Power, assurance and probability of success in clinical trial design

February 2020

Contents

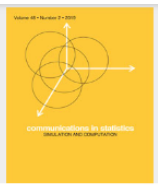
Main idea

- The notion of statistical power is by definition “power against a certain alternative”.
- In clinical trial design, statistical power is defined as the probability of rejecting the null hypothesis at a pre-specified true clinical treatment effect, i.e. conditional on the true but actually unknown effect.
- In practice, the true effect is not a fixed value, and therefore the planned trial could be underpowered or overpowered.
- In order to incorporate the uncertainties of this observed treatment effect, a Bayesian assurance has been proposed as an alternative to the conventional statistical power.
- This is defined as the unconditional probability of rejecting the null hypothesis i.e. without fixing any specific effect level.
- We will explain the transition from conventional statistical power to Bayesian assurance.

PHARMACEUTICAL STATISTICS
Pharmaceut. Statist. 2005; **4**: 187–201
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/pst.175

Assurance in clinical trial design

Anthony O'Hagan[1,*,†], John W. Stevens[2] and Michael J. Campbell[3]
[1] *Centre for Bayesian Statistics in Health Economics, University of Sheffield, UK*
[2] *AstraZeneca R&D Charnwood, Loughborough, UK*
[3] *School of Health and Related Research, University of Sheffield, UK*



Type I and II errord

- A conventional frequentist design depends on:

  a) The hypothesis to be tested: H_0 : Drug “ineffective” or $\delta=0$

  b) Type I error: The probability of rejecting the null hypothesis H_0: $\delta=0$, when it is true

  c) Type II error: The probability of “accepting” the null hypothesis when it is false, i.e. when the alternative hypothesis, H_1,= Drug “has a certain effect” δ is true

- *Power (1-Type II error) is the probability of rejecting the null hypothesis if the treatment effect equals the assumed value δ. Power = Probability of success?*

H_1 = The Alternative Hypothesis

Of course, we do not know the true effect δ in advance. How do we deal with that?

- The specification of the treatment effect "*may be based on a judgement concerning the minimal effect which has clinical relevance ….. or on a judgement concerning the anticipated effect of the treatment*." ICH.E9

- However, the specified treatment effect may need to be set larger than this value due to e.g. what is commercially viable (TPP).

- In practice, power can also be affected by
  - availability of patients
  - financial constraints

Power and success

- Consider a simple one-sided test:
- R= reject the null hypothesis
- Power = P(R| H_1) where H_1 : Effect = δ, i.e. Power is a conditional probability as a function of δ
- Under H_1 we can interpret power as a "probability of success" i.e. the probability of rejecting the null hypothesis **IF** the drug is actually working
- However, the true treatment effect is uknown

Assurance

- Assurance is the unconditional probability that the trial will end with the desired outcome regardless of what the effect = δ is.
- In most cases, Assurance can be expressed as the expectation of the (conditional) power with respect to a prior distribution of θ.
- It entails a Bayesian perspective because it requires a prior distribution for θ
- In that sense, assurance can be seen as average a priori power.

$$P(R) = \int P(R,\theta)d\theta$$
$$= \int P(R \mid \theta).P(\theta)d\theta$$
$$= E_\theta[P(R \mid \theta)]$$

- In general, assurance can have more complex relation to power

Prior information

The specification of prior information

- May be an approximate judgement
- May be based on little prior data
- Can be strong or weak
- Phase IIb can provide prior for Phase III etc

- The effort necessary to specify the prior distribution depends on the context, including
  - The role of the assurance in contributing to decisions regarding future development
  - The financial benefit and/or risk at stake

At the design stage

- Adopting a Bayesian approach at the design stage is reasonable because
  - It takes the uncertainty about the effect into account
  - The design of a trial is an internal decision
  - The design should take into account all available information
- However: It is not necessary to combine the prior information with the trial data at the end analysis stage!

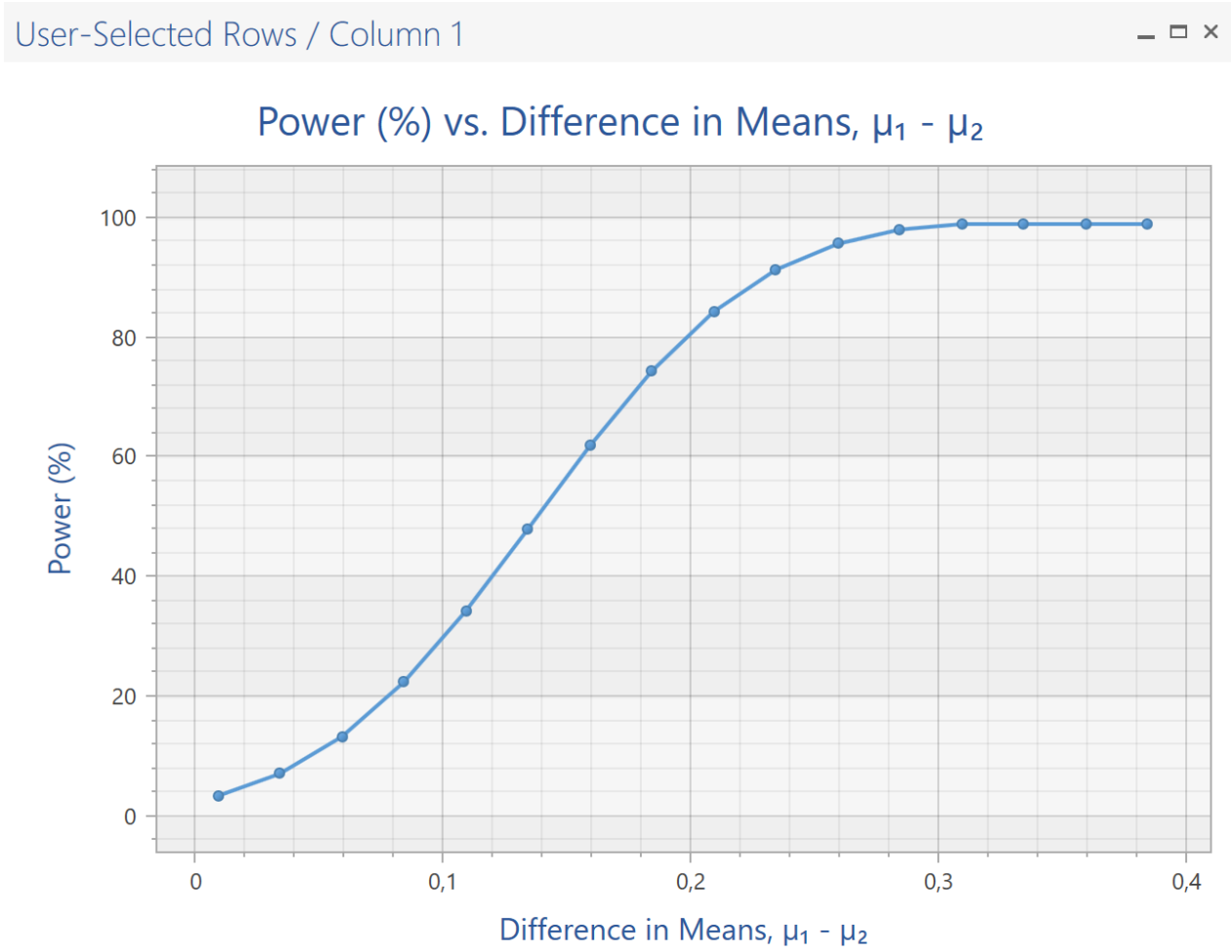
Example

- Phase 2a superiority trial two arms trial to compare the effect of reducing CRP in patients with e.g. Rheumatoid Arthritis (RA)
- The outcome is reduction in CRP after four weeks relative to baseline
- The analysis will involve a 1-sided test at the 2.5% significance level
- The variances of CRP reduction in the two treatment groups are assumed to be known with values $\sigma^2_1 = \sigma^2_2 = 0.25^2$
- We require 80% power to detect a treatment effect of $\delta = 0.20$

Sample size calculation

MTT0-1 / Two Group t-test of Equal Means

	1
Test Significance Level, α	0,025
1 or 2 Sided	1
Group 1 Mean, μ_1	
Group 2 Mean, μ_2	
Difference in Means, $\mu_1 - \mu_2$	0,200
Common Standard Deviation, σ	0,250
Effect Size, $\delta=\lvert\mu_1-\mu_2\rvert/\sigma$	0,800
Power (%)	80%
n per Group	**26**



Example

- Since we assume that the population variances are known, the model for the data is:

$$x_{ij} \sim N(\mu_i, \sigma_i^2)$$

- A hypothesis test will be based on the sample mean difference:

$$\bar{x}_2 - \bar{x}_1 \sim N(\delta, \tau^2)$$

- where

$$\delta = \mu_2 - \mu_1 \qquad \tau^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Prior distribution

- Suppose that the prior distribution for the population mean difference is conjugate normal)

$$\delta = \mu_2 - \mu_1 \sim N(m, v)$$

- Then the unconditional distribution is

$$\bar{x}_2 - \bar{x}_1 \sim N(m, \tau^2 + v)$$

from which we can calculate the assurance of rejecting the null hypothesis

- For the calculation of assurance, we suppose that

$$E[\delta] = 0.2 \qquad Var[\delta] = 0.0625$$

Assurance

- A one-sided 100α% significance test of the null hypothesis that $\delta = 0$ against the alternative that $\delta > 0$ will reject the null hypothesis if

$$R=\{\bar{x}_2 - \bar{x}_1 > \tau Z_\alpha\}$$

- where Z_α is the upper 100α% significance point of the standard normal distribution.

- The assurance of this outcome is

$$\bar{x}_2 - \bar{x}_1 \sim N(\delta, \tau^2)$$
$$\delta = \mu_2 - \mu_1 \sim N(m, v)$$
$$\bar{x}_2 - \bar{x}_1 \sim N(m, \tau^2 + v)$$

$$\gamma = \underbrace{P(\bar{x}_2 - \bar{x}_1 > \tau Z_\alpha)}_{P(R)} = \Phi\left(\frac{-\tau Z_\alpha + m}{\sqrt{\tau^2 + v}}\right)$$

Come from the prior

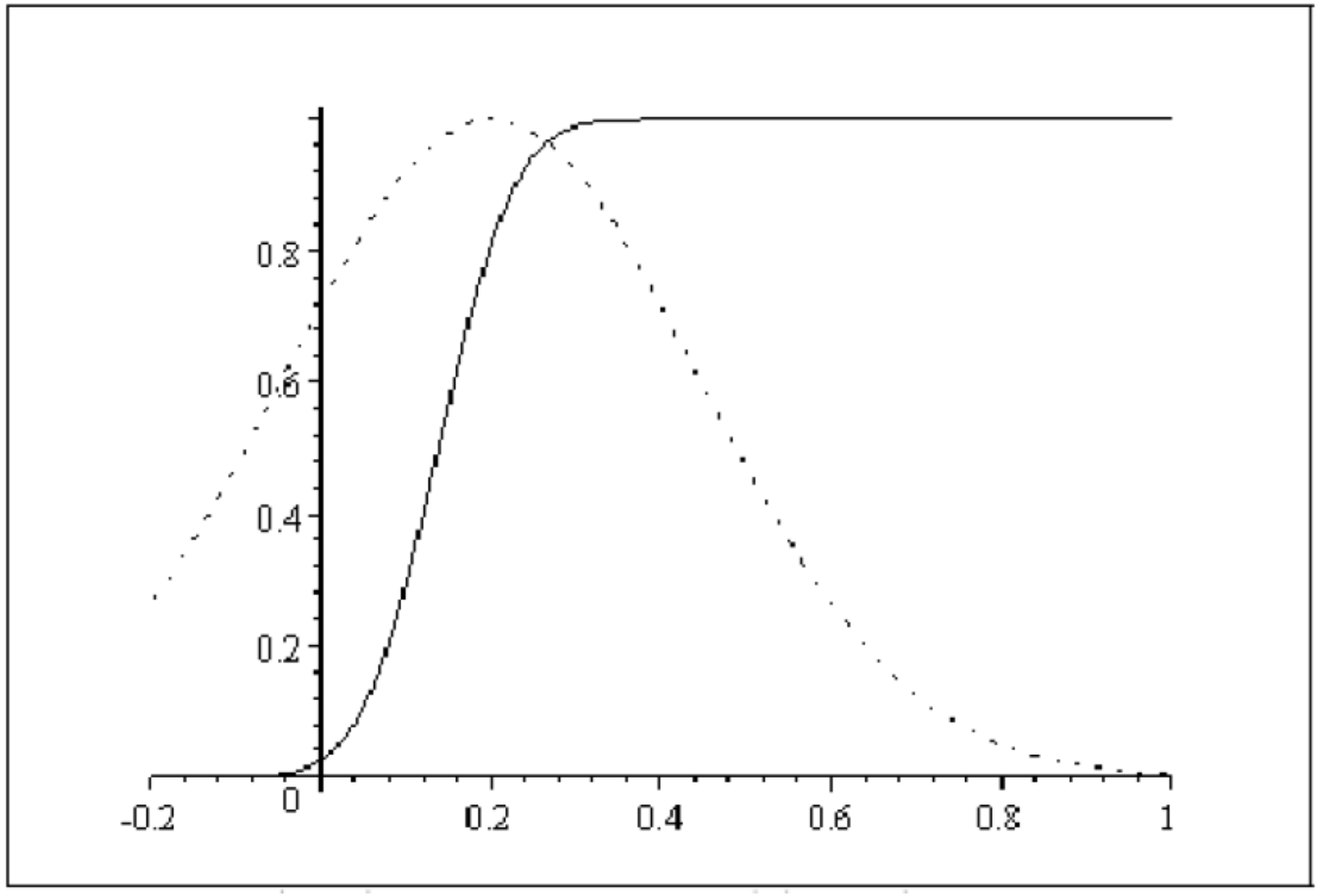
$$P(R) = \int P(R,\theta)d\theta = \int P(R|\theta).P(\theta)d\theta = E_\theta[P(R|\theta)]$$

Example

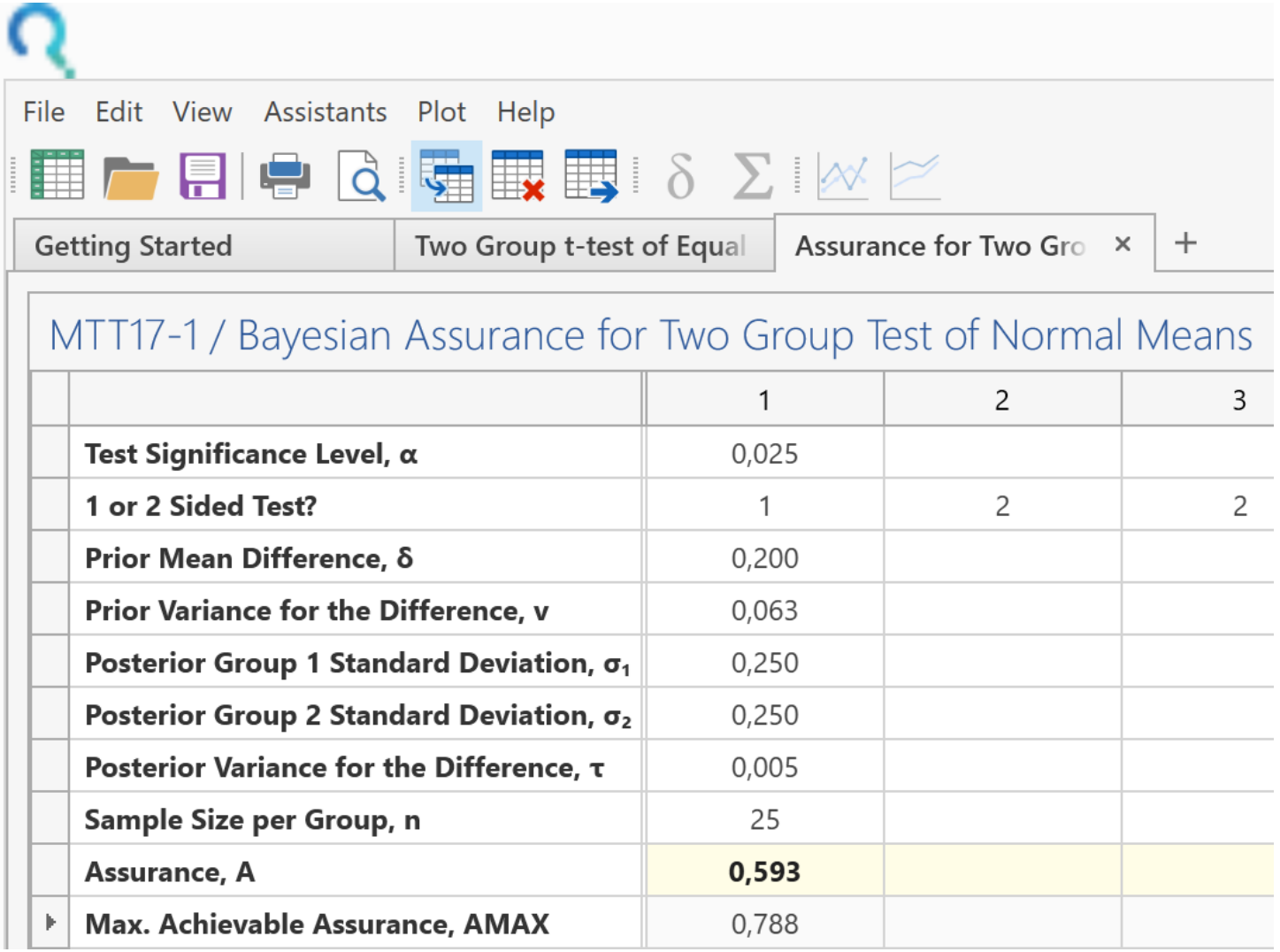
- If n=∞, τ^2=0 so the assurance cannot exceed

$\Phi\left(m/\sqrt{v}\right)$ = $\Phi(0.2/0.25)$=0.793

- For n=25 the assurance is 0.595 which is 75% of the maximum 0.80 (power)



Power (solid line) and prior (dotted line)



	1	2	3
Test Significance Level, α	0,025		
1 or 2 Sided Test?	1	2	2
Prior Mean Difference, δ	0,200		
Prior Variance for the Difference, v	0,063		
Posterior Group 1 Standard Deviation, σ_1	0,250		
Posterior Group 2 Standard Deviation, σ_2	0,250		
Posterior Variance for the Difference, τ	0,005		
Sample Size per Group, n	25		
Assurance, A	**0,593**		
Max. Achievable Assurance, AMAX	0,788		

In general

- In simple situations the solution can be derived analytically.
- Answers will need to be derived using (Bayesian) clinical trial simulation when:
    - The prior is in a non-conjugate form
    - The definition of a successful outcome involves several variables and/or several trials

Simulations

- The process involves:
  - Sampling (simulation) of θ from the prior distribution
  - Sampling data (or the sufficient statistics) using the design, statistical model and the sampled value of θ
  - Determining which outcomes occur
  - Repeating many times and recording the proportion proportion of simulations in which each outcome occurs
- The process is similar to conventional clinical trial simulation except for the step of sampling from the prior distribution

Clinical relevance and risk/benefit

$$\gamma = P(\bar{x}_2 - \bar{x}_1 > \tau Z_\alpha) = \Phi\left(\frac{-\tau Z_\alpha + m}{\sqrt{\tau^2 + v}}\right)$$

- The discussion above refers to clinical significance and is also called predictive probability of statistical significance as well as Bayesian power (PPoS1)
- Of course, clinical relevance of the observed difference δ between treatments is also required for success.
- We define the probability of clinical relevance, PPoS2, as the probability exceeding a pre-defined minimal clinically relevant threshold
- The decisions regarding licensing are taken based on a benefit-risk assessment. Several quantitative methodologies have been proposed e.g. Multi-Criteria Decision Analysis (MCDA) leading to (PPoS3)

Summary

- Assurance can be a good addition to measure of the sponsor’s risk
- But: Determining assurance cannot be done without specifying prior distributions for the unknown parameters
- Assurances can “easily” be determined using Bayesian clinical trial simulation

References

- O’Hagan A, Stevens JW. Bayesian assessment of sample size size for clinical trials of cost-efectiveness. Medical Decision Making 2001; 21(3): 219-230
- https://www.statsols.com/nquery/sample-size-and-power-calculation-procedures#bayes

Confidentiality Notice