2018 YCCSA SUMMER SCHOLARSHIP PROJECT SUBMISSION

This form is for prospective project supervisors to submit their projects to be included in the YCCSA Summer Scholarships Programme for 2018.

It is the purpose of the YCCSA Summer School that any projects submitted are novel and interdisciplinary in nature.

Date	
Supervisors' Names and Departments / Affiliation and Contact Email	*Ana Duarte, Centre for Health Economics, ana.duarte@york.ac.uk* *James Lomas, Centre for Health Economics, james.lomas@york.ac.uk* *Simon Hickinbotham, Department of Archaeology, simon.hickinbotham@york.ac.uk*
Project Title	*Modelling tail probabilities with non-normally distributed biomarkers – Development of an R software package*
Project Description	*This project aims to improve Health Technology Assessments (HTAs) via the provision of a new open source software package to model risk.* *One of the applications of these risk models is to characterise biomarker distributions and are often based on individual patient data (IPD), which are used to calculate the probabilities of biomarker values beyond a clinical diagnostic threshold based on observed proportions in the data. The risk models must consider the effects of highly unlikely events, which is a problem because data on low probability events is inevitably in short supply. The problem is compounded by high variation in the characteristics of these events. Recent work by Jones and colleagues[1,2] uses flexible parametric models to estimate the full conditional distribution, capturing the non-normal characteristics of healthcare costs and enabling the derivation of tail probabilities for particular populations. A similar framework can be applied to estimate the full conditional distribution of biomarker levels in the human body, as recent research demonstrates.[3] Overall, this approach improves generalisability of conclusions, allows the construction of risk models, and disengages decision models from direct calculations from the IPD database.* *Despite the advantages of using these methods, they are not commonly used in economic evaluation. One of the barriers to the wider use of the methods is that model selection can be an onerous and technically demanding process, which requires an extensive knowledge of econometrics. A software package that could automatise model testing to facilitate model selection would facilitate the use of these methods to the wider HTA community.* *The software package will be developed in R, a free software environment for statistical computing and graphics. The use of R in the HTA community is widespread because it is free, cross-platform, and open source, making it an extremely flexible tool and allowing for methodological developments to be readily translated into easy to use applications. Importantly, the open source feature is particularly valuable in a collaborative environment, as it allows inspection and modification of the source code from anyone. This*

	also increases transparency and reproducibility of the programming aspects of research, which is of particular importance to HTA agencies, as well as to improve interdisciplinary communication with scientific communities.[4] *For the first step of the project, the student will replicate in R the code developed in Stata for previous work conducted to model distributions of glucose levels in pregnant women so as to identify those who would benefit from hypoglycemiant treatment. This previous work tested flexible fully parametric maximum-likelihood models to estimate the distributions of continuous diagnostic markers (fasting and pre-load glucose levels after the administration of a diagnostic test),*[3] *and was an application of methods initially developed to model and predict healthcare costs.*[1,2] *The second step will be to develop a R software package that allows testing of the full range of models used in the previous step, drawing on existing packages such as flexsurv*[5] *and select the best fitting model automatically using an IPD sample. The software should output the parameters that will allow specifying the best fitting model in an external decision making model to predict costs and outcomes. Finally, the software package will be refined to allow working with a wide range of biomarkers and corresponding diagnostic tests, allowing for multiple measures and test cut-offs.*
Required Skills	*The student should have some experience programming and an interest on mathematical/statistical modelling.*
Supervision and Collaboration Arrangements	*All supervisors will be actively involved in co-supervision. A Duarte and J Lomas will provide input on economic modelling and econometrics. S Hickinbotham will provide input on programming.*
Project Dates	*The summer school runs for 9 weeks, starting on Monday, 09 July 2018 and finishing on Friday, 07 September 2018.*
Other Information	
References	1. *Jones AM, Lomas J, Moore P, Rice N. A quasi-Monte Carlo comparison of developments in parametric and semi-parametric regression methods for heavy-tailed and non-normal data: with an application to healthcare costs. Journal of the Royal Statistical Society, Statistics in Society, Series A. 2016;179:951–974. doi:10.1111/rssa.12141* 2. *Jones AM, Lomas J, Rice N. Healthcare cost regressions: going beyond the mean to estimate the full distribution. Health economics. 2015;24(9):1192-1212.* 3. *Nikolaidis G, Duarte A, Griffin S, Farrar D, Lomas J. 'Beyond the Mean' in Biomarkers Modelling for Economic Evaluations: a case study on Gestational Diabetes Mellitus (GDM). in submission to Pharmacoeconomics.* 4. *Lowndes JSS, Best BD, Scarborough C, et al. Our path to better science in less time using open data science tools. Nature Ecology & Evolution. 2017;1:0160.* 5. *Jackson, C. flexsurv: A Platform for Parametric Survival Modeling in R. Journal of Statistical Software. 2016;70(8):1-33. doi:10.18637/jss.v070.i08*

When complete, please email the form to sarah.christmas@york.ac.uk