Sample size calculations using Bayesian optimisation

Duncan T. Wilson[a], Richard Hooper[b], Rebecca E. A. Walwyn[a], Sarah R. Brown[a], Julia Brown[a], Amanda J. Farrin[a]

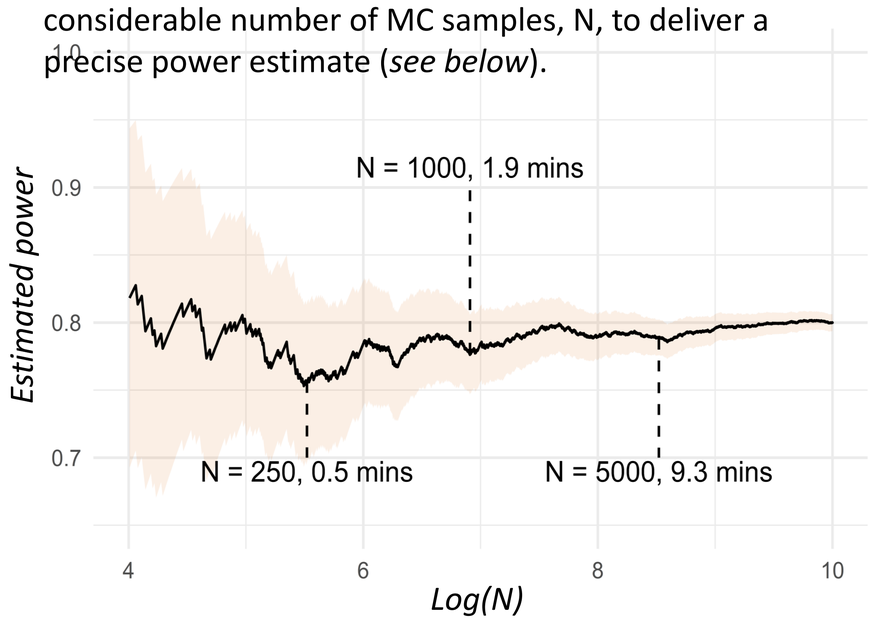
a - Clinical Trials Research Unit, Leeds Institute of Clinical Trials Research, University of Leeds
b – Pragmatic Clinical Trials Unit, Blizard Institute, Queen Mary University of London

Leeds Institute of Clinical Trials Research

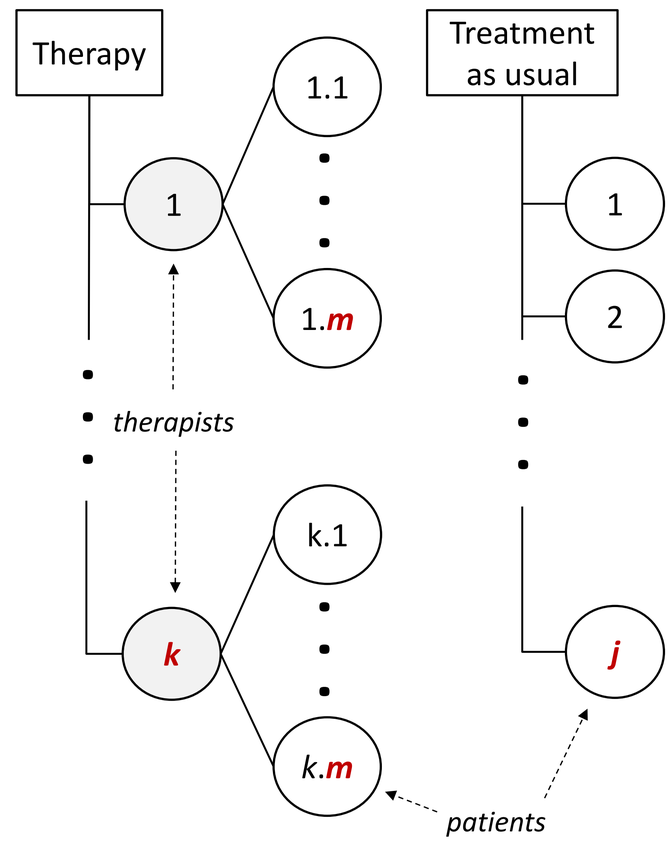
UNIVERSITY OF LEEDS

Background

- For complex trial designs and analyses, simple analytic formulae for power calculations are not always available.
- In such cases we can always fall back on **Monte Carlo estimates of power**, and use these when determining the optimal sample size **[1]**.
- This can be **computationally demanding**, requiring a considerable number of MC samples, N, to deliver a precise power estimate (*see below*).

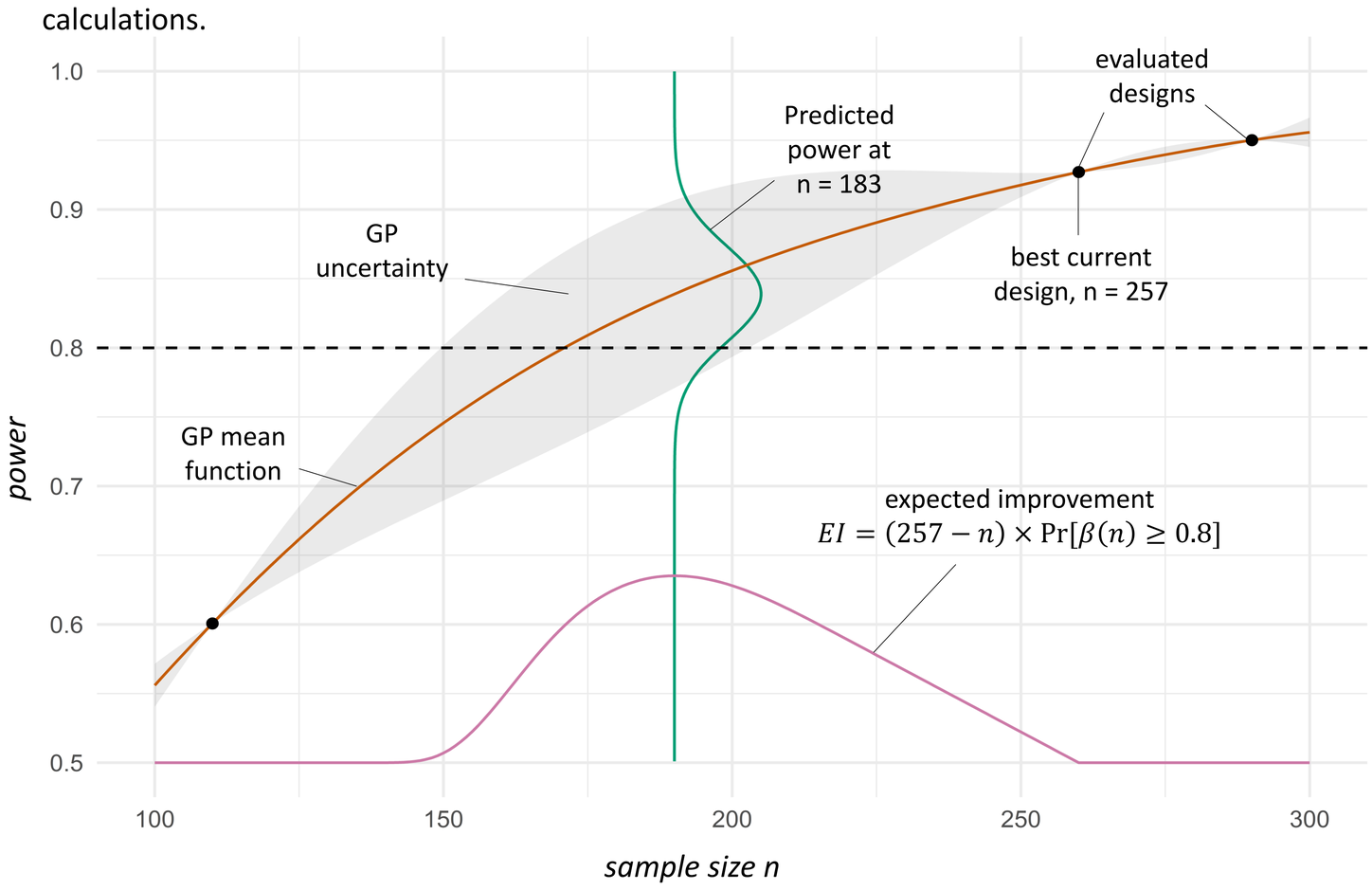


- In these complex design, we often have several sample size parameters to choose the values of, and several criteria we want to minimise.
- For example, consider a partially nested trial of a psychotherapy intervention, where there are ***k*** therapists in the intervention arm, each treating an average of ***m*** patients, and there are ***j*** patients in the control arm (*see right*).
- For $k \in \{3, 30\}$, $m \in \{3, 40\}$, $j \in \{100, 500\}$, we have **over 500,000 possible designs** to choose from.
- We want to find a set of designs which are adequately powered, and which offer different trade-offs between minimising the number of therapists and minimising the total number of patients.
- We will fit a **partially nested heteroskedastic model** for a continuous patient outcome, accounting for clustering in the intervention arm **[2]**. A likelihood ratio test will be used to test the hypothesis of no treatment effect, and so we need to use MC estimates of power.
- To solve this problem in a timely manner, we need to use **highly efficient optimisation algorithms**.



Methods

- Because estimating power takes so long, we can only do so for a small (< 200) number of designs.
- However, given some initial power estimates, we can construct a **surrogate model** f of the true power function β.
- We use a **Gaussian process** (GP) surrogate model, which is flexible and leads to tractable calculations.
- A GP model represents our belief about the power of a design through a normal distribution, giving both a point prediction (the mean) and a measure of the uncertainty in that prediction (*see below*).
- GP models are commonly used in a wide variety of fields, and several R packages for fitting GPs are available.



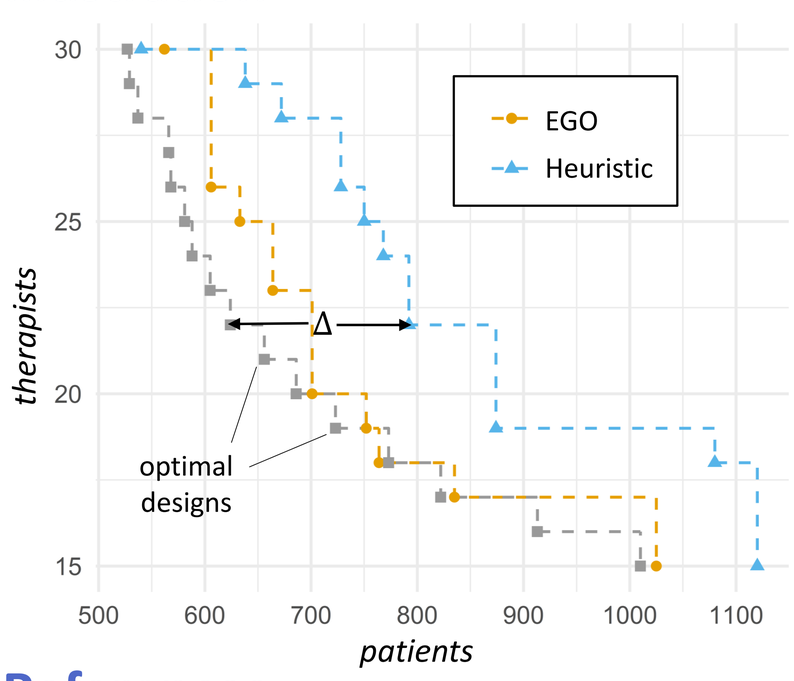
Model-assisted Efficient Global Optimisation (EGO)

- Given a GP model of the power function and the uncertain predictions it provides, we can ask questions like:
  - If I estimate the power of a new design, what is the probability that it will be sufficient?
  - Compared with the best design I have found so far, what improvement can I expect to see if I estimate the power of this new design?
- When minimising a single criteria, we can guide the search process by estimating the power of the design which gives the largest **expected improvement [3]** (*see left*).
- Our partially nested design is more complex – we want to minimise both the number of therapists *k* and the total number of patients *n*.
- At each iteration in the algorithm (*see below*) we select a random weight *w* and define the quality of a design as $wk + (1 - w)n$. We then minimise this single criteria, subject to power.
- By selecting a random weight *w* at each iteration, we find a range of designs with different trade-offs between the two criteria.

The algorithm

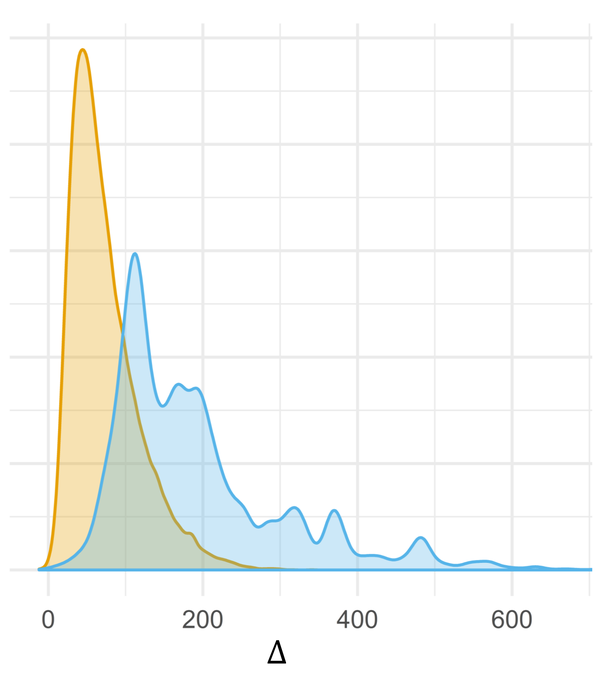
1. Choose an initial set of designs X.
2. Compute the Monte Carlo power estimate at each $x \in X$.
3. Using a random weight *w*, calculate the value y of each $x \in X$.
4. Fit a Gaussian process model $f(x)$ as a surrogate for $\beta(x)$.
5. Find the value y of the best design $x \in X$ which is almost certainly adequately powered, according to the model f.
6. Find the design $x \notin X$ with largest expected improvement over y.
7. Compute the Monte Carlo estimate of the power at x and add to X.
8. Repeat steps 3 – 7 until the computational budget is exhausted.

Illustration



We used the EGO method to determine sample size for the above partially nested psychotherapy example. For comparison, we also used a simple heuristic. The two methods are contrasted here.

- In a single application, we compared the performance of EGO and the heuristic.
- The EGO algorithm finds **more efficient designs** – for equal numbers of therapists, EGO designs can require as many as 220 fewer patients, a reduction of around 20%.
- The EGO designs are quite close to the optimal designs (*see left*).
- Over 1000 applications, we count the differences Δ between the obtained and optimal designs' *n*.
- On average, the EGO algorithm requires **120 fewer patients** than the heuristic (*see right*).
- EGO is also more likely to locate a design for each feasible k – the heuristic will miss a feasible k around 15% of the time.



References

[1] Landau, S. & Stahl, D. (2013), Sample size and power calculations for medical studies by simulation when closed form expressions are not available, *Statistical Methods in Medical Research*, 22, 324-345. **[2]** Roberts, C. & Roberts, S. A. (2005), Design and analysis of clinical trials with clustering effects due to treatment, *Clinical Trials*, 2, 152-162. **[3]** Jones, D. R. (2001), A Taxonomy of Global Optimization Methods Based on Response Surfaces, *Journal of Global Optimization*, 21, 345-383.

Acknowledgements

Duncan Wilson is funded by an MRC Skills Development Fellowship

MRC Medical Research Council