Logistic regression models for patient-level prediction based on massive observational data: Do we need all data?

Luis H. John, Jan A. Kors, Jenna M. Reps, Patrick B. Ryan, Peter R. Rijnbeek

https://doi.org/10.1016/j.ijmedinf.2022.104762

Presenter: Henrik John
• Clinical Prediction Modelling: Is more data always better?

• Large datasets can put strong demands on computing resources.
• Large prediction analyses with the scale of LEGEND can have prohibitive computation times.
• Generate learning curves for 81 prediction problems, across three claims databases (CCAE, MDCD, MDCR).
• Apply threshold criterion.
RESULTS: SAMPLE SIZE
RESULTS: NO. OF PREDICTORS
CONCLUSION

• A considerable reduction in sample size and model complexity paired with a minimal loss of performance is possible.

• Paper: https://doi.org/10.1016/j.ijmedinf.2022.104762
• Contact: l.john@erasmusmc.nl