EPAD Deliverable 2.3

Initial version of optimized model fitting software

Executive Summary

MRC-BSU, tasked with risk stratification in EPAD, is exploring different modelling strategies before the start of the proof-of-concept trail. Currently the modelling effort is focussed on latent-class mixed models (LCMM). These models are fitted using the R package lcmm. The computational parts of the package are implemented entirely in Fortran 90.

We performed a thorough analysis of the performance of the original LCMM code base using a synthetic dataset supplied by MCR-BSU containing 10k individuals, thought to be an upper bound on the size required for EPAD. Estimating the parameters using the original code base takes roughly 7 hours. When consecutively evaluating different models as part of model exploration, such a long runtime is unacceptable as the wait-time for the modeller doing the exploration is too high. Several performance issues were identified in the code base and some issues have been resolved. More efficient utilization of the hardware was achieved through vectorization and parallelism. This lead to a substantial decrease in computation time from the original 7 hours to 7 minutes (a factor 60 speed-up) on the same hardware, without compromising the accuracy of the results and using modest computing infrastructure: a single 16-core machine is enough to get these gains.

This deliverable gives the performance report and the user manual, and describes the optimised version of LCMM which has been made.

For more information: info@epad.org

This work has received support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking (EPAD grant nº115736).