MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

Sample Size Calculation

a closer look at 'hybrid' methods

K Kunzmann, M Grayling, KM Lee,
D Robertson, K Rufibach, JMS Wason

2020/11/04 (updated: 2020-11-04)


How many?

42 will do!

Preliminaries

- simple one-arm Wald test with test statistic

$$Z_n := \frac{\hat{\theta}_n - \theta_0}{\mathrm{SE}\left[\hat{\theta}_n\right]}$$

- w.l.o.g. $\theta_0 = 0$, $\mathcal{H}_0 = \theta \leq 0$, and effect size standardized
- assume that $\mathrm{SE}\left[\hat{\theta}_n\right]$ is deterministic function of n
  (think: σ known in the case of θ being the mean of a normal distribution)
- maximal one-sided type one error rate is externally given as $\alpha = 0.025$
  (i.e. critical value is 1.96, reject $\mathcal{H}_0 \Leftrightarrow Z_n > 1.96$)
- everything generalizes to
  - multi-arm comparisons
  - asymptotically normal test statistics (hazard ratios!)
  - etc.

Standard Approach

- probability to reject [1], $\Pr_\theta[Z_n > 1.96]$, is a function of θ and n
- pick point alternative $\theta_{alt} > 0$ and acceptable type-two error rate β
- chose smallest n such that $\Pr_{\theta_{alt}}[Z_n > 1.96] = 1 - \beta$
- n is monotone function of θ_{alt}: if $\theta_{alt} \nearrow$ then $n \searrow$
- in the following: $\beta = 0.2$, i.e. we aim for 80% power
- for $n = 42$ this works if $\theta_{alt} \approx 0.43$

wait, where does the point alternative come from?

[1] determining the sample size via power arguments is common but not the only criterion, could also use width of confidence interval, utility, ...

Choice of Point Alternative

MCID

- Minimal Clinically Important Difference (MCID)
- externally given
- $\theta \geq \theta_{\mathrm{MCID}} : \Pr_\theta[Z_n > 1.96] \geq 1 - \beta$
  i.e. all relevant effects are detected with minimum power (due to monotonicity)
- expensive: $n \to \infty$ as $\theta_{\mathrm{MCID}} \to 0$

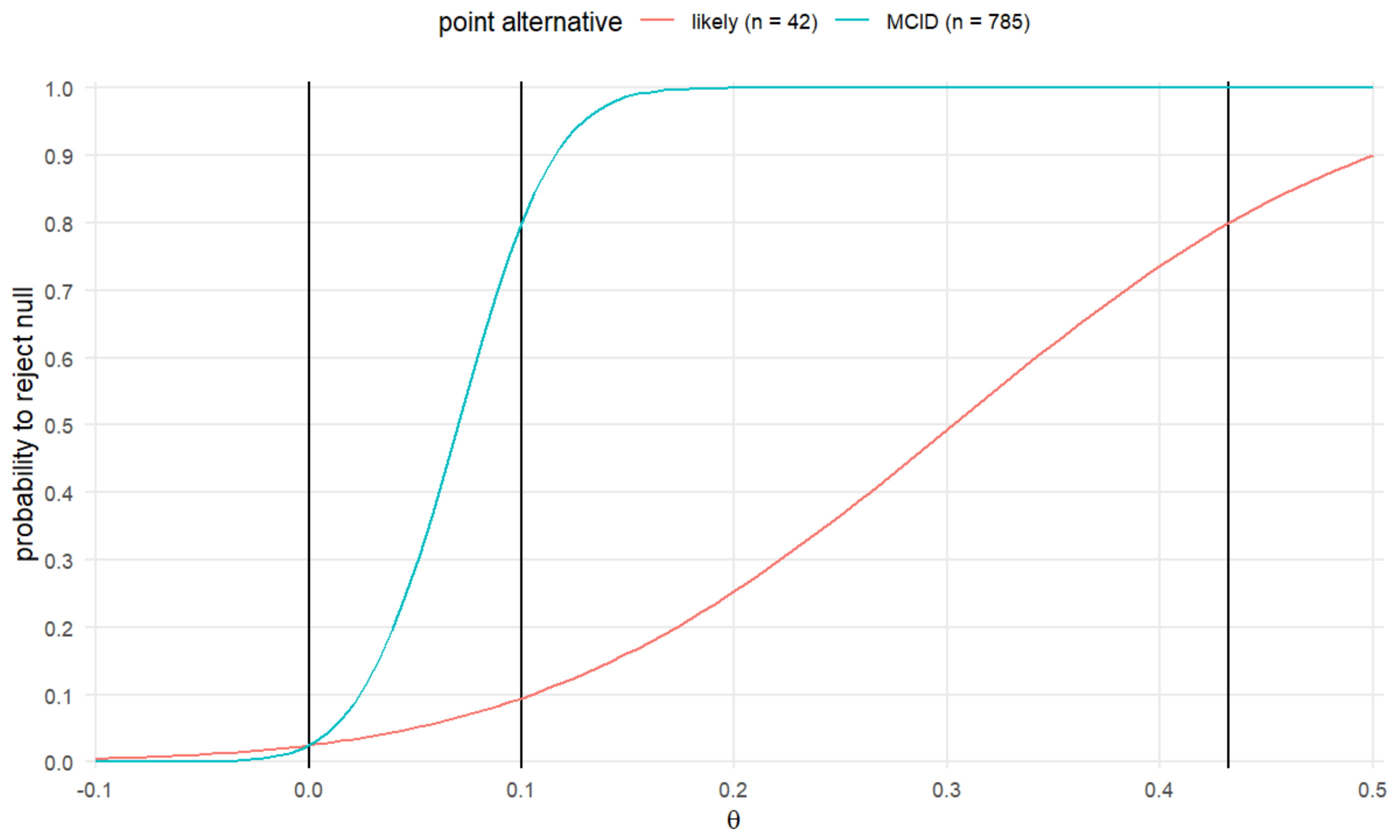
'likely' alternative

- sometimes it is 'very likely' that

  $$\theta >> \theta_{\mathrm{MCID}}$$

- deriving n from θ_{MCID} is ineffective
- $\rightsquigarrow$ consider all $\theta \geq \theta_{\mathrm{MCID}}$ and their 'relative likelihood'
- **what is 'relative likelihood'?**

'Power Plays'



point alternative
likely (n = 42)
MCID (n = 785)
probability to reject null
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
0.0
0.1
0.2
0.3
0.4
0.5
θ

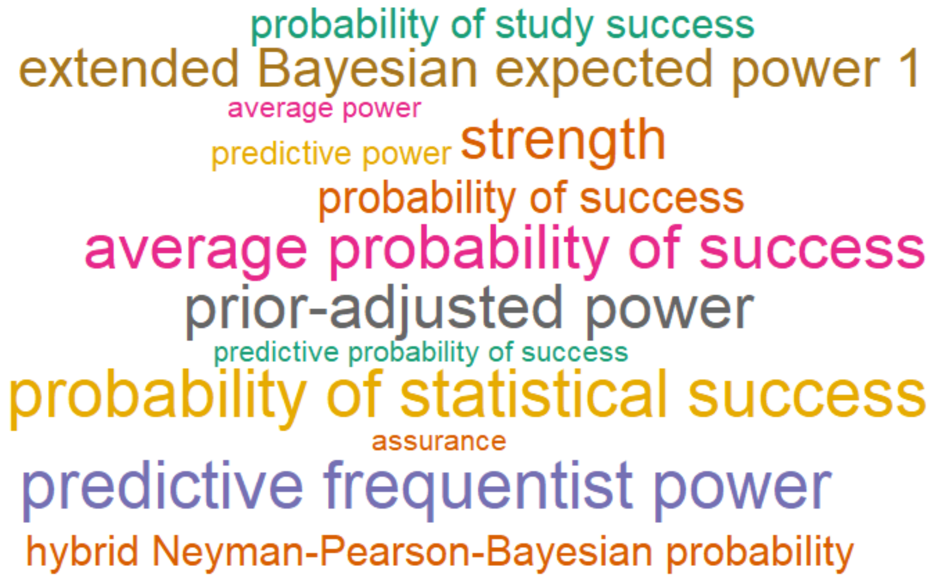
'Hybrid' Approaches

- 'hybrid' between **Bayesian** and **frequentist** methods[1]
- model unknown parameter as random variable Θ
- uses prior distribution to quantify planning uncertainty about Θ

$$\Pr[\Theta \leq x] = \int^{x} \underbrace{\varphi(\theta)}_{\text{prior density}} \,\mathrm{d}\,\theta$$

- $\varphi(\theta)$ is **weight function** encoding 'relative likelihood' values of θ
- final analysis remains entirely frequentist!

[1] not really Bayesian since Bayes Theorem is never invoked!



- see [https://arxiv.org/abs/2006.15715] for extensive literature review

'Bayesian' Point Alternative

Justify Point Alternative

- *"[...] a somewhat arbitrarily chosen location parameter of the [prior] distribution (for example the mean, the median or the 70th percentile) [...]"* could be used to determine n [1]
- simply Bayesian justification of the choice of point alternative θ_{alt}
- naive: $\theta_{alt} = \mathrm{E}[\Theta]$
- problems:

  1. probability to reject is non-linear function of effect size!

     $$\underbrace{\mathrm{Pr}_{\mathrm{E}[\Theta]}[Z_n > 1.96]}_{\text{probability to reject at expected effect}} \neq \underbrace{\mathrm{E}\big[\mathrm{Pr}_{\Theta}[Z_n > 1.96]\big]}_{\text{expected probability to reject}}$$

  2. $\mathrm{E}[\Theta]$ may lie in $\mathcal{H}_0$ $\rightsquigarrow$ required sample size undefined

[1] D J Spiegelhalter and L S Freedman. A predictive approach to selecting the size of a clinical trial, based onsubjective clinical opinion.Statistics in Medicine, 5(1):1–13, 1986.

Prior-Quantile Approach (1)

- use $(1-\gamma)$-quantile of the prior distribution *conditional* on a relevant effect:

$$\mathrm{Q}_{1-\gamma}[\Theta \mid \Theta \geq \theta_{\mathrm{MCID}}] := \inf_x Pr\big[\Theta \geq x \mid \Theta \geq \theta_{\mathrm{MCID}}\big] \geq \gamma$$

- *"*$\mathrm{Q}_{1-\gamma}[\Theta \mid \Theta \geq \theta_{\mathrm{MCID}}]$ *is the* θ *such that* $\mathrm{Pr}[\Theta \geq \theta] = \gamma$ *"*
- chose n such that power at $\theta_{alt} = \mathrm{Q}_{1-\gamma}[\Theta \mid \Theta \geq \theta_{\mathrm{MCID}}]$ is 80%
- $\mathrm{Q}_{1-\gamma}[\Theta \mid \Theta \geq \theta_{\mathrm{MCID}}] \geq \theta_{\mathrm{MCID}} \rightsquigarrow$ sample size always well-defined

$\rightsquigarrow$ nothing but a more formal way of justifying $\theta_{alt} \geq \theta_{\mathrm{MCID}}$

$\rightsquigarrow$ compatible with all sample size formulas!

Prior-Quantile Approach (2)

- how is that better than using $\theta_{alt} = \mathrm{E}[\Theta \mid \Theta \geq \theta_{\mathrm{MCID}}]$?
- recall that

$$\underbrace{\Pr_{\mathrm{E}[\Theta]}[Z_n > 1.96]}_{\text{probability to reject at expected effect}} \neq \underbrace{\mathrm{E}\left[\Pr_\Theta[Z_n > 1.96]\right]}_{\text{expected probability to reject}}$$

- $\theta \mapsto \Pr_\theta[Z_n > 1.96]$ is monotononously increasing, thus

$$\Pr[\Theta \geq \theta^* \mid \Theta \geq \theta_{\mathrm{MCID}}] = \gamma \quad \wedge \quad \Pr_{\theta^*}[Z_n > 1.96] \geq 1 - \beta$$

$$\Rightarrow \quad \Pr\left[\underbrace{\Pr_\Theta[Z_n > 1.96]}_{\text{'random power'}} \geq 1 - \beta \mid \Theta \geq \theta_{\mathrm{MCID}}\right] = \gamma$$

- in words: *"When calculating sample size based on the $1-\gamma$ quantile of the prior conditional on a relevant effect such that the power at this point alternative is $1-\beta$, the probability that 'random power' exceeds $1-\beta$ is γ."*
- requires additional parameter γ, how to justify? Is median ($\gamma = 0.5$) ok?

Probability of Success & Expected Power

Probability of Success (1)

- what constitutes a 'success' ?
- surprisingly vague in the literature, often needs to be reverse-engeneered from formula for 'probability of success'
- common: 'success' := 'reject $\mathcal{H}_0$', i.e.

$$\begin{aligned}\mathrm{PoS}'(n) &:= \Pr[Z_n > 1.96] \\ &= \int \Pr_\theta[Z_n > 1.96]\,\varphi(\theta)\,\mathrm{d}\,\theta\end{aligned}$$

- integration includes the null hypothesis - **type-one errors are successes!**

Greedy!

Probability of Success (2)

'greedy' definition:

'success' := 'reject $\mathcal{H}_0$'

$$\begin{aligned}\mathrm{PoS}'(n) :&= \Pr[Z_n > 1.96] \\ &= \int \Pr_\theta[Z_n > 1.96]\varphi(\theta)\,\mathrm{d}\,\theta\end{aligned}$$

long-term definition:

'success' := 'reject $\mathcal{H}_0$' **and** $\Theta \geq \theta_{\mathrm{MCID}}$

$$\begin{aligned}\mathrm{PoS}(n) :&= \Pr[Z_n > 1.96, \Theta \geq \theta_{\mathrm{MCID}}] \\ &= \int_{\theta_{\mathrm{MCID}}} \Pr_\theta[Z_n > 1.96]\varphi(\theta)\,\mathrm{d}\,\theta\end{aligned}$$

$$\mathrm{PoS}'(n) = \underbrace{\Pr[Z_n > 1.96, \Theta \leq 0]}_{\text{probability of type-one error}} + \underbrace{\Pr[Z_n > 1.96, 0 < \Theta < \theta_{\mathrm{MCID}}]}_{\text{probability of irrelevant rejection}} + \mathrm{PoS}(n)$$

in practice, $\Pr[Z_n > 1.96, \Theta \leq 0]$ small due to strict type-one error rate control at α and most prior mass in $\mathcal{H}_0$ concentrated on θ_0.

Probability of Success (3)

- critically depends on definition of 'success'
- numerical difference between $\mathrm{PoS}(n)$ and $\mathrm{PoS}'(n)$ mostly negligible in practice (but not always) [1]
- remember: if you go with $\mathrm{PoS}'(n)$, type-one errors are successes!
- pick carefully, be explicit, justify!

[1] https://arxiv.org/abs/2006.15715

Sample Size Calculation with PoS

- idea: determine n such that $\mathrm{PoS}(n) = 1 - \beta$
- note that

$$\begin{aligned}
\mathrm{PoS}(n) =& \Pr[Z_n > 1.96, \Theta \geq \theta_{\mathrm{MCID}}] \\
=& \Pr[Z_n > 1.96 \mid \Theta \geq \theta_{\mathrm{MCID}}] \ \Pr[\Theta \geq \theta_{\mathrm{MCID}}] \\
\leq& \Pr[\Theta \geq \theta_{\mathrm{MCID}}]
\end{aligned}$$

- **problem:** what if $1 - \beta > \Pr[\Theta \geq \theta_{\mathrm{MCID}}] \rightsquigarrow n = ?$ $\lightning$
- **solutions:**
  1. either re-scale usual $1 - \beta$ to make it problem-specific (how?) ...
  2. ... or re-scale $\mathrm{PoS}(n)$ to $[0, 1]$

Enter: Expected Power

- luckily, there is a natural way to re-scale $\mathrm{PoS}(n)$ to $[0, 1]$

$$
\begin{aligned}
\mathrm{PoS}(n) &= \Pr[Z_n > 1.96 \mid \Theta \geq \theta_{\mathrm{MCID}}]\ \Pr[\Theta \geq \theta_{\mathrm{MCID}}] \\
\Leftrightarrow \quad \underbrace{\mathrm{PoS}(n) / \Pr[\Theta \geq \theta_{\mathrm{MCID}}]}_{\in[0,1]} &= \Pr[Z_n > 1.96 \mid \Theta \geq \theta_{\mathrm{MCID}}] \\
&= \mathrm{E}\left[\underbrace{\Pr_{\Theta}[Z_n > 1.96] \mid \Theta \geq \theta_{\mathrm{MCID}}}_{\text{'random power'}}\right] \\
&=: \mathrm{EP}(n)
\end{aligned}
$$

- $\mathrm{EP}(n)$ is 'expected power' and $\mathrm{PoS}(n) = \Pr[\Theta \geq \theta_{\mathrm{MCID}}]\ \mathrm{EP}(n)$
- joint probability of rejection and relevant effect $\mathrm{PoS}(n)$ vs. conditional probability of rejection given relevant effect $\mathrm{EP}(n)$

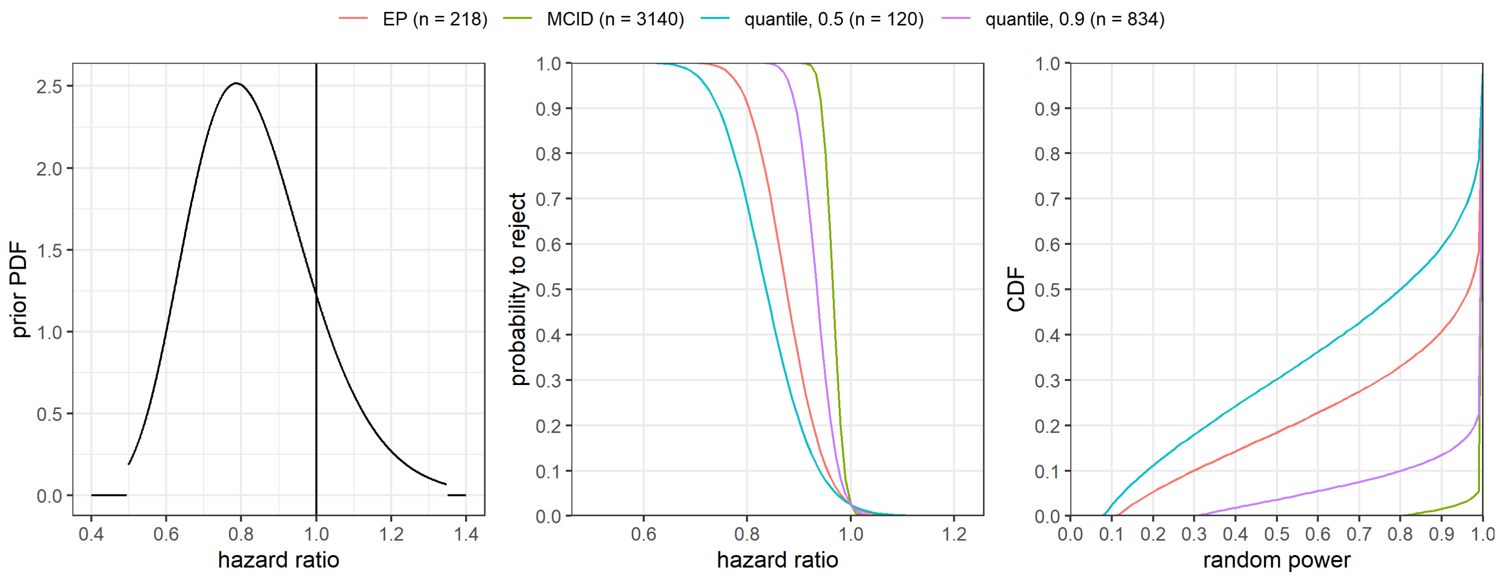
Just a Different Weight Function...

$$
\begin{aligned}
&\Pr[\Theta \geq \theta_{\mathrm{MCID}}]\ \mathrm{EP}(n) \\
&= \Pr[\Theta \geq \theta_{\mathrm{MCID}}] \int_{\theta_{\mathrm{MCID}}} \Pr_\theta[Z_n > 1.96]\, \varphi(\theta \,|\, \Theta \geq \theta_{\mathrm{MCID}})\, \mathrm{d}\,\theta \\
&= \int_{\theta_{\mathrm{MCID}}} \Pr_\theta[Z_n > 1.96]\, \underbrace{\Pr[\Theta \geq \theta_{\mathrm{MCID}}]\, \varphi(\theta \,|\, \Theta \geq \theta_{\mathrm{MCID}})}_{\varphi(\theta)}\, \mathrm{d}\,\theta \\
&= \mathrm{PoS}(n)
\end{aligned}
$$

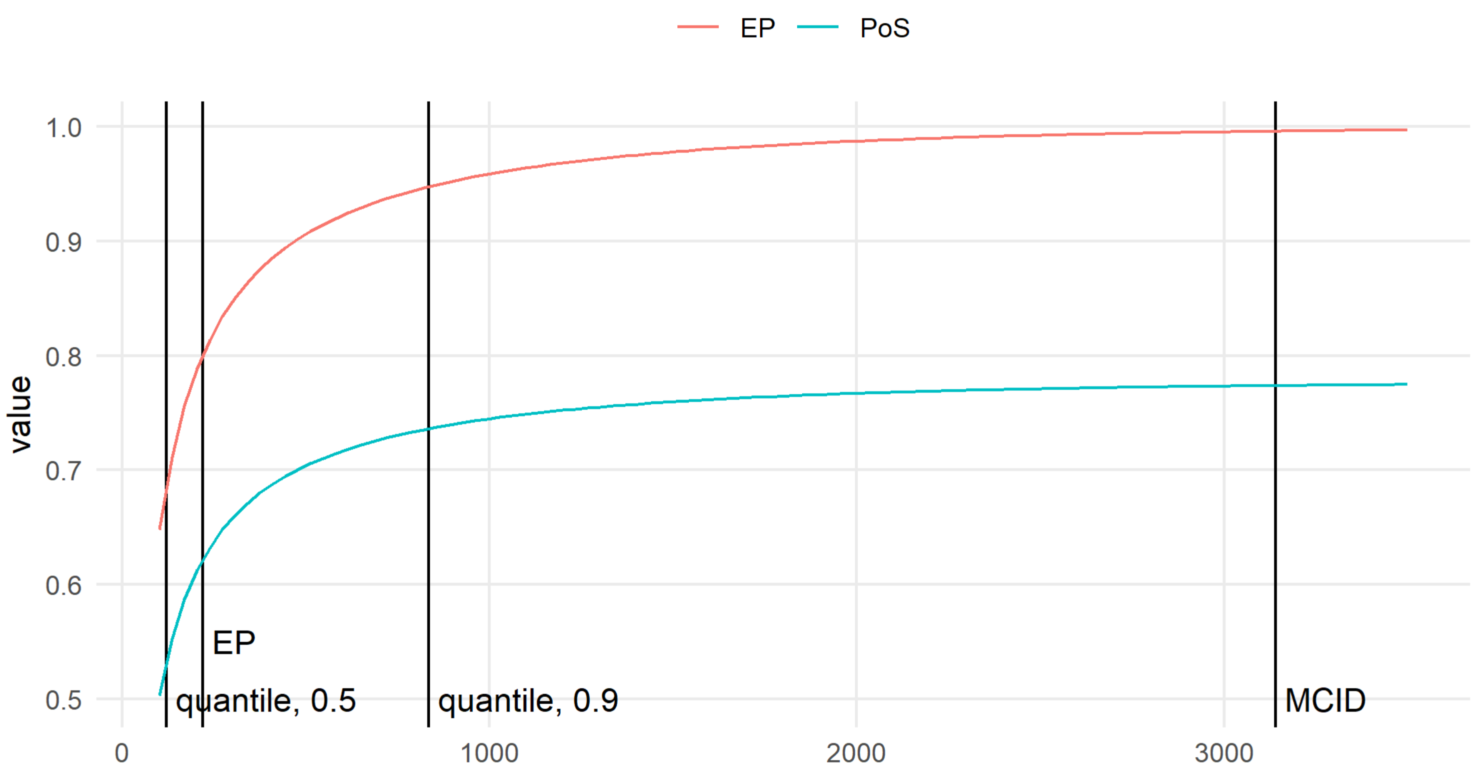
Example

Single-Arm, Survival Endpoint

- effect on **hazard ratio** scale $\mathrm{HR} = \exp(-\theta)$
- $\mathrm{HR}_{\mathrm{MCID}} = 0.95$, $\mathrm{HR}_0 = 1$
- $\alpha = 0.025$, $\beta = 0.2$



Expected Power and Pos vs n



Shiny/Code/Pre-Print

Thanks to Google Cloud (https://cloud.google.com/), OVH (https://www.ovh.com/), GESIS Notebooks (https://notebooks.gesis.org) and the Turing Institute (https://turing.ac.uk) for supporting us!

binder

Starting repository: kkmann/sample-size-calculation-under-uncertainty/master

- pre-print: [https://arxiv.org/abs/2006.15715]
- code: [https://github.com/kkmann/sample-size-calculation-under-uncertainty]
- Shiny latest
- follow-up: predictive power and sample size recalculation [https://arxiv.org/abs/2010.06567]

Summary

- 'Bayesian' choice of θ_{alt} not straight-forward: power function is non-linear!
- 'quantile approach' is principled and simple to implement but requires additional parameter γ, confusing!
- definition of 'probability of success' is treacherous, not suitable for sample size calculation!
- 'expected power' is natural extension of power calculation to situation with prior information
- 'hybrid' methods make distinction between 'likely' & 'relevant' effects transparent
- Bayes Theorem not used - relevant when looking at **interim adaptations** [1]
- **no magic bullet** but consistent framework for incorporating uncertainty!

[1] https://arxiv.org/abs/2010.06567

Thanks!

www.mrc-bsu.cam.ac.uk

@kevin_kunzmann

Slides created with the R package xaringan.