

Using Cloud Computing to Improve the Run-Time of Individual Patient **Simulation Models**

Joseph Whitaker¹, Alex Diamantopoulos¹, Claudia Rinciog¹

¹Symmetron Limited, London, England • Poster inquiries: crinciog@symmetron.net • www.symmetron.net • Presented at ISPOR 2022 Vienna Annual Meeting

Background and Objectives

- Patient-level simulation (PLS) approaches such as discrete event simulation (DES) typically capture important nuances of an individual patient's profile and history.
- The accuracy benefits must often be balanced against a computational trade-off in model run-time, which can be compounded in a probabilistic sensitivity analysis (PSA) of parameter uncertainty.

Objective: To explore whether retail cloud computing solutions could be used to improve the run-time of a DES model.

Methods

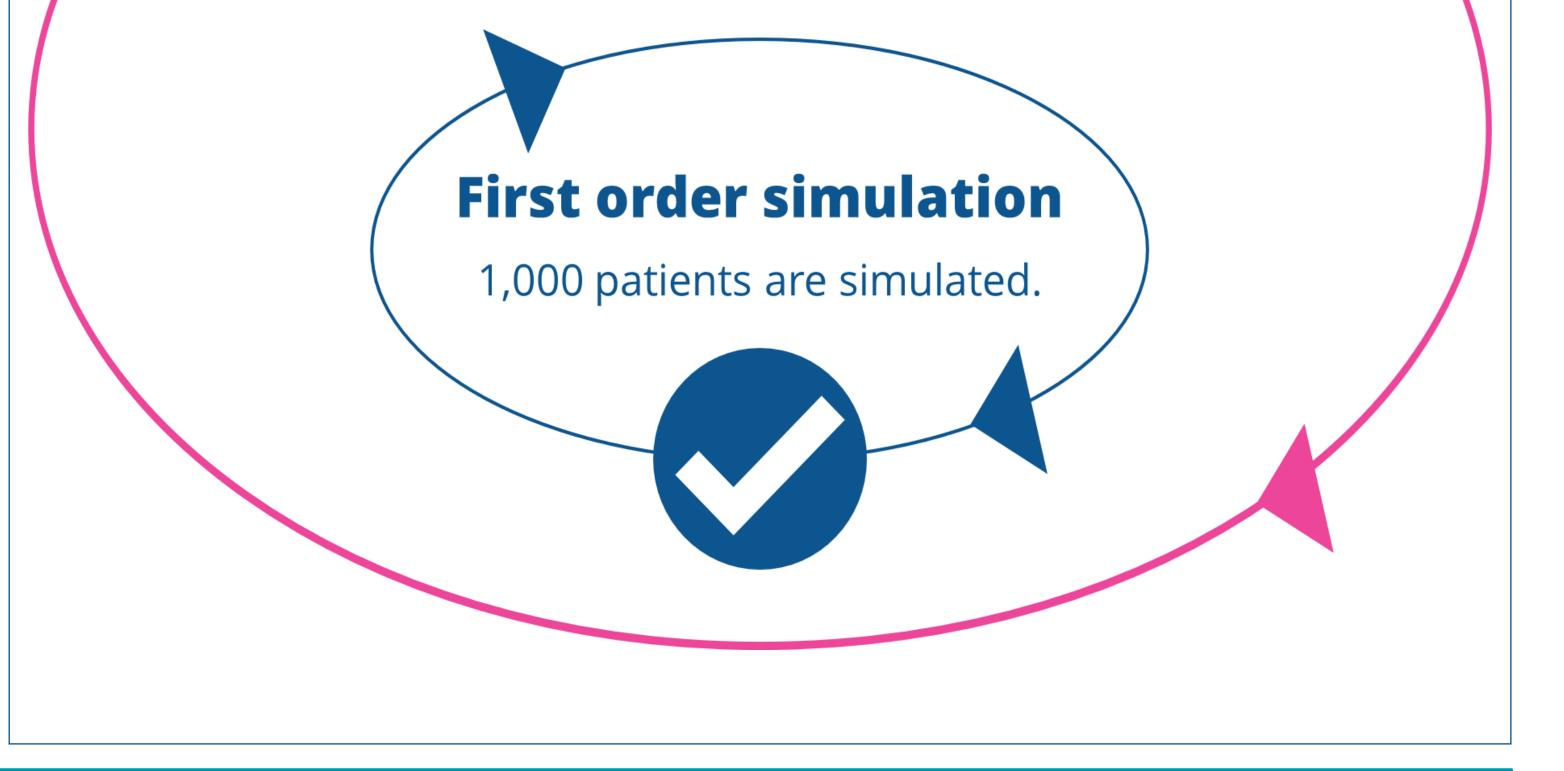
A parallelisation approach was utilised, whereby multiple central processing units (CPUs) were used simultaneously to run a cost-effectiveness analysis of stroke prevention techniques in patients with atrial fibrillation (AF).

Figure 1. The typical process of conducting probabilistic sensitivity analysis in patient-level simulation

Second order simulation

SA46

The first order simulation is repeated 1,000 times with randomly sampled model input values.



- A DES structure in R was ideal due to the complex aetiology of strokes in patients with AF and the nuanced interplay of the associated risk factors. As standard in economic evaluations for health technology assessment (HTA) submissions, parameter uncertainty was assessed using PSA.
- Analyses on a laptop utilising 1 and 6 CPUs and cloud computing analyses utilising 8, 32 and 128 CPUs were explored.
- To assess the improvement in run-time, the DES model was executed using 1,000 patients for 1,000 PSA runs, 1,000,000 patient runs in total (Figure 1).
- The time and total cost per analysis associated with each cloud computing approach were also recorded.

Results

- When using a laptop with no parallelisation, the model took 4.3 hours to complete. This was reduced to 1.6 hours (a 63% reduction in run-time) by using a typical parallelisation approach with 6 CPUs (**Table 1**).
- Using the cloud computing approach, the model runs with 8, 32 and 128 CPUs and produced run-times of 52.4, 15.7 and 5.5 minutes (reductions of 80%, 94% and 98%, respectively) (Table 1).
- The cloud computing approach allowed for up to a 20-fold increase in computational power, as depicted in Figure 2.
- The total cost associated with the cloud computing approaches ranged from United States Dollars (USD) 0.35-0.59 per analysis (Figure 3).

Table 1. Run-time reduction strategies in a DES model

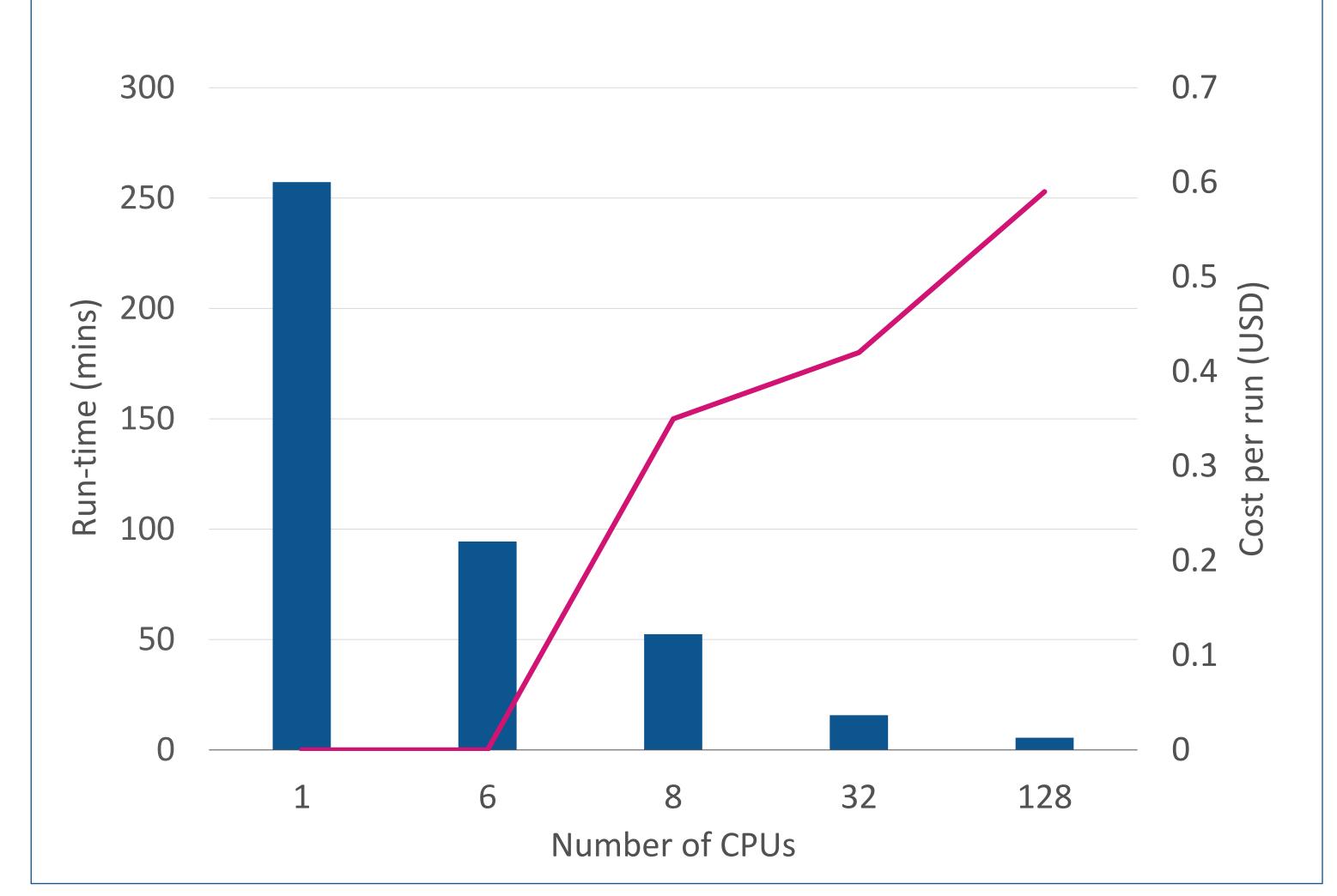
Run method	CPUs	Run time (mins)	Run time (hours)	Factor vs 1 CPU	Cost per run (USD)
Laptop	1	257.2	4.29	_	Free
Laptop	6	94.4	1.57	2.27	Free
Cloud computing	8	52.4	0.87	4.91	0.35
Cloud computing	32	15.7	0.26	16.41	0.42
Cloud computing	128	5.5	0.09	46.76	0.59

Figure 2. Computational power of run methods at maximum capacity

Laptop: 6 CPUs

Cloud-computing: 128 CPUs

Figure 3. Run-time reduction strategies and the associated costs per run



Conclusions

- In this analysis, parallelisation reduced model run-time by up to 98%, costing only USD 0.59 per run at maximum efficiency (128 CPUs).
- Considering the substantial time investment required for PLS analyses without parallelisation, USD 0.59 is comparatively negligible and demonstrates an opportunity for optimisation within the fast-paced HEOR environment.
- Since R is amenable to parallelisation it may be a useful software platform for computationally intensive PLS analyses.
- Cloud-based computing can be considered to be transferable technology that can be applied in a wide range of computationally-heavy analyses.
- The significant reduction in model run-time illustrates how cloud-based solutions can provide a considerable and inexpensive improvement of the run-time of PLS, thereby improving their practicality and usefulness when considering it as an approach for HTA modelling.

Key message: Parallelisation maximises efficiencies and promotes model agility allowing seamless collaboration between all stakeholders involved in an HTA submission for new health interventions.



Declaration of fundings: This study was funded in full by Symmetron Limited.