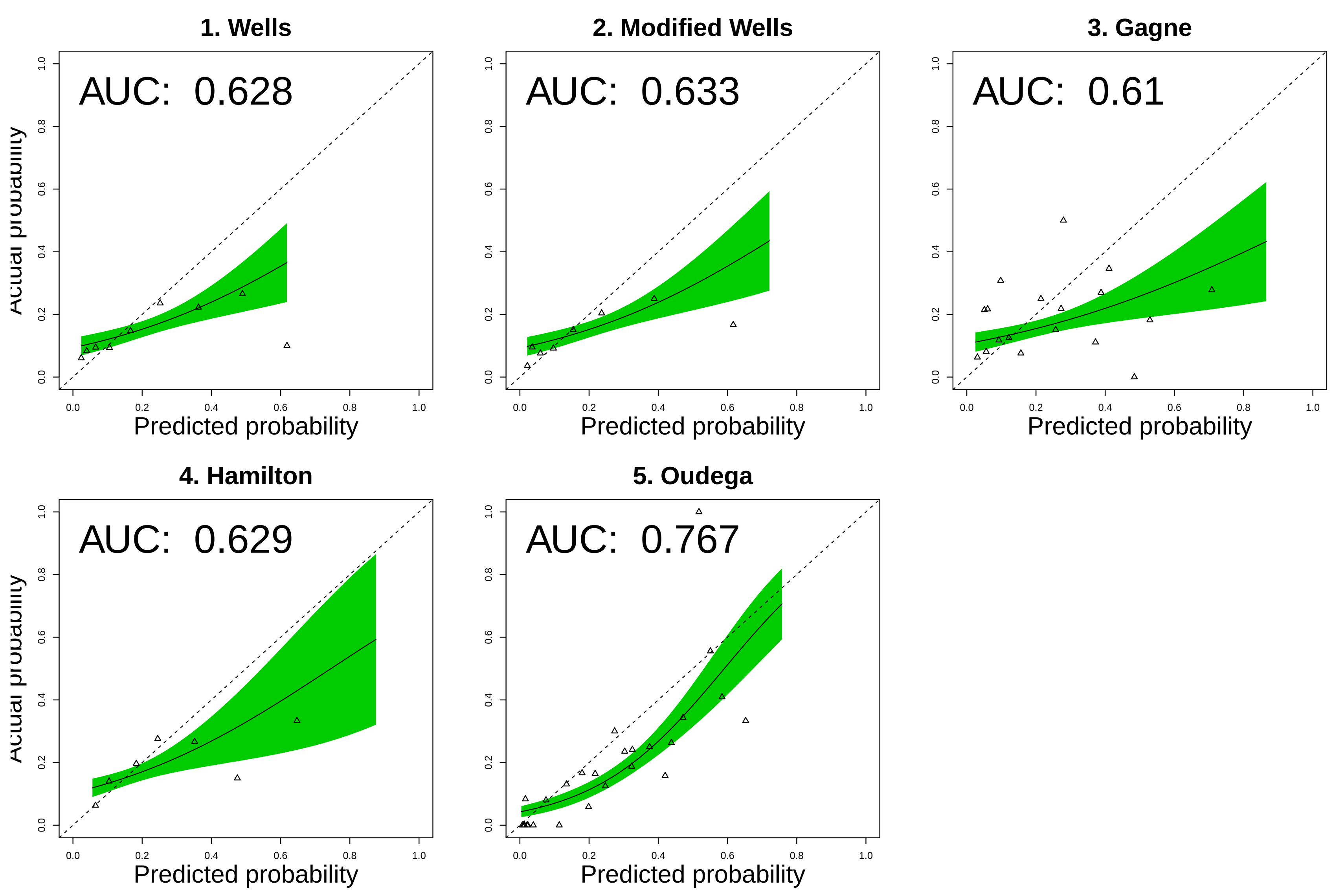

Figure: Performance of 5 previously published prediction models in an external validation dataset ($N = 791$)

Contact

Thomas Debray, MSc
Julius Center for Health Sciences and Primary Care
University Medical Center Utrecht
Tel: +31 (0) 88 75 68640
Email: t.debray@umcutrecht.nl

Case Study (cont'd)

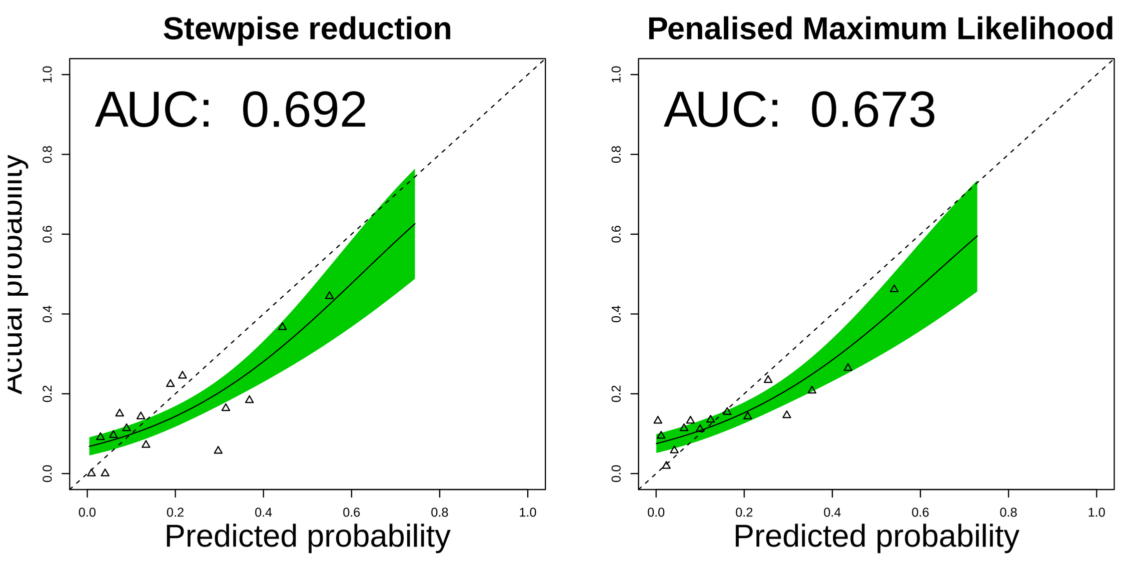

Figure: Performance of 2 novel prediction models ($N = 1\,028$) in the external validation dataset.

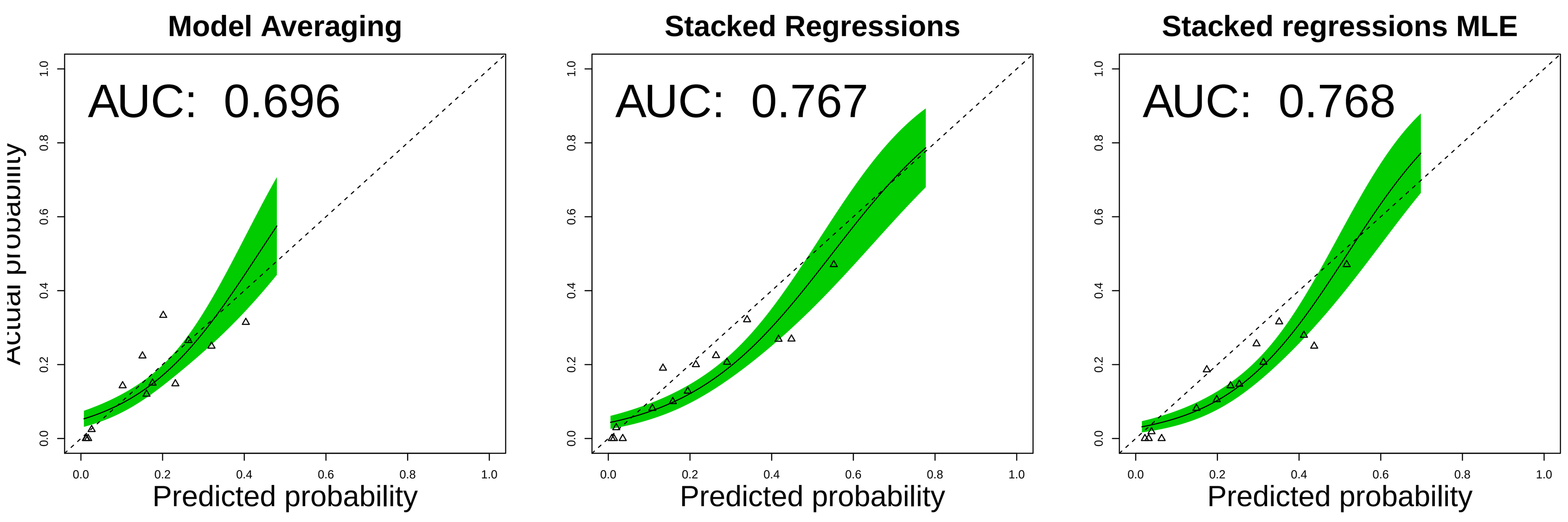

Figure: Performance of updated prediction models in the external validation dataset

- weights Model Averaging model 1-5: 0.00; 0.00; 0.00; 0.00; 1.00
- weights Stacked Regressions model 1-5: 0.00; 0.00; 0.00; 0.12; 0.90
- weights Stacked Regressions (MLE) model 1-5: 0.00; 0.00; 0.00; 0.08; 0.71 ($\alpha_0 = -0.11$)
- Note: Stacked Regression weights do not necessarily add up to 1

Discussion

- Extension of model validation and updating: multiple prediction models are validated and combined into a single updated model.
- Models that validate poorly will not contribute much (or not at all) in the aggregation process. This allows to identify useful models and combine their strengths.
- A simulation study indicates that the MLE approach of stacked regressions achieves superior calibration and discrimination in a majority of scenarios.
- Aggregation of literature models is possible with relatively few IPD.

Limitations

- Variable selection needed in updated models
- Limited accounting for heterogeneity
- Limited use of IPD at hand. When literature models perform poorly, IPD models could be included in the aggregation process

References

1. Debray TP, Koffijberg H, Vergouwe Y, Moons KG, Steyerberg EW. Aggregating published prediction models with individual participant data: a comparison of different approaches. Statistics in Medicine 2012; 10.1002/sim.5412.
2. Steyerberg EW, Eijkemans MJ, Van Houwelingen JC, Lee KL, Habbema JD. Prognostic models based on literature and individual patient data in logistic regression analysis. Statistics in Medicine 2000; 19(2), 141–160.
3. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial. Statistical Science 1999; 14(4), 382–417.
4. van der Laan MJ, Polley EC, Hubbard AE. Super learner. Statistical Applications in Genetics and Molecular Biology 2007; 6(1), Article 25.
5. Breiman L. Stacked regressions. Machine Learning 1996; 24(1), 49–64.
6. Ting KM, Witten IH. Issues in stacked generalization. Journal of Artificial Intelligence Research 1999; 10(1), 271–289.

This work was supported by Netherlands Organisation for Scientific Research (NWO) (9120.8004, 918.10.615 and 916.11.126)