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

Richard D Riley¹ & Gary S. Collins²

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

²Centre 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.³⁶ 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.⁷⁸

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).¹⁰¹¹
**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.\(^\text{12,13}\)

**Reference List**