# EVALUATION OF THE PARETO FRONTIER APPROACH FOR MODEL CALIBRATION

Paul Peter Schneider[1], Wael Mohammed[1], Robert Smith[1]

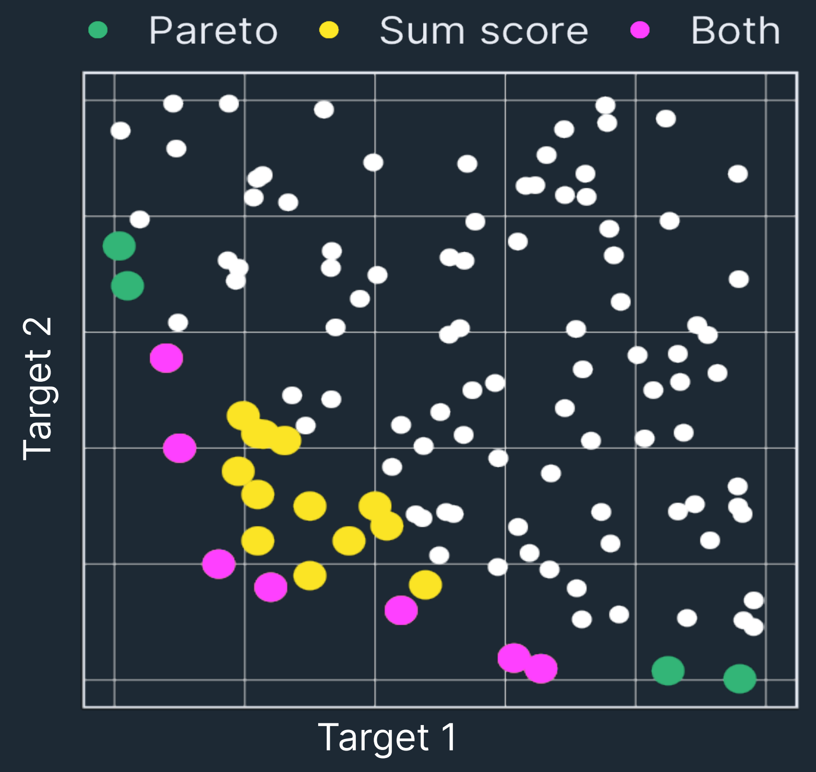
[1]ScHARR, University of Sheffield, Sheffield, UK

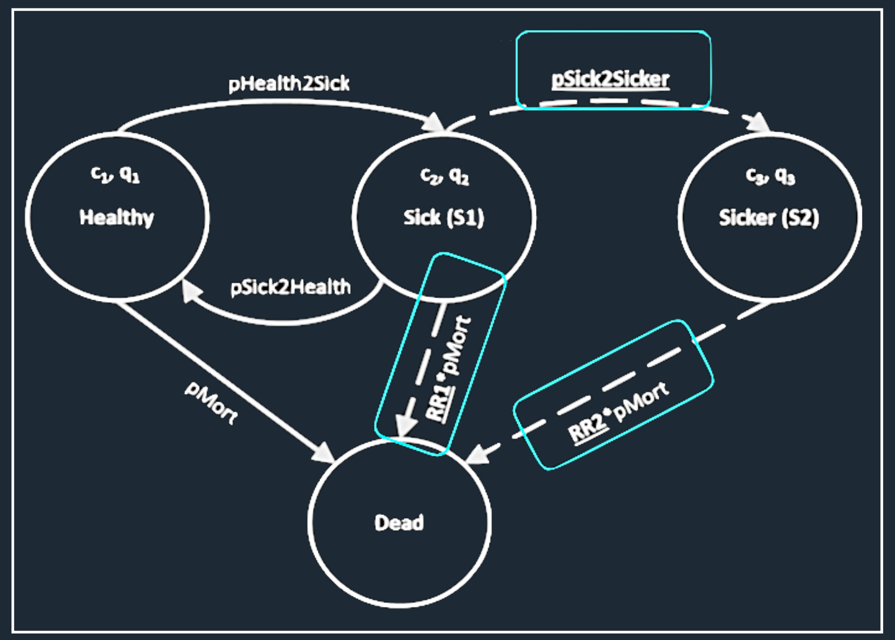
**Key reference:** Enns EA, Cipriano LE, Simons CT, Kong CY. Identifying best-fitting inputs in health-economic model calibration: a Pareto frontier approach. MDM. 2015 Feb;35(2):170-82.

OBJECTIVE: We conducted a simulation study to assess the performance of the Pareto Frontier approach against a conventional distance-based (unweighted) sum score.

## PARETO FRONTIER

- The Pareto Frontier is model a calibration method, recently proposed by Enns et al. (2015)
- A set of input parameters is on the Frontier, if you cannot improve the fit on one target without reducing it on another (see right figure)




## SICK-SICKER MODEL

- We used the same cohort state transition model that Enns et al. presented in their paper (see left figure)
- It has 3 unknown parameters that need to be calibrated
- We tested 4 target sets, consisting of 2-5 targets

## IMPLEMENTATION

The study was conducted in R v4.0. The rPref package was used to identify Pareto optimal sets. We used a 64-cores AWS instance and parallelisation to execute the >500 mio. model runs.

The source code is available at: github.com/bitowaqr/pareto_frontier

Paul Peter Schneider
University of Sheffield, UK
p.schneider@sheffield.ac.uk
@waq0r

The University Of Sheffield.

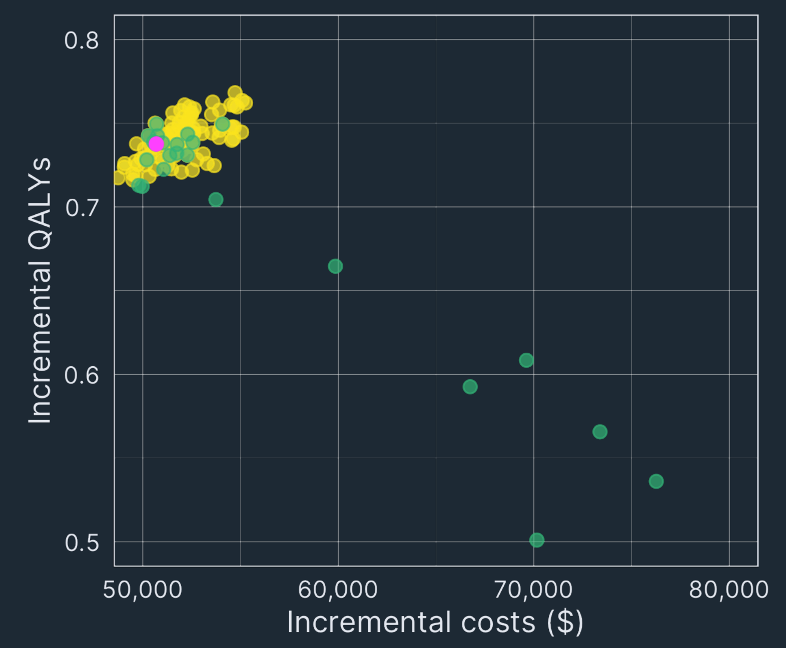
wellcome

## SIMULATION PSEUDO CODE

```
for i = 1 to 10,000  {

 1. Specify a true model:
   – Randomly draw values for all (known
     and unknown) model parameters
   – Compute the true incremental net
     monetary benefit (iNMB)

 2. Generate calibration target sets:
   – Run a micro-simulation to generate
     stochastic targets

 3. Run model calibration:
   – Generate 50,000 candidate input sets
   – For each set, compute differences
     between model outputs and targets
   – For each of the 4 target sets,
     select the inputs that:
       – lie on the Pareto Frontier
       – are among 1% with the lowest sum
         of absolute errors

 4. Evaluate calibration performance:
   – Compute the mean iNMB across
     selected input sets and compare it
     against the true iNMB
}
```

## SAMPLE RESULTS FOR $i = 1$

- The right figure shows exemplary results for one simulation run
- Here, the sum score calibration performed better than the Pareto Frontier approach: the mean absolute error in iNMB was 977 vs. 19,091.




## MAIN RESULTS


Mean (SD) absolute error in iNMB

| Target set | Pareto Frontier | Sum score |
|---|---|---|
| 1 | 11,600 (7,032) | **7,505 (6,381)** |
| 2 | 7,403 (4,151) | **5,368 (4,104)** |
| 3 | **21,095 (8,097)** | 28,185 (14,337) |
| 4 | 12,461 (5,444) | **3,905 (2,859)** |

- The sum score method provided more accurate mean iNMB predictions for 3 of 4 target sets
- Models calibrated with the Pareto Frontier approach performed better only when using Target Set 3*
- The mean (SD) number of sets on the Frontier was 601 (984)
- Identifying Pareto optimal inputs was computationally demanding

*Note: Target set 3 consisted of 3 proportions (range: 0-1) and 1 ratio (range: 0-Inf.). When target trade-offs are (mis-)specified like this, it is not surprising that a sum score performs poorly.

## CONCLUSION

1) The Pareto Frontier model calibration method generally performed worse than the simple, distance-based sum score.
2) However, when trade-offs between targets are misspecified, the Pareto Frontier may provide less biased results.