MDG: Decision-theoretic analysis of pilot trials

Duncan T. Wilson

Clinical Trials Research Unit
Leeds Institute of Clinical Trials Research
University of Leeds

d.t.wilson@leeds.ac.uk

7th March 2018

Developing methods to design and analyse pilot trials of complex interventions, considering...

- Frequentist methods (extending phase II designs, known nuisance parameters, multivariate testing);
- Bayesian methods (unconditional probabilities of hypotheses, uncertainty in parameters);
- Decision-theoretic methods (specifying a utility function, modelling the confirmatory RCT).

Previously...

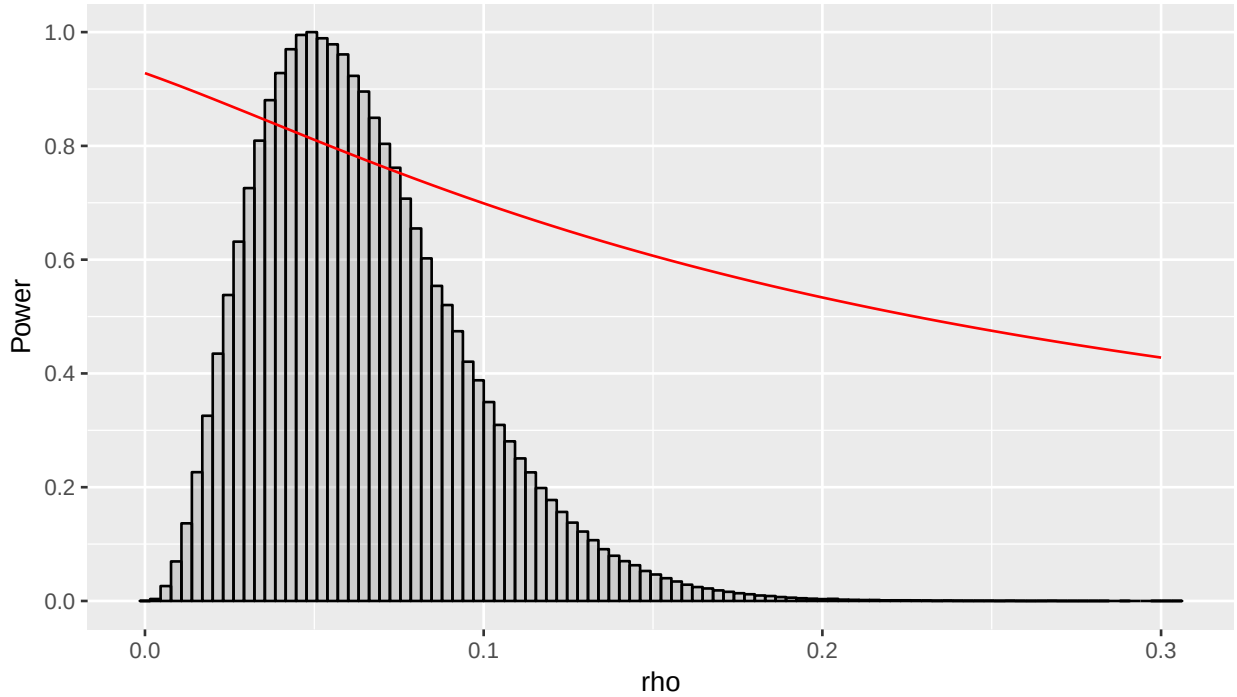
Last time I spoke at an MDG we discussed how a Bayesian approach can help us handle uncertainty in, e.g., an ICC ρ of a cluster trial:

$$n = n_i[1 + (m-1)\rho]$$

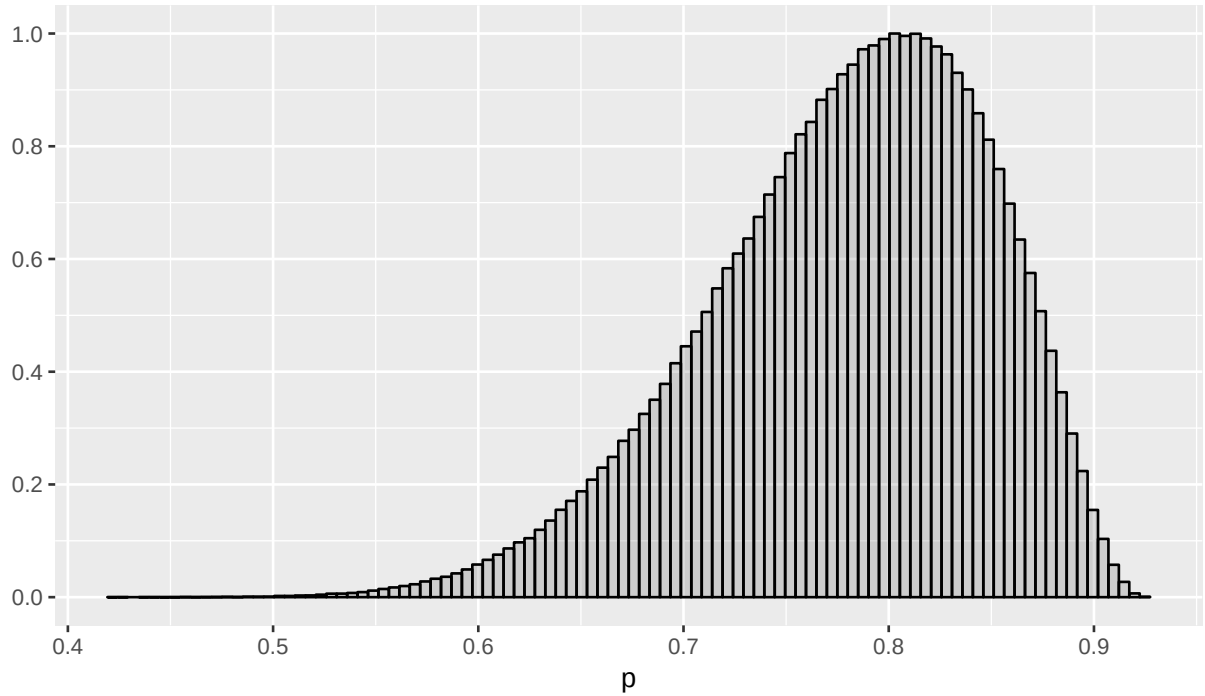
n_i = sample size required when clustering is ignored

Defining a (subjective prior) of $\rho \sim Beta(4, 58)$ (say), will give a posterior after seeing the pilot data.

Uncertainty



Uncertainty about ρ propagates to uncertainty about power for fixed n:



Focussed on methods of assurance - calculating the unconditional probability of an event. Most commonly, the rejection of the null in a freq. test, denoted A:

$$\text{assurance} = Pr[A] = \int Pr[A \mid \theta] p(\theta) d\theta$$

$$p(\theta) = \text{probability density function of } \theta$$

(Other types of events considered too, e.g. rejecting the null *and* the true effect is meaningful)

Utility

Although assurance is arguably easier to interpret than conditional power, it isn't clear how to balance it against type I error and sample size costs when designing a trial. Alternative:

- Specify a utility function $u(d, \theta)$ (d = trial design)
- We can then choose the design which maximises the expected utility:

$$\arg\max_{d} \ \mathbb{E}_{\theta}[u(d, \theta)] \tag{1}$$

- We will parametrise the design by $d = (\alpha, n)$, the type I error rate and the sample size.

Why might we want to use a decision-theoretic approach? Axioms:

1. The relation $\succeq$, where $c_1 \succeq c_2$ means c_1 is at least as good as c_2, is a weak ordering (i.e. complete and transitive) on $\mathcal{C}$, the space of consequences.
2. There exists $c^*, c_* \in \mathcal{C}$ such that $c^* \succ c_*$ and $\forall c \in \mathcal{C}, c^* \succeq c \succeq c_*$.
3. ...

Axioms imply that there exists a $u : \mathcal{C} \to \mathbb{R}$ such that $\forall c_1, c_2 \in \mathcal{C}$:

$$c_1 \succeq c_2 \Leftrightarrow u(c_1) \geq u(c_2) \tag{2}$$

Scope for today

We will consider:

- A pilot trial has been conducted;
- We have a (empirical or mathematical) posterior distribution on the unknown treatment difference θ;
- A 'confirmatory' balanced parallel group RCT is to be designed, i.e. $d = (\alpha, n)$ is to be chosen;
- The primary endpoint and corresponding statistical test have been selected;

What designs are admissible? How do we choose the optimum?

A single-attribute value function

We have two attributes to the problem - the efficacy θ (benefit), and the sample size n (cost). We assume these are *preferentially independent*:

$$(\theta, n') \succeq (\theta', n') \tag{3}$$

$$\Rightarrow (\theta, n) \succeq (\theta', n) \ \forall \ n \tag{4}$$

We then focus on defining a single-attribute *value* function $v(a, \theta)$, where a is the action taken as a result of the test - stop $(-)$ or go $(+)$.

A single-attribute value function

$$v(+, \theta) = \theta \tag{5}$$

$$v(-, \theta) = 0 \tag{6}$$

A single-attribute value function

Center at θ^*, the *minimal clinically important difference*:

$$v(+, \theta) = \theta - \theta^* \tag{7}$$

$$v(-, \theta) = 0 \tag{8}$$

A single-attribute value function

Account for the *opportunity cost* - the value of the best alternative action:

$$v(+, \theta) = (\theta - \theta^*) - 0 = \theta - \theta^* \quad (9)$$

$$v(-, \theta) = 0 - (\theta - \theta^*) = \theta^* - \theta \quad (10)$$

A single-attribute utility function

Transform from value to utility, accounting for *attitude to risk* - risk neutral just gives

$$u(+, \theta) = \theta - \theta^* \tag{11}$$

$$u(-, \theta) = \theta^* - \theta \tag{12}$$

Denote the power function by $g(d, \theta)$. Then the expected utility is:

$$\mathbb{E}_\theta[u(d,\theta)] = \int \{Pr[+|d,\theta]u(+,\theta) + Pr[-|d,\theta]u(-,\theta)\}p(\theta)d\theta \tag{13}$$

$$= \int \{g(d,\theta)(\theta - \theta^*) + [1 - g(d,\theta)](\theta^* - \theta)\}p(\theta)d\theta. \tag{14}$$

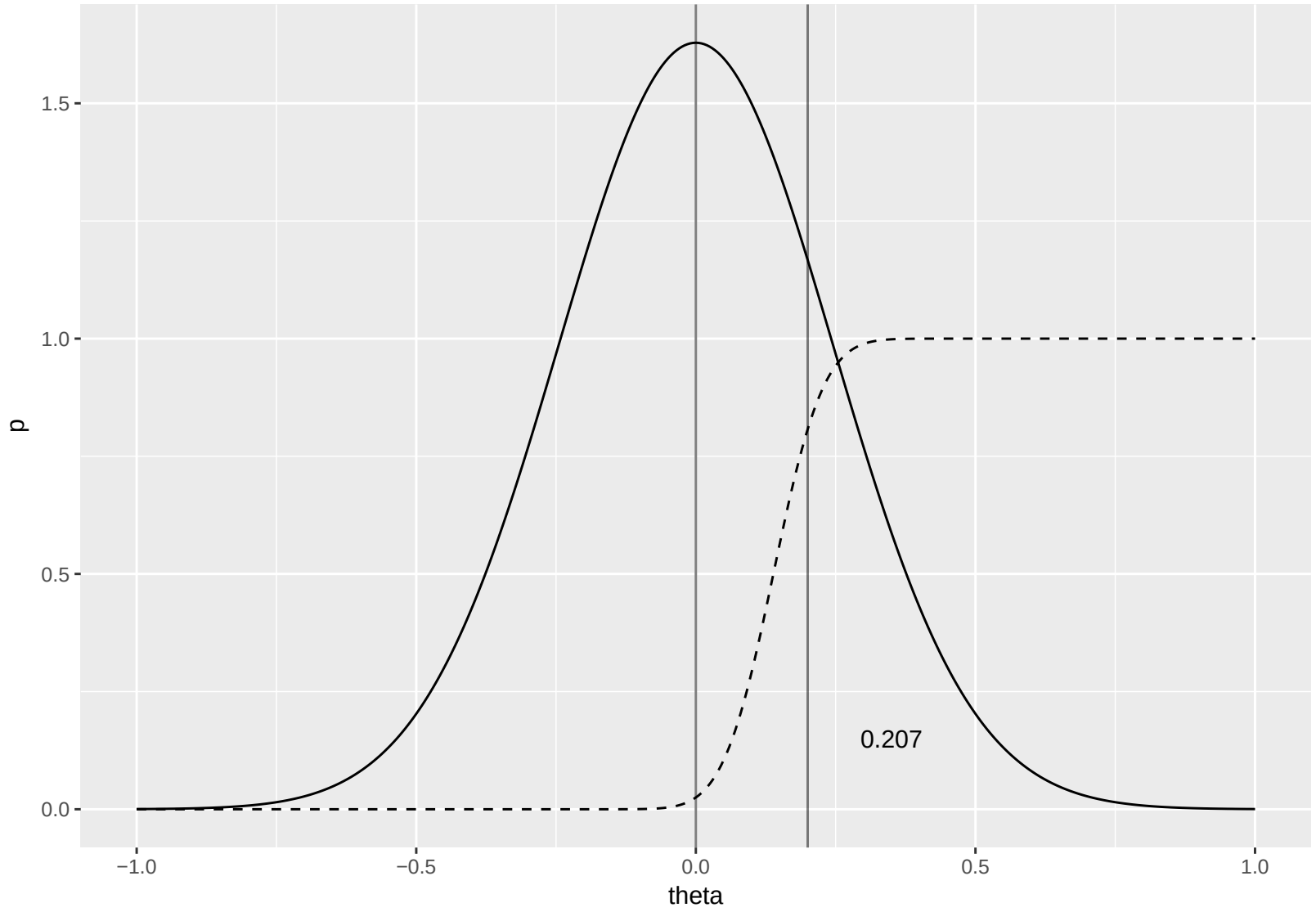
How to interpret $\mathbb{E}_\theta[u(d,\theta)]$?

Example

From O'Hagan, Stevens and Campbell (2005), "Assurance in clinical trial design", *Pharmaceutical statistics*:

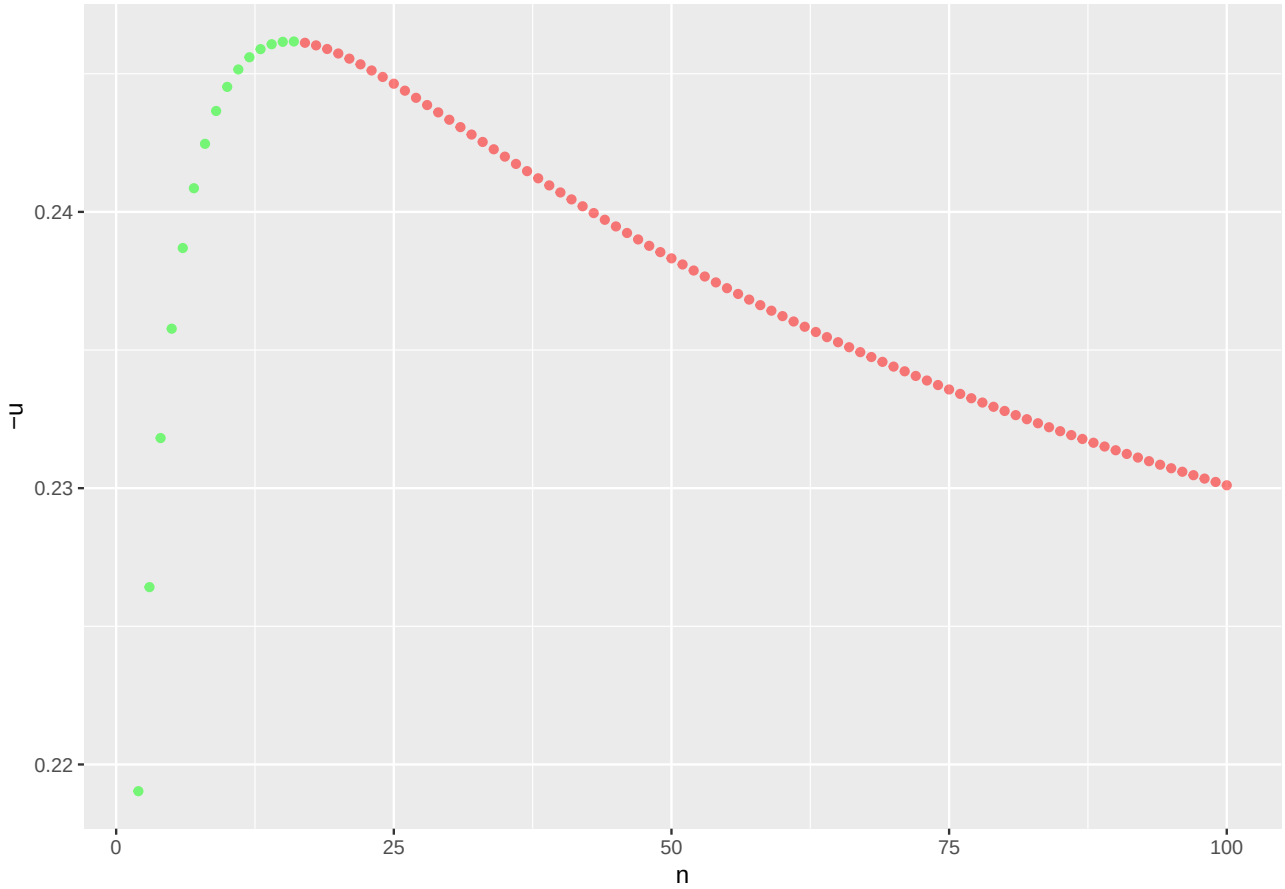
- Two arm trial with continuous outcome $y_i \sim N(\mu_i, \sigma^2)$ for groups $i = 0, 1$;
- Known $\sigma = 0.25$, unknown $\theta = \mu_1 - \mu_0$;
- Prior distribution $\theta \sim N(m, s^2)$, where $m = 0$, $s = 0.244$.
- $n = 25$ per arm gives 80% power to detect $\theta = 0.2$.

Example



Example - fixed α

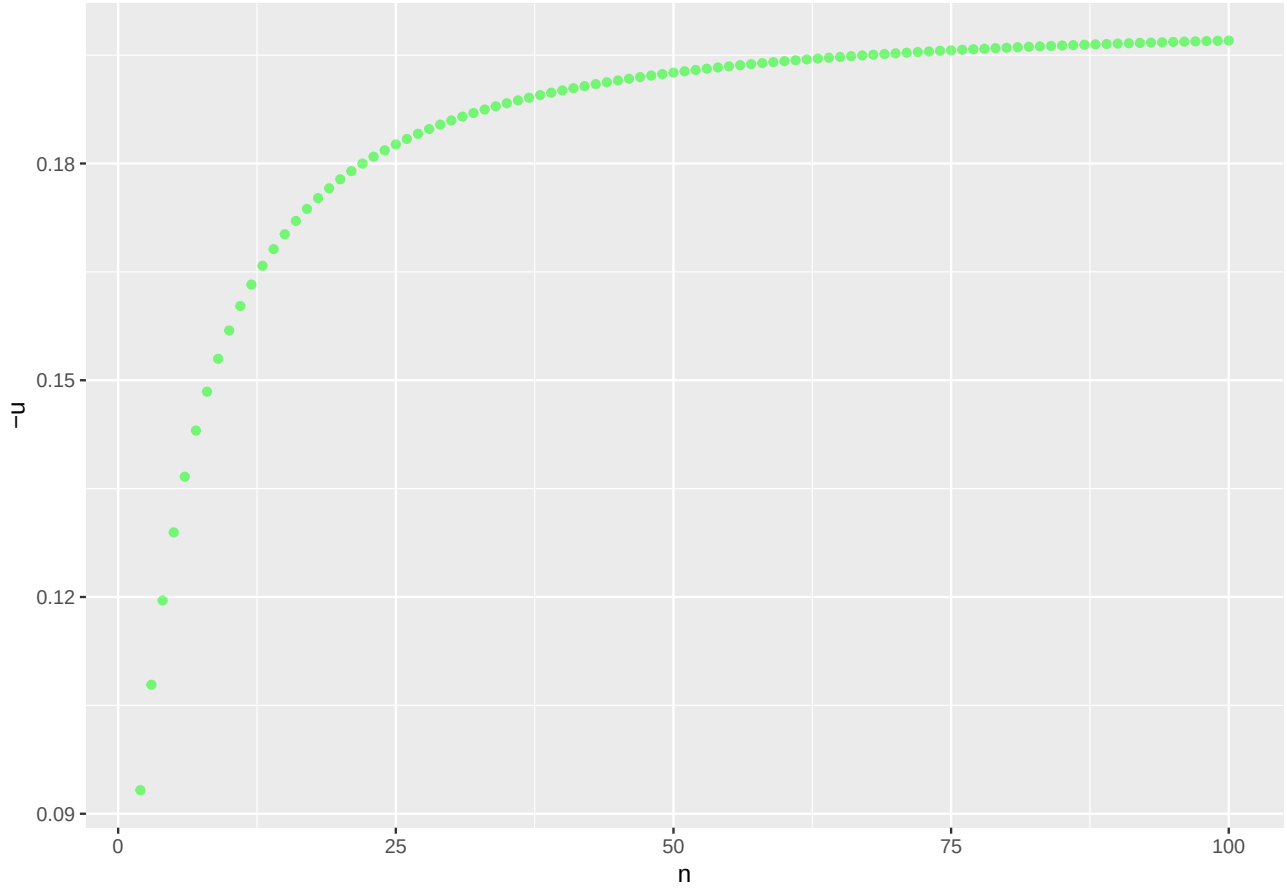
Expected utility for different n, all with $\alpha = 0.025$ (one-sided) and **MCID = 0.2**:



Maximum utility at $n = 16$.

Example - fixed α

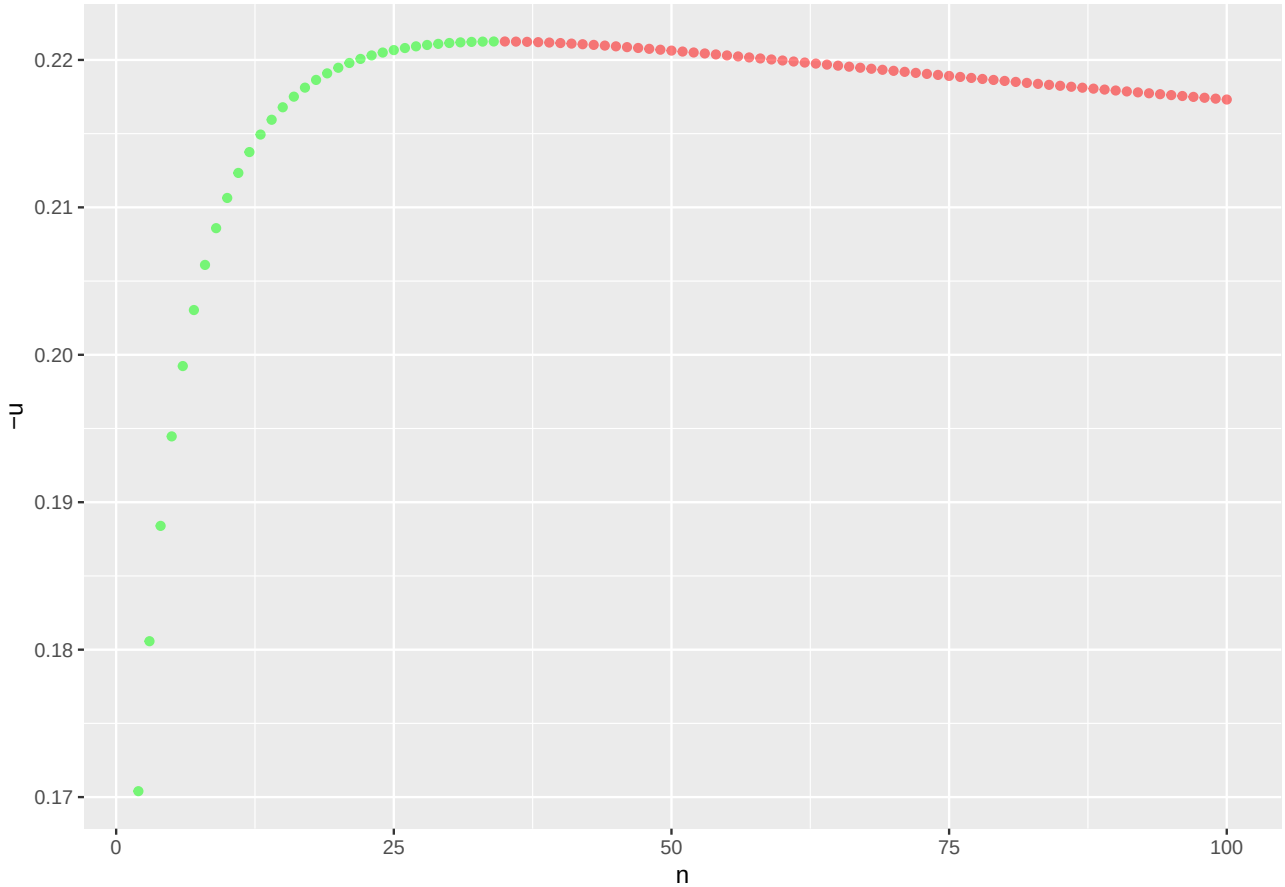
Expected utility for different n, all with $\alpha = 0.025$ (one-sided) and **MCID = 0.05**:



Maximum utility at $n > 100$.

Example - fixed α

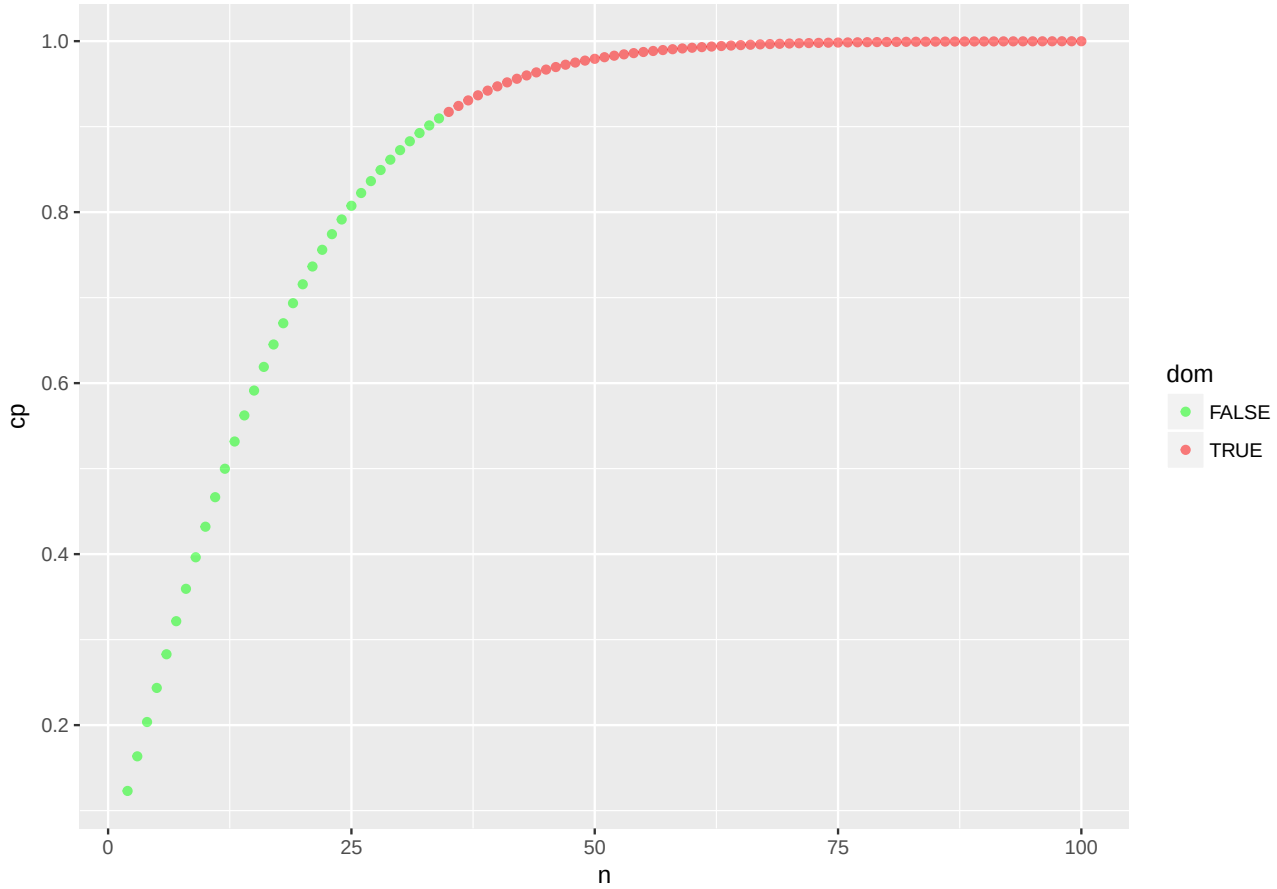
Expected utility for different n, all with $\alpha = 0.025$ (one-sided) and **MCID = 0.142**:



Maximum utility at $n = 34$.

Example - fixed α

Corresponding conditional power at $\theta = 0.2$, for **MCID = 0.142**:



Aside - the MCID

Cook *et al.*, "Specifying the target difference in the primary outcome for a randomised controlled trial: guidance for researchers", *Trials*, 2015.

"...the difference in the primary outcome value that the study is designed to detect *reliably*." [my italics]

"the smallest difference ...which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management."

Our definition implies that for $\theta \approx \theta^*$, we are indifferent between recommending and rejecting the treatment.

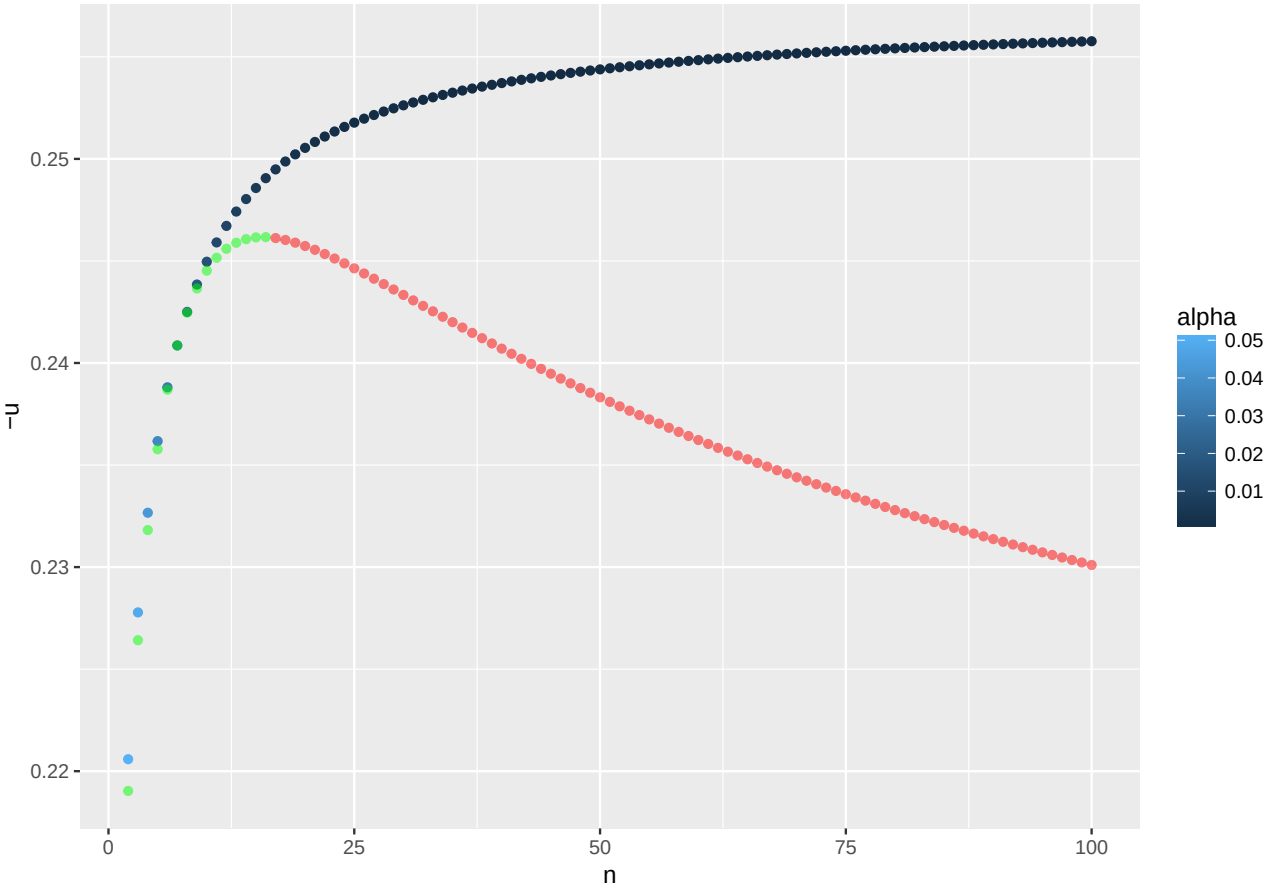
Aside - the MCID

Why $\theta^* = 0.142$?

- The original design was to give 80% power at $\theta = 0.2$ for $n = 25$.
- With the same sample size, we have $\approx 50\%$ power at $\theta = 0.142$.
- Should a frequentist design always have 50% power at (our) MCID? Is this the point of equipoise? Are MCIDs defined in practice as the effect we want to detect with 80% power? And if so, can we justify adjusting sample size and power around this MCID?

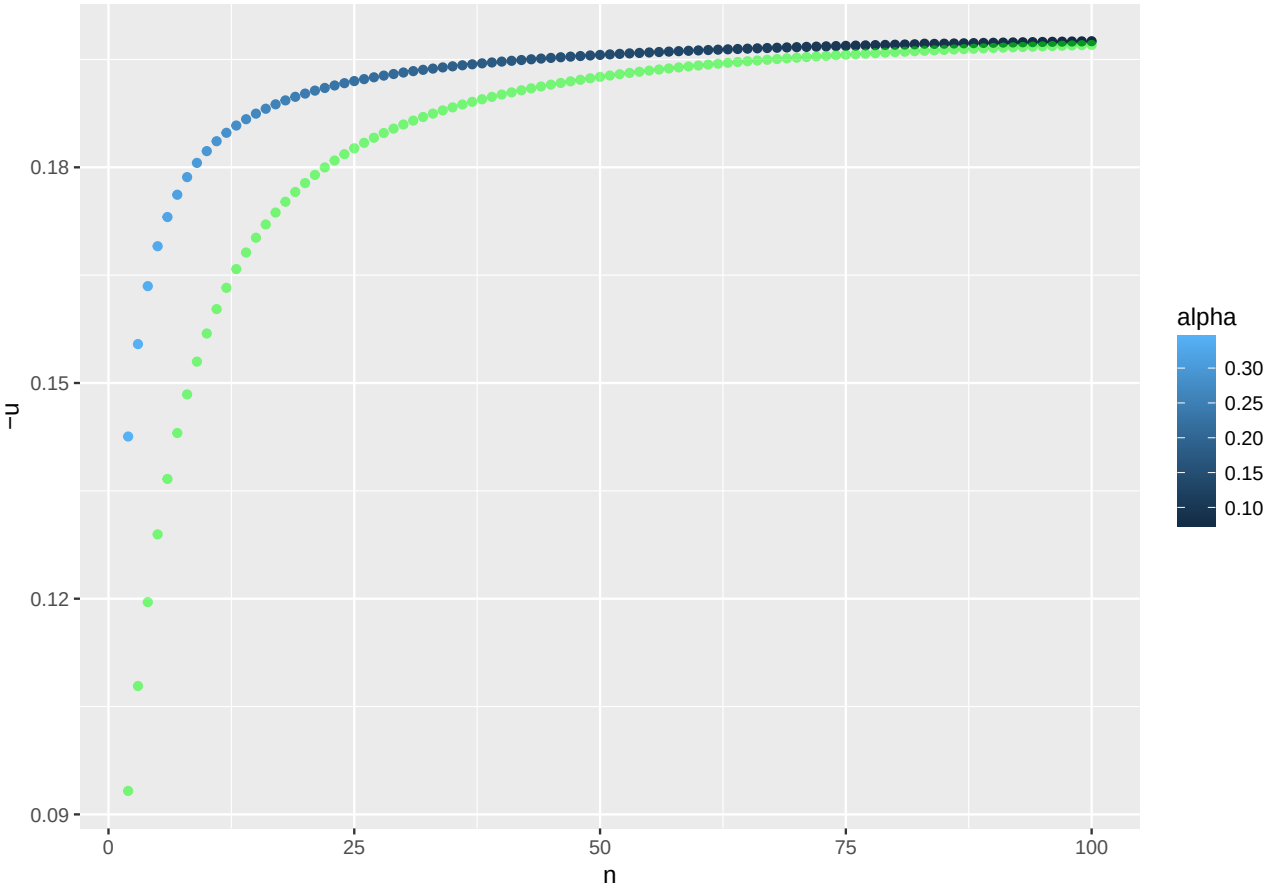
Example - free α

Maximising expected utility over α for each n - **MCID = 0.2**:



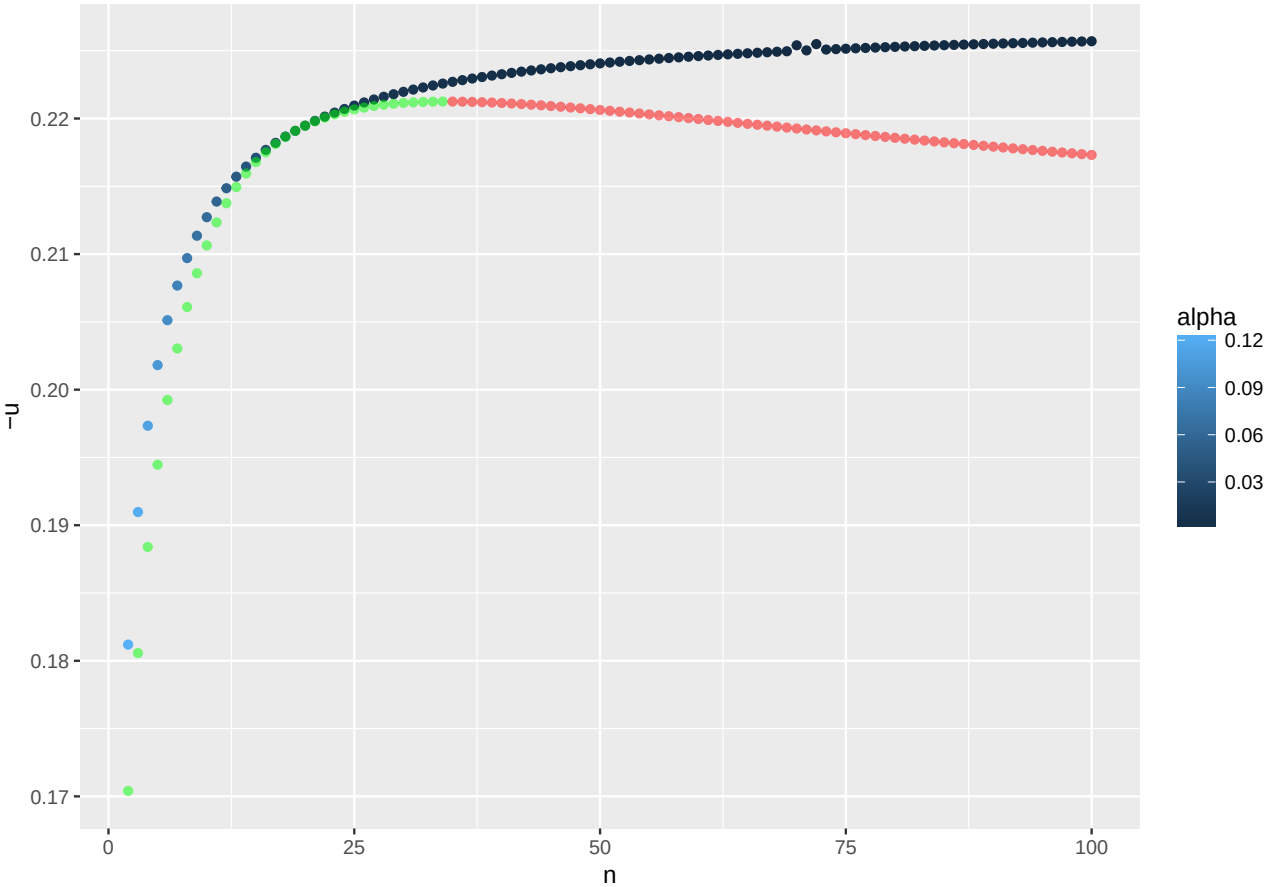
Example - free α

Maximising expected utility over α for each n - **MCID = 0.05**:



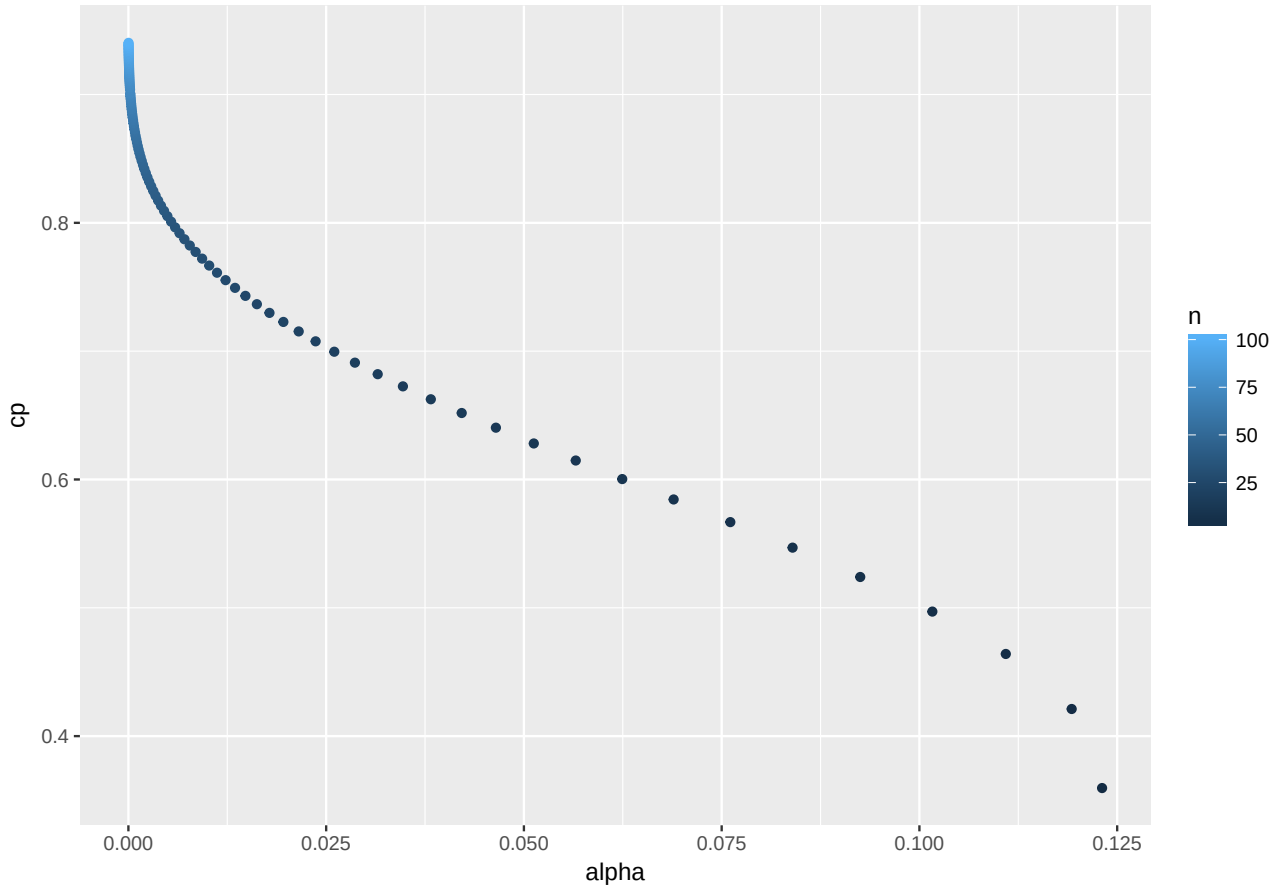
Example - free α

Maximising expected utility over α for each n - **MCID = 0.142**:



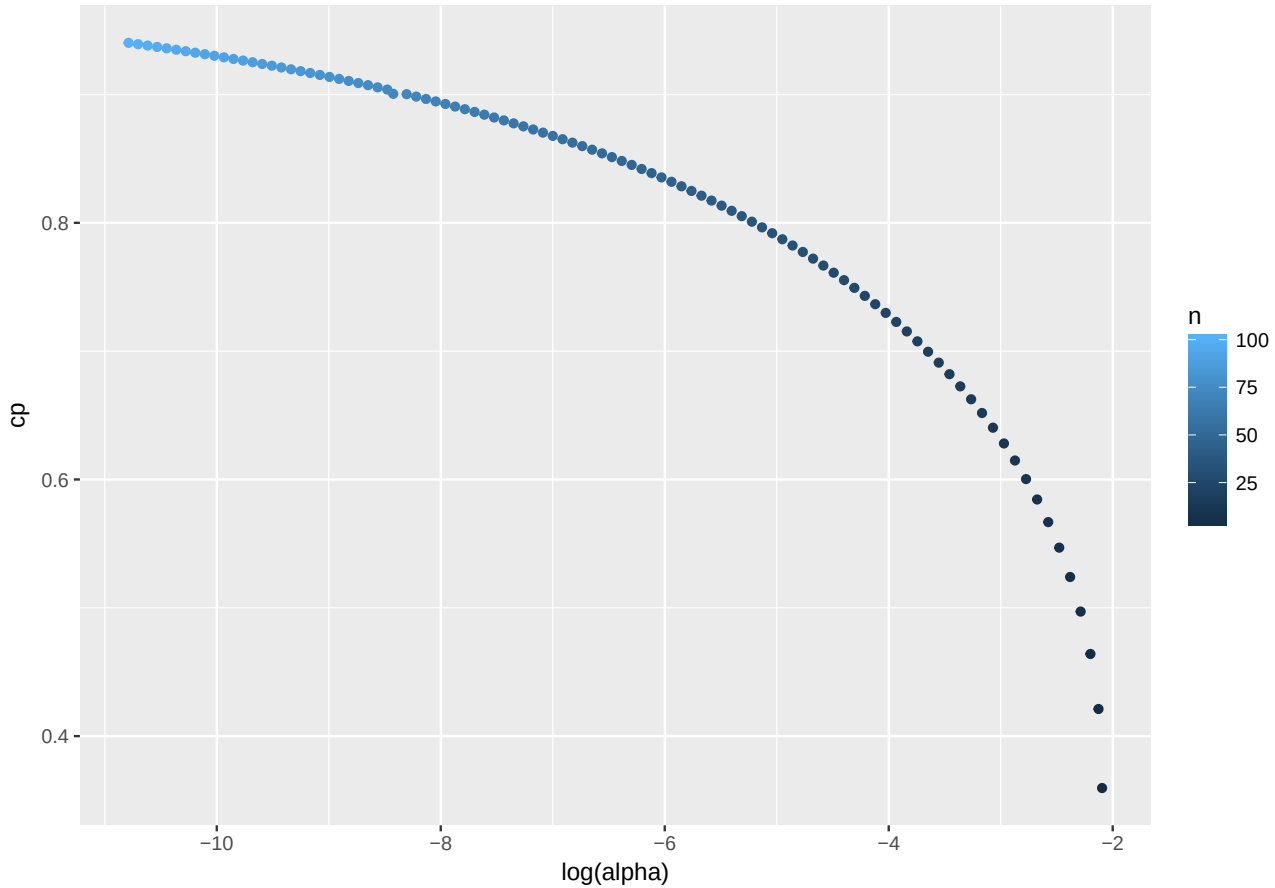
Example - free α

Operating characteristics (power at $\theta = 0.2$) of these efficient designs:



Example - free α

Operating characteristics (power at $\theta = 0.2$) of these efficient designs:



Critical values

Each optimal design specifies n and a critical value c to be used in the test. We see that:

- As n increases, the c approaches θ^* (from above);
- The optimal design for each n specifies a critical value which, if observed, would give us a posterior distribution on θ which gives $Pr[\theta > \theta^* | \bar{x} = c] = 0.5$.

Latter is a consequence of having a symmetric utility and a symmetric posterior distribution on θ.

Non-linear utility

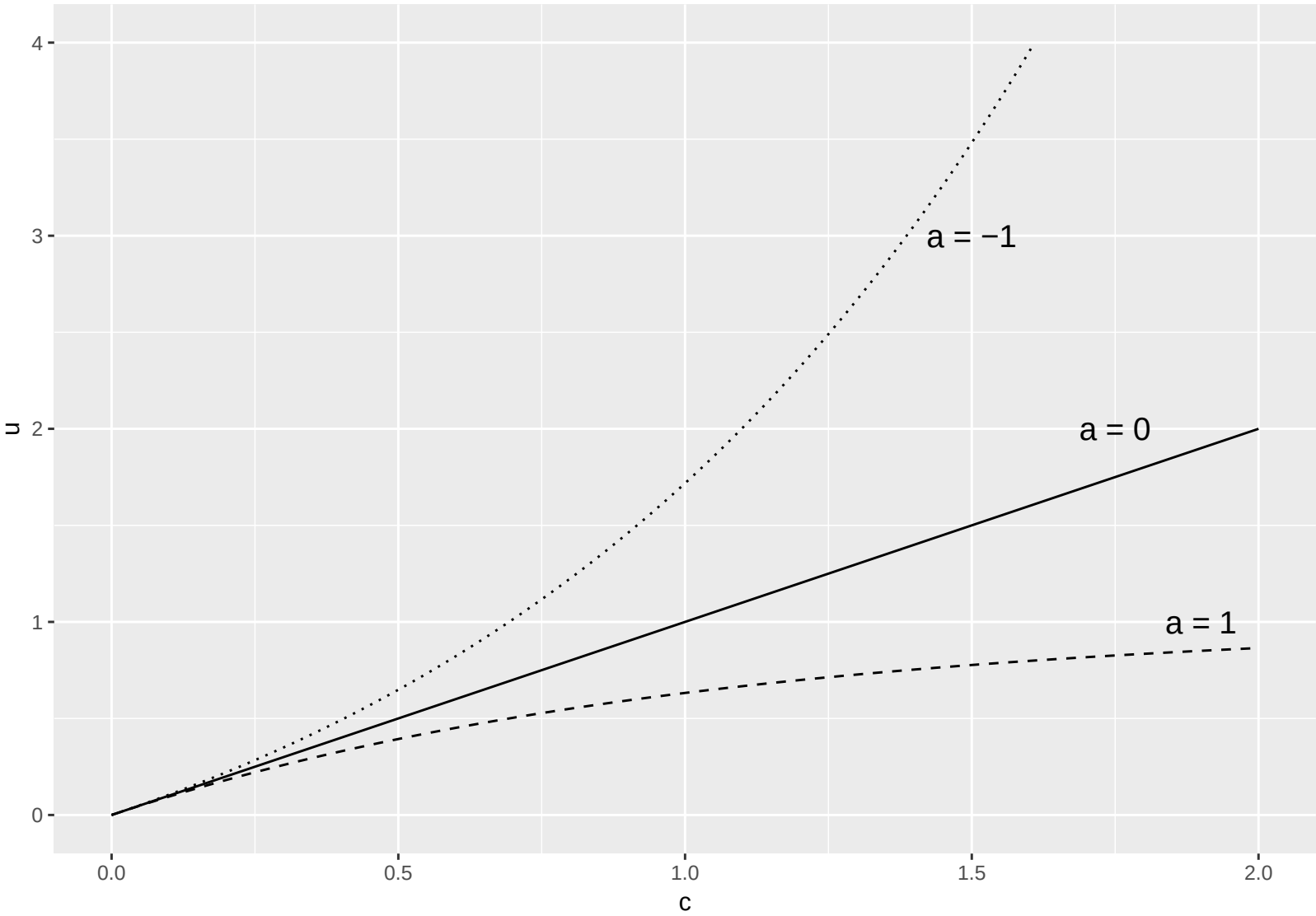
We transformed a value function into a utility function, but a risk-neutral approach made no difference. Consider the exponential utility:

$$u(c) = \begin{cases} (1 - e^{-ac})/a & \text{if } a \neq 0 \\ c & \text{if } a = 0 \end{cases} \tag{15}$$

- $a > 0 \Rightarrow$ risk *averse*;
- $a < 0 \Rightarrow$ risk *seeking*.

Uniquely, implies *constant absolute risk aversion*.

Non-linear utility



Recall that we assumed θ and n were preferentially independent. If we further assume that out attitude to risk on each attribute doesn't depend on the other, we have an additive utility function:

$$u(\theta, n) = u_1(\theta) + k u_2(n). \tag{16}$$

Given the individual utilities u_1, u_2 we can determine k by asking for the θ such that

$$(\theta, n_*) \sim (\theta_*, n^*) \tag{17}$$

What has all this got to do with pilot trials??

- Provides a simple way to use all information (recruitment, follow-up, variability, efficacy, safety, adherence . . .);
- Allows for trade-offs between intervention attributes;
- Models the link between the pilot and main study, allowing for optimal pilot design.

Using the same example but with $\sigma = 1$:

Pilot			Confirmatory				
α_1	β_1	n_1	α_2	β_2	n_2	$\mathbb{E}[n]$	$\mathbb{E}[u]$
0.41	0.97	209	0.0001	0.90	1227	781.3	0.393
0.43	0.94	150	0.018	0.77	408	341.1	0.391
0.43	0.78	43	0.121	0.65	119	98	0.380

Nuisance parameters

Nuisance parameters are those which will influence the power of the confirmatory trial, but otherwise are not in the utility function. Recall,

$$\mathbb{E}_\theta[u(d,\theta)] = \int \{g(d,\theta)(\theta - \theta^*) + [1 - g(d,\theta)](\theta^* - \theta)\} p(\theta) d\theta. \tag{18}$$

- Now a multi-dimensional integration over vector θ;
- Not analytically tractable $\rightarrow$ Monte Carlo approximation;
- Can use samples from the MCMC analysis of pilot data (for every design evaluated);
- Challenges when analytic expression for $g(d,\theta)$ unavailable.