A parallel microsimulation package for modelling cancer screening policies.

Andreas Karlsson*, Niten Olofsson†, Erwin Laure‡ and Mark Clements*

*Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
†Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden & Combitech AB
‡Center for High Performance Computing, Royal Institute of Technology, Stockholm, Sweden & Swedish e-Science Research Centre (SeRC)

Conducted within the e-Science for Cancer Prevention and Control (eCPC) project, funded by the Swedish e-Science Research Centre (SeRC).

Karolinska Institutet

Summary

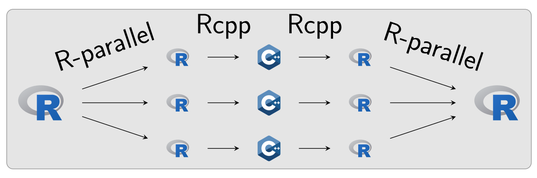
- Microsimulation with stochastic life histories is an important tool in the development of public policies.
- We provide an R package for flexible and parallel discrete-event microsimulations.
- The close coupling between R and C++ offers advantages such as ease of software dissemination, simple data management and analysis in R combined with the speed of C++.
- The performance gained by the hybrid `OpenMP/MPI` came at the cost of significant re-factoring of the existing code.
- In a case study, we showed that four-yearly prostate cancer testing would have a similar effectiveness and a marked decrease in costs compared with two-yearly testing and current testing.

Microsimulation R package

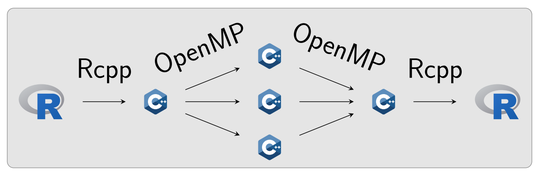
The microsimulation R package is open source and released under a GPL licence [1]. For speed, the simulation engine and model is specified in C++ , with pre- and post-processing in R, using the `Rcpp` package to seamlessly pass data structures between R and C++ [2].
The C++ core uses the `SSIM` library for discrete-event simulation [3]. The `SSIM` library is lightweight, providing classes for processes and events and static methods for the simulation.
We use common random numbers from the `RngStreams` library, where the random number stream is divided into sub-streams and each individual assigned a sub-stream [4].
Most discrete-event simulation (DES) frameworks focus on large and process-oriented simulations, rather than lightweight, event-oriented simulations which is more efficient for modelling rare events [5].

Parallel Performance

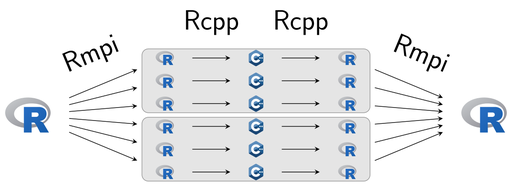
We investigated four forms of parallelism: (i) shared memory parallelism using R; (ii) shared memory parallelism at the C++ level; (iii) distributed memory parallelism using R; and (iv) a hybrid with shared/distributed memory parallelism.



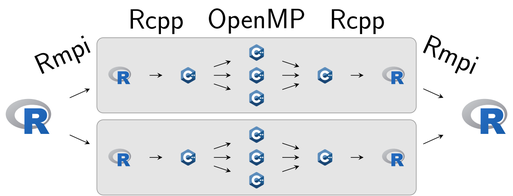
(i)



(ii)

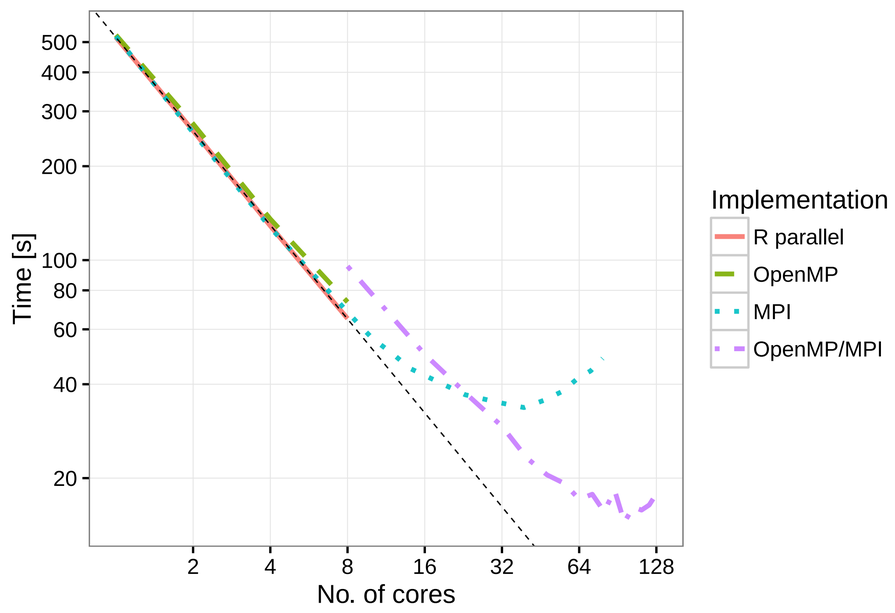


(iii)



(iv)

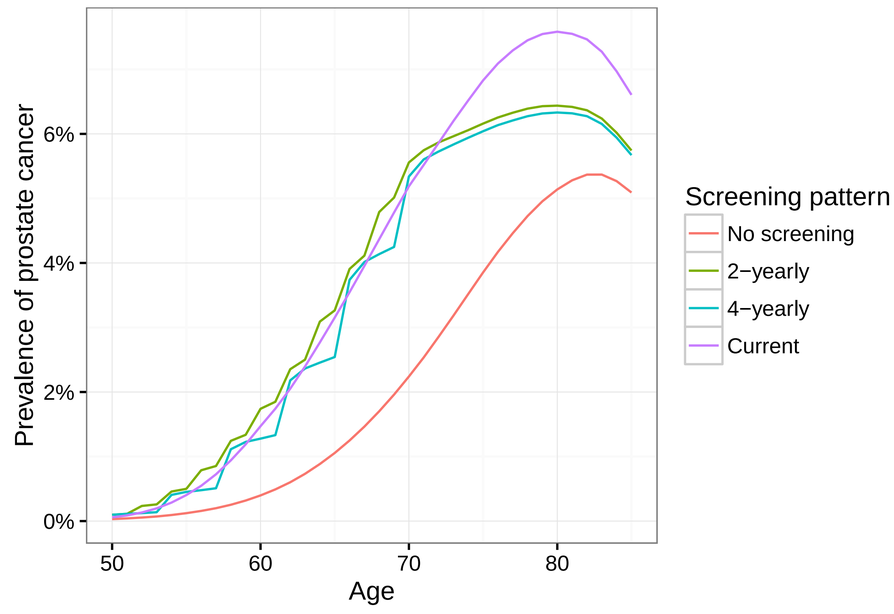
Our benchmarking of the four implementations was done on a cluster with eight cores per node using a maximum of 16 nodes. For the benchmarking, we decided on a simulation size of 10^7 individuals.



Execution time using R's `parallel` package, `OpenMP` alone, `MPI` alone and `OpenMP/MPI` as a hybrid solution. The black dashed line represents ideal scaling.

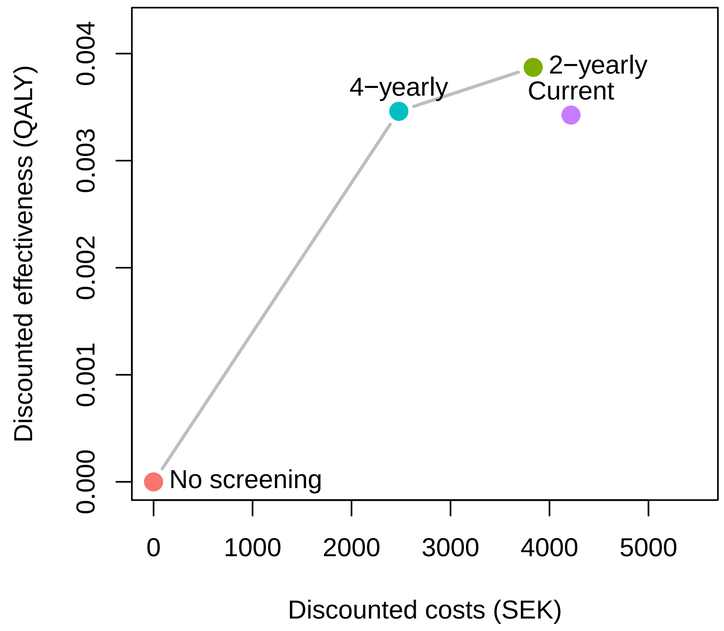
Case Study

Using the microsimulation package, we implemented a prostate cancer model based on a well validated model from the Fred Hutchinson Cancer Research Center [6]. Leveraging extensive Swedish population-based on cancer incidence, treatment modalities, mortality and prostate cancer survival we calibrated the model to the Nordic context.



Among other clinical measures we predicted the prevalence of men living with a prostate cancer diagnosis under specific re-testing protocols. Naturally, *no screening* resulted in a lower prevalence, as the asymptomatic men remain undetected, while *2-yearly* and the *4-yearly* testing resulted in an increased prevalence during the screening ages which then dropped well below the prevalence of the *current* testing pattern.

Cost-effectiveness Analysis



We also investigated the cost-effectiveness of prostate cancer testing under these re-testing protocols. For each scenario, we calculated the discounted expected costs and discounted expected effectiveness. Effectiveness was measured in terms of quality adjusted life-years (QALYs) and costs were measured in terms of Swedish kronor (1SEK $\approx$ 0.11USD).

References

M. Clements and A. Karlsson. (2016), Microsimulation package for R, [Online]. Available: `https://github.com/mclements/microsimulation`.

D. Eddelbuettel and R. Franccois, "Rcpp: Seamless R and C++ Integration," *Journal of Statistical Software*, vol. 40, no. 8, pp. 1–18, 2011. [Online]. Available: `http://www.jstatsoft.org/v40/i08/`.

A. Carzaniga and M. Rutherford. (2015), SSim - A Simple Discrete-Event Simulation Library, [Online]. Available: `http://www.inf.usi.ch/carzaniga/ssim/`.

P. L'Ecuyer, R. Simard, E. J. Chen, and W. D. Kelton, "An Object-Oriented Random-Number Package with Many Long Streams and Substreams," *Operations Research*, vol. 50, no. 6, pp. 1073–1075, 2002. [Online]. Available: `http://dx.doi.org/10.1287/opre.50.6.1073.358`.

A. Law, *Simulation Modeling and Analysis (McGraw-Hill Series in Industrial Engineering and Management)*, 5th ed. McGraw-Hill Education, 2014.

R. Gulati, J. L. Gore, and R. Etzioni, "Comparative effectiveness of alternative prostate-specific antigen–based prostate cancer screening strategies: Model estimates of potential benefits and harms," *Annals of Internal Medicine*, vol. 158, no. 3, pp. 145–153, 2013.

Contact: andreas.a.karlsson@ki.se