

**Bayesian survival analysis for prognostic index development, with many covariates and missing data**

Malcolm Farrow<sup>1</sup> and Xiaohui Zhao<sup>1</sup>

<sup>1</sup> School of Mathematics & Statistics, Newcastle University, UK

We consider the Bayesian analysis of a large survival data set with more than 100 explanatory variables and 2025 patients. The data refer to patients with diffuse large B-cell lymphoma and form part of a much larger data set on lymphoma patients collected by the Scotland and Newcastle Lymphoma Group. The aim of the analysis is to produce a prognostic system offering advantages over existing prognostic indices. The system is intended for use in healthcare and also by the pharmaceutical industry in clinical trial design. It will make possible the use of more variables, and a more developed model, than existing indices, and thereby, it is hoped, will give improved prognostic precision, but will also allow computation of prognoses when only some of these variables are observed.

Missing values occur in many of the variables in the data set and omitting cases with missing values would seriously reduce the number of cases available and might distort our inference. We need to consider how we model the dependence of survival time on so many covariates and, in particular, how we construct a missing data model for such a large and diverse set of variables, both for the initial analysis and for the use of the system with new patients when only some covariates are observed.

We compare different approaches, which involve factorising the joint probability density of survival time and the covariates in different ways and the introduction of some latent variables.