Harnessing repeated measurements of predictor variables: A review of existing methods for clinical risk prediction

Lucy M. Bull[1,2], Mark Lunt[1], Glen P. Martin[3], Kimme Hyrich[1,4], Jamie C. Sergeant[1,2]

[1]Centre for Epidemiology Versus Arthritis, Centre for Musculoskeletal Research, Manchester Academic Health Science Centre, University of Manchester, Oxford Road, Manchester M13 9PL, UK
[2]Centre for Biostatistics, Manchester Academic Health Science Centre, University of Manchester, Oxford Road, Manchester M13 9PL, UK
[3]Division of Informatics, Imaging and Data Science, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Oxford Road, Manchester M13 9PL, UK
[4]National Institute for Health Research Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Oxford Road, Manchester M13 9PL, UK

Background: Clinical prediction models (CPMs) can predict the risk of health outcomes, such as disease onset or progression, for individual patients. The majority of existing CPMs only harness cross-sectional patient information. Incorporating repeated measurements into CPMs may provide an opportunity to enhance their performance.

Aims: To systematically review the literature to understand and summarise existing approaches for harnessing repeated measurements in the development of CPMs, and empirically investigate the suitability of identified methods to real-world data using an illustrative example in rheumatoid arthritis (RA).

Methods: Medline, Embase, and Web of Science were searched for articles reporting the development of a multivariable CPM for patient-level prediction, and modelling repeated measurements of at least one predictor. Information was extracted on: the method, its specific aim, reported advantages and limitations, and software available to apply the method. For the illustrative example, CPMs were developed to predict serious infections for RA patients starting anti-TNF therapy. Preliminary analyses include a comparison of compatible cross-sectional and longitudinal CPMs to predict a patient's 12-month risk of serious infection at various time points during follow-up.

Results: The database search revealed 217 relevant articles. Seven methodological frameworks were identified: time-dependent covariate modelling, generalised estimating equations, landmark analysis, two-stage modelling, joint-modelling, trajectory classification and machine learning. Each of these frameworks satisfies at least one of three aims: to better specify predictor-outcome relationship over time; to infer a covariate value at a pre-specified time, and to account for the effect of covariate change. Identified features in available RA observational cohort data motivated the comparison of six applicable methods.

Conclusion: The applicability of identified methods depends on the motivation for including longitudinal information and the method's compatibility with the clinical context and available patient data, for both model development and risk estimation in practice.