

# merlin: A framework and implementation(s) for extended mixed effects regression of linear, non-linear and user-defined models

Michael J Crowther

Associate Professor of Biostatistics Biostatistics

Research Group

Department of Health Sciences

University of Leicester

Leicester, UK

and

Department of Medical Epidemiology and Biostatistics

Karolinska Institutet

Stockholm, Sweden

michael.crowther@le.ac.uk

## Abstract

The rise in availability of electronic health record data raises both challenges and opportunities for new and complex analyses. In this work, I will introduce an extended framework for the analysis of many types of data (continuous, binary, count, survival,), encompassing any number of outcomes of any type, each of which could be repeatedly measured (longitudinal), with any number of levels, and with any number of random effects at each level. Many standard distributions are described, as well as non-standard user-defined non-linear models. The extension focuses on a complex linear predictor for each outcome model, allowing sharing and linking between sub-models in an extremely flexible way, either by linking random effects directly, or the expected value of one outcome (or function of it) within the linear predictor of another. Non-linear and time-dependent effects are also seamlessly incorporated to the linear predictor through the use of splines or fractional polynomials, or anything else you may wish. I further allow level-specific random effect distributions and numerical integration techniques to improve usability, relaxing the normally distributed random effects assumption to allow multivariate t-distributed random effects. I'll consider some special cases of the general framework, describing some new models in the fields of survival analysis and joint longitudinal-survival models, illustrated with some examples in cancer research, and discuss various further potential uses of the implementation. User friendly, and easily extendable, soft-

ware is provided by way of the `-merlin-` package, in both Stata and R, which I will illustrate throughout the talk. Utility functions will be shown to allow the user to extend to non-standard distributions in an extremely simple way, providing a framework for extensive methods development, and the ability to fit previously proposed models very simply, which can be particularly important when authors haven't provided usable software packages.