International Journal of Medical Informatics
Volume 163, July 2022, 104762

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

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

Show more

Share Cite

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

Presenter: Henrik John

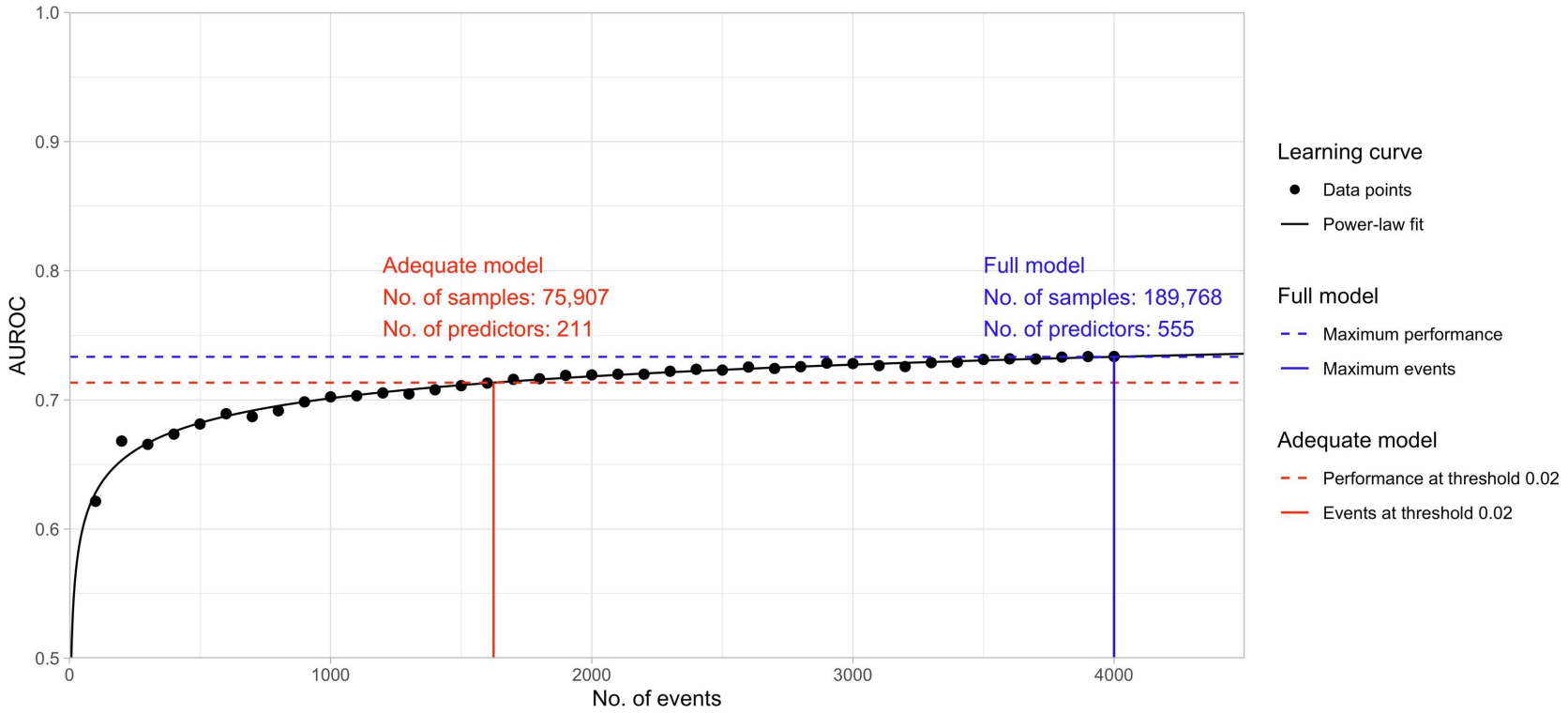
BACKGROUND

- Clinical Prediction Modelling: Is more data always better?

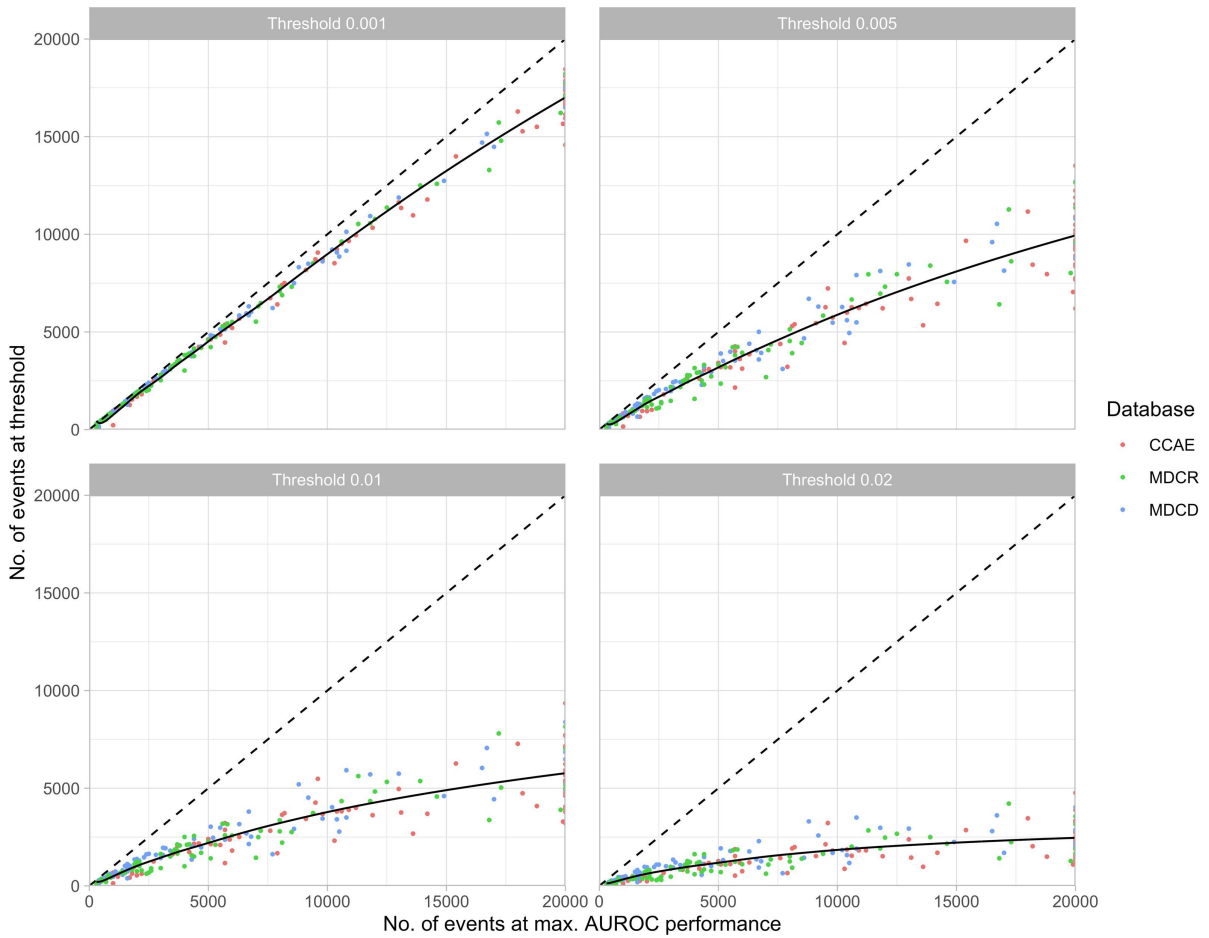
- Large datasets can put strong demands on computing resources.
- Large prediction analyses with the scale of LEGEND can have prohibitive computation times.

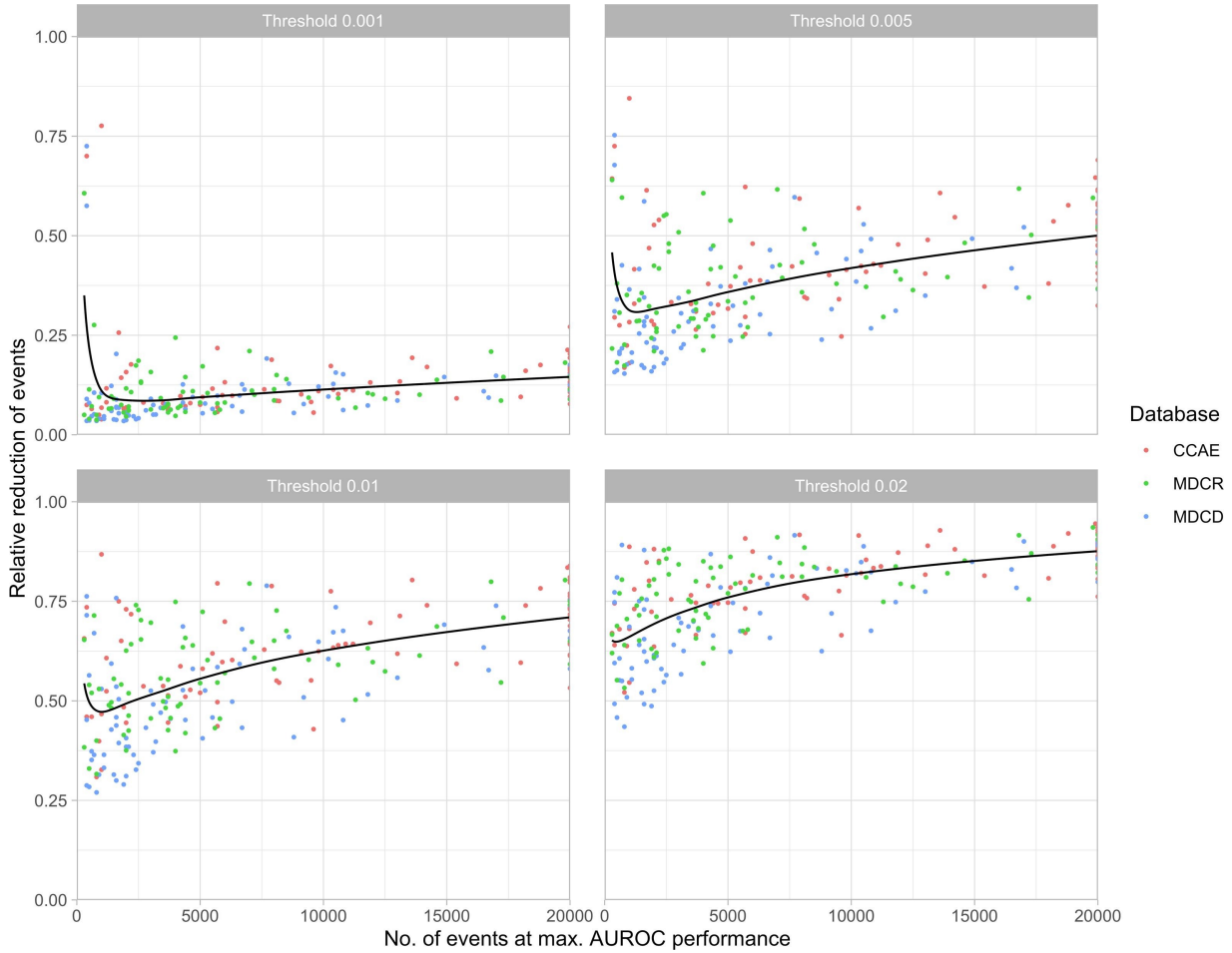
METHODS

- Generate learning curves for 81 prediction problems, across three claims databases (CCAE, MDCD, MDCR).
- Apply threshold criterion.

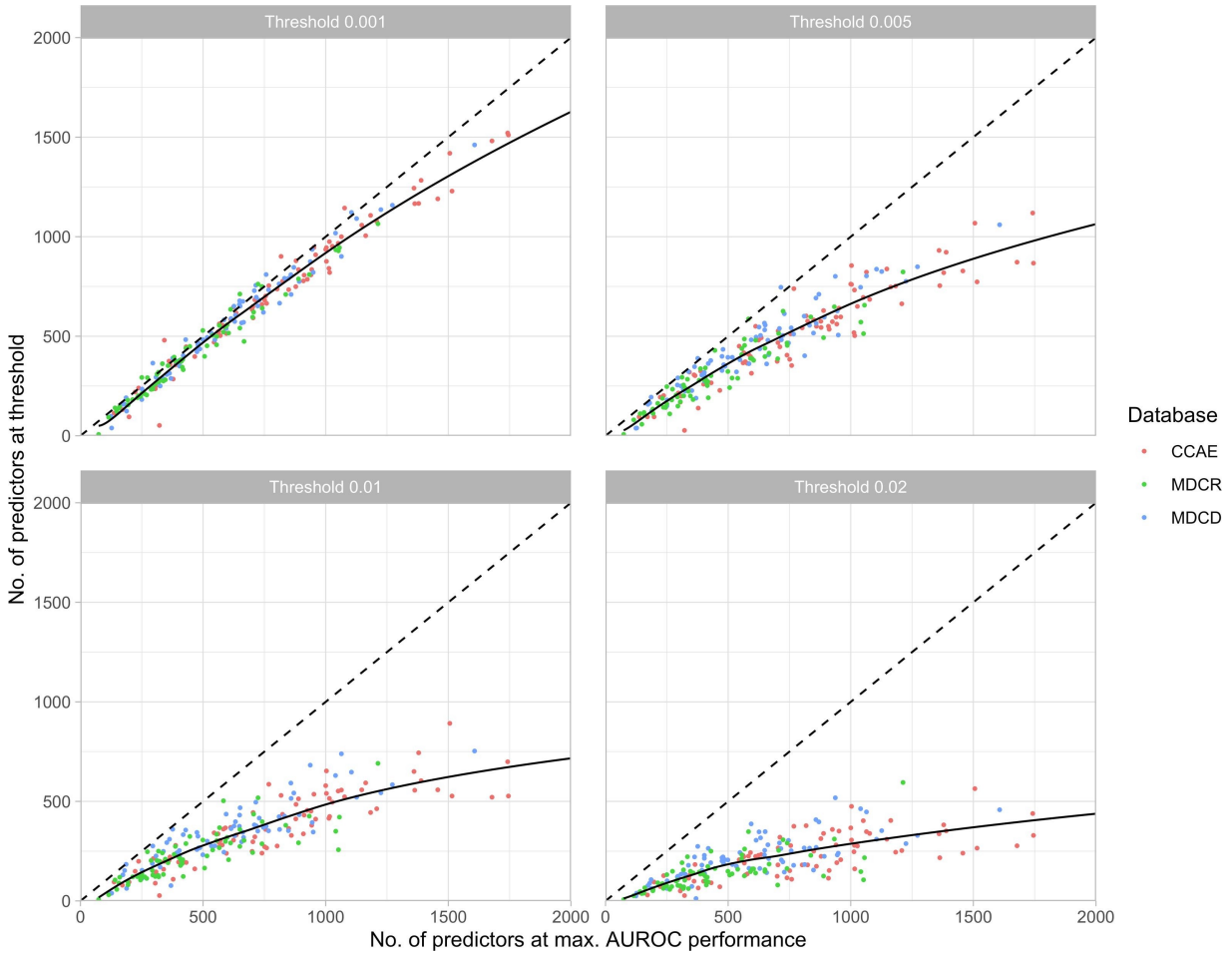


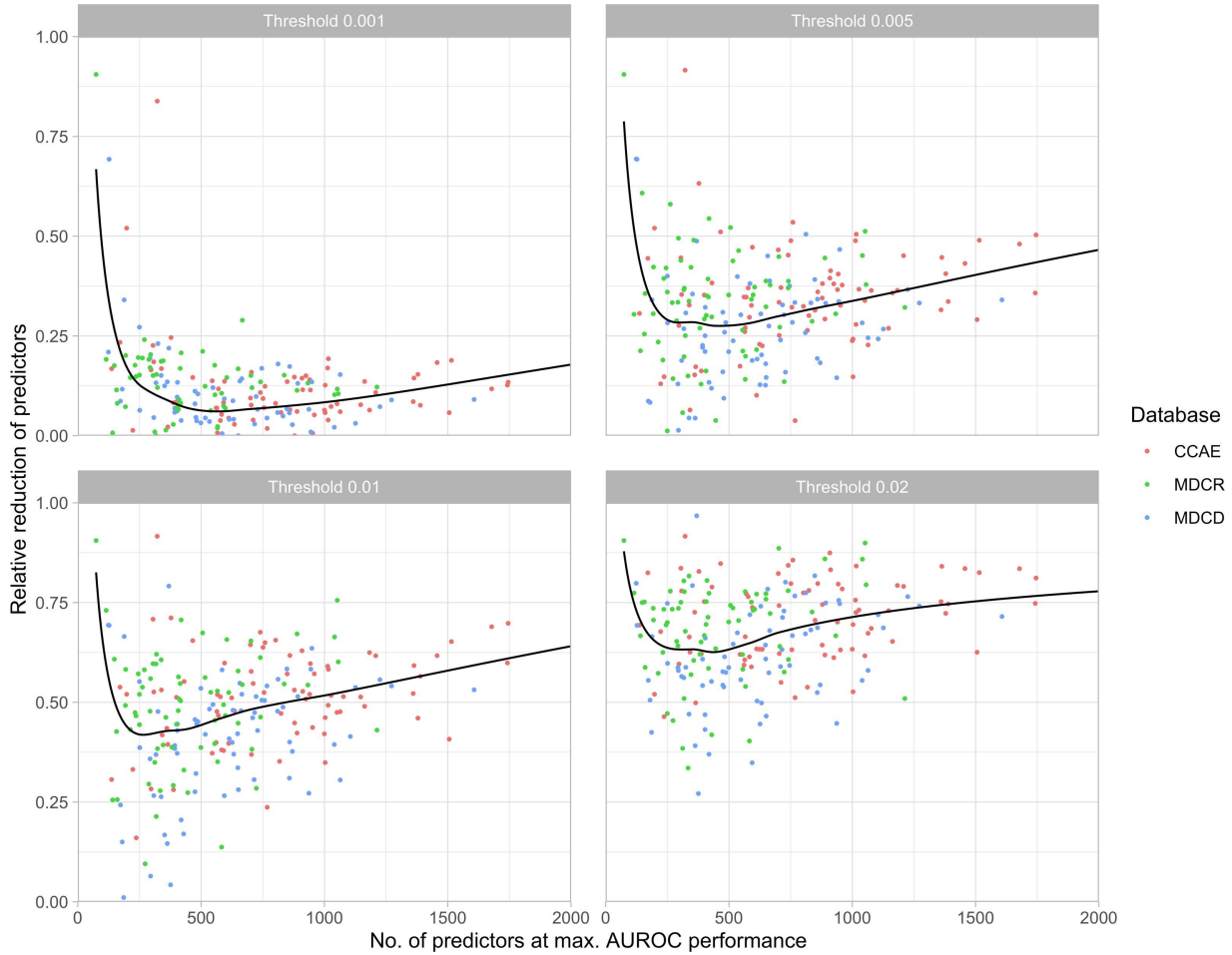
RESULTS: SAMPLE SIZE





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