DAGStat meeting 2022
Education for Statistics in Practice (March 30th 2022)

Prognosis Research in Healthcare: *initiatives to improve methodology standards*

Richard D Riley[1] & Gary S. Collins[2]

[1]Centre for Prognosis Research, School of Medicine, Keele University, Keele. UK.
r.riley@keele.ac.uk @Richard_D_Riley

[2]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford.
gary.collins@csm.ox.ac.uk @GSCollins

ABSTRACT

In healthcare, prognosis research is the study of future outcomes in individuals with a particular disease or health condition.[1] Statistical methods are fundamental to prognosis research, for example to appropriately summarise, explain and predict outcomes, and to provide reliable results that inform clinical practice and personalized healthcare decisions. However, methodology standards within prognosis research are often sub-standard, with issues including small sample sizes, overfitting, dichotomisation of continuous variables, lack of validation, and selective or incomplete reporting.[2] These problems have exacerbated with the growth of AI and machine learning methods.

Nevertheless, positive initiatives are being made to help improve prognosis research. In this talk, we will describe a number of these initiatives and encourage participants to adopt and disseminate better practice. Over two 90-minute sessions, we will cover four broad topics and illustrate the issues using real examples.

The PROGRESS Framework
The PROGnosis RESearch Strategy (PROGRESS) provides a framework of four key themes within prognosis research: overall prognosis, prognostic factors, prognostic (prediction) models, and predictors of treatment effect.[3-6] We will describe the rationale for this framework, outline the scope of each theme and why they are important, explain the limitations of current statistical practice in each, and signpost guidance for methodological improvements.

Reporting Guidelines for Prognosis Research
It is crucial for prognosis research studies to be fully and transparently reported (e.g. in terms of their rationale, design, methodology and findings), so that their findings can be critically appraised and utilised as appropriate. We will provide evidence along with examples of poor reporting in current prognosis and prediction studies (including machine learning studies), and showcase new and upcoming reporting guidelines that aim to address current shortcomings.[7 8]

Sample Size Calculations For Prognostic Model Research
Sample size calculations are rarely undertaken in prognosis model research; if they are, overly-simplistic rules of thumb are often used. In terms of sample size for model development, a well-known rule of thumb is to have at least 10 events per predictor variable, but we will describe a more principled approach based on minimising expected overfitting and ensuring precise parameter estimation.[9] In terms of sample size for model validation, a rule of thumb is to ensure at least 100 events and 100 non-events. Again, a more principled approach is possible, and we will describe how it uses the distribution of the model's linear predictor, and targets precise estimation of key model performance measures (calibration, discrimination and clinical utility).[10 11]

Individual Participant Data Meta-Analysis for Prognosis Research

One way to increase sample size is to undertake an IPD meta-analysis project, where the participant-level data from multiple existing studies are obtained, checked, harmonised, and synthesized. We will describe what an IPD meta-analysis project entails, and give examples for how it has improved prognosis research. In particular, we demonstrate how it enables non-linear prognostic relationships to be modelled; allows the development and validation of prognostic models across multiple settings and populations; and allow a more powerful and appropriate assessment of predictors of treatment effect.[12 13]

Reference List

1. Riley RD, van der Windt D, Croft P, et al., editors. *Prognosis Research in Healthcare: Concepts, Methods and Impact*. Oxford, UK: Oxford University Press, 2019.
2. Hemingway H, Riley RD, Altman DG. Ten steps towards improving prognosis research. *BMJ* 2009;339:b4184.
3. Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ* 2013;346:e5595.
4. Riley RD, Hayden JA, Steyerberg EW, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med* 2013;10(2):e1001380.
5. Steyerberg EW, Moons KG, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med* 2013;10(2):e1001381.
6. Hingorani AD, Windt DA, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ* 2013;346:e5793.
7. McShane LM, Altman DG, Sauerbrei W, et al. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer* 2005;93(4):387-91.
8. Collins GS, Reitsma JB, Altman DG, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med* 2015;162:55-63.
9. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;368:m441.
10. Riley RD, Collins GS, Ensor J, et al. Minimum sample size calculations for external validation of a clinical prediction model with a time-to-event outcome. *Stat Med* 2021
11. Riley RD, Debray TPA, Collins GS, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Stat Med* 2021;40(19):4230-51.
12. Riley RD, Debray TPA, Fisher D, et al. Individual participant data meta-analysis to examine interactions between treatment effect and participant-level covariates: Statistical recommendations for conduct and planning. *Stat Med* 2020;39(15):2115-37.
13. Riley RD, Ensor J, Snell KI, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ* 2016;353:i3140.