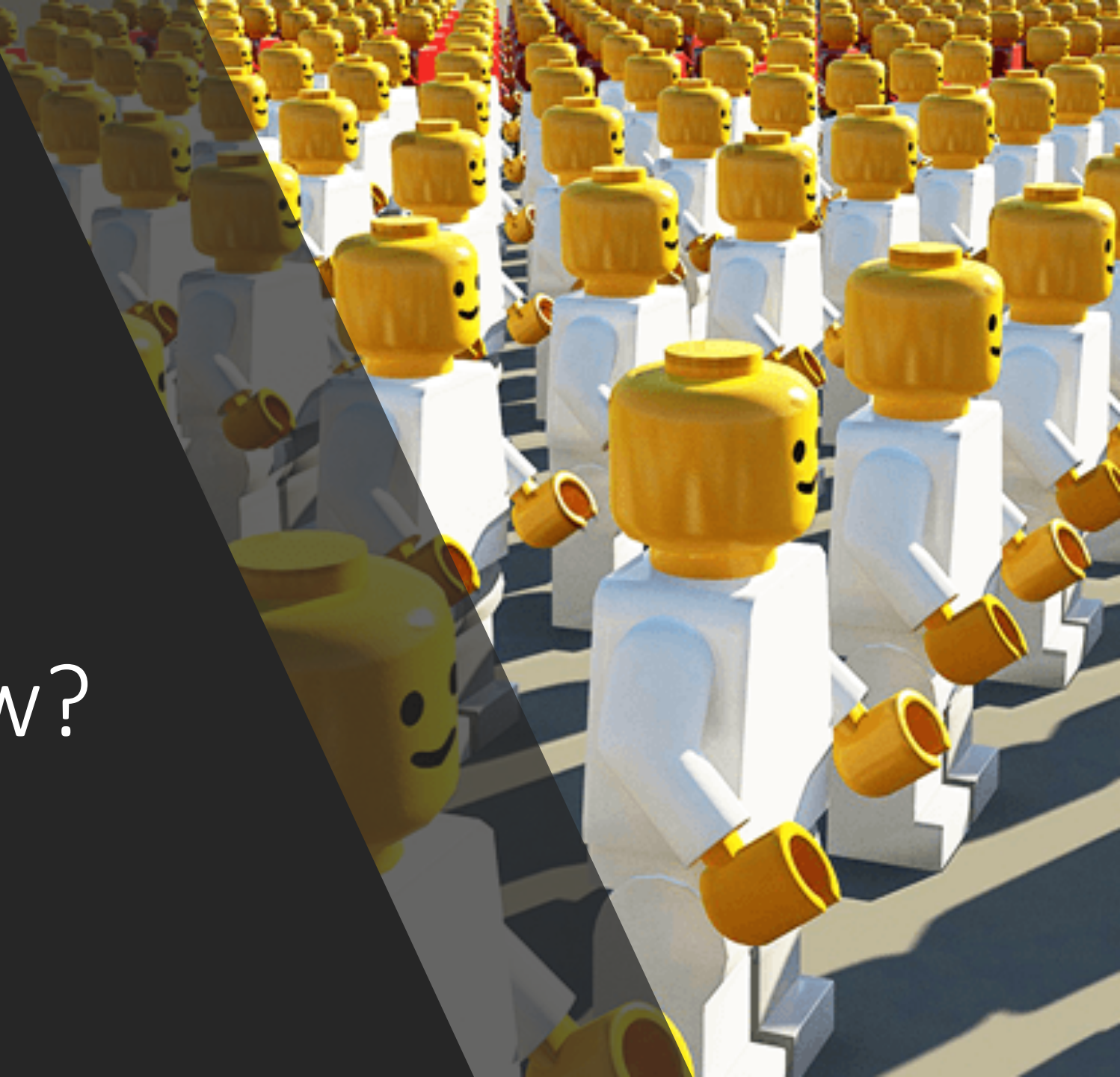
Graeme L. Hickey

University of Liverpool

Sample size determination: why, when, how?



@graemeleehickey

www.glhickey.com

graeme.hickey@liverpool.ac.uk

Why?

Scientific: might miss out on an important discovery (testing too few), or find a clinically irrelevant effect size (testing too many)

Ethical: might sacrifice subjects (testing too many) or unnecessarily expose too few when study success chance low (testing too few)

Economical: might waste money and time (testing too many) or have to repeat the experiment again (testing too few)

Also, generally required for study grant proposals

When?

- Should be determined in advance of the study
- For randomised control trials (RCTs), must be determined and specified in the study protocol before recruitment starts

What not to do

Use same sample size as another (possibly similar) study

Might have just gotten lucky

Base sample size on what is available

Extend study period, seek more money, pool study

Use a nice whole number and hope no one notices

Unless you want your paper rejected

Avoid calculating a sample size because you couldn't estimate the parameters needed

Do a pilot study or use approximate formulae, e.g. SD ≈ (max – min) / 4

Avoid calculating a sample size because you couldn't work one out

Speak to a statistician

Example

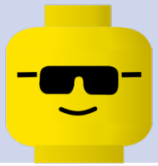
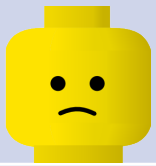
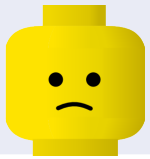
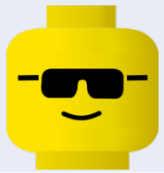
- A physician wants to set a study to compare a new antihypertensive drug relative to a placebo
- Participants are randomized into two treatment groups:
  - Group N: new drug
  - Group P: placebo
- The primary endpoint is taken as the mean reduction in systolic blood pressure (BP_{sys}) after four weeks

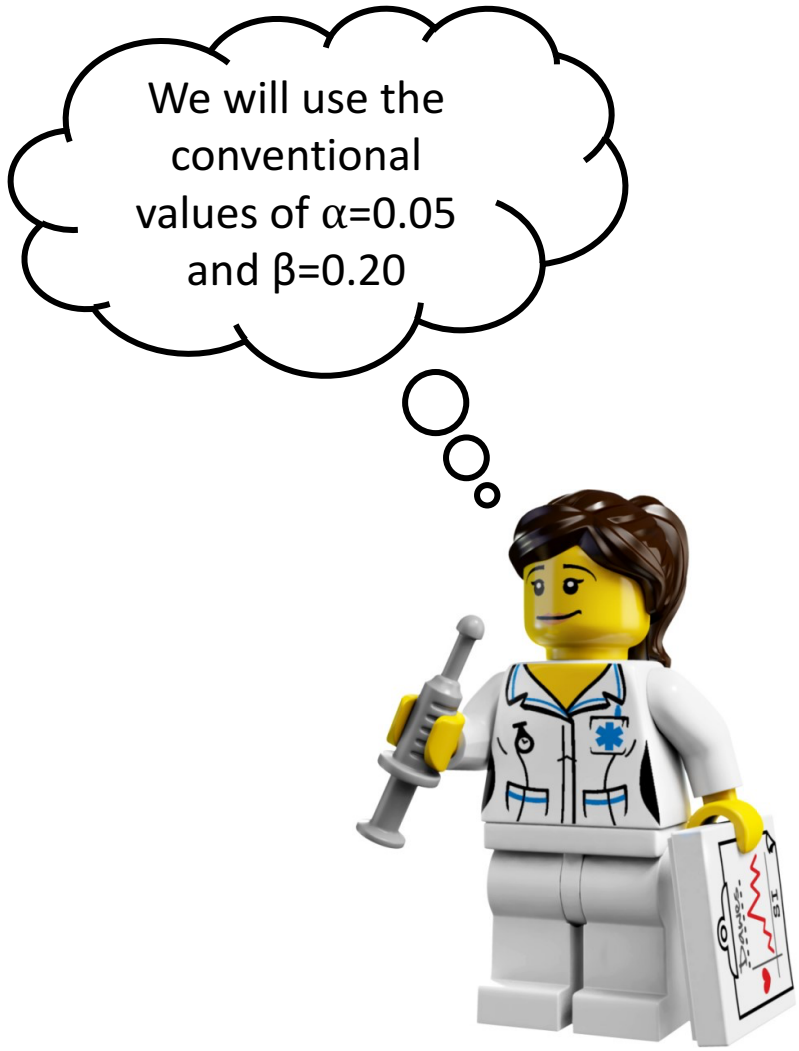
What do we need?

Item	Definition	Specified value
Type I error (α)		
Power ($1 - \beta$)		
Minimal clinically relevant difference		
Variation		

Errors

Hypothesis test

Truth	No evidence of a difference	Evidence of a difference
No difference	True Negative 	False positive **Type I error (α)** 
Difference	False negative **Type II error (β)** 	True Positive 



What do we need?

Item	Definition	Specified value
Type I error (α)	The probability of falsely rejecting H_0 (false positive rate)	0.05
Power ($1 - \beta$)	The probability of correctly rejecting H_0 (true positive rate)	0.80
Minimal clinically relevant difference		
Variation		

Minimal clinically relevant difference

- Minimal difference between the studied groups that the investigator wishes to detect
- Referred to as minimal clinically relevant difference (MCRD) – different from statistical significance
- MCRD should be biologically plausible
- Sample size $\propto$ $MCRD^{-2}$
  - E.g. if *n*=100 required to detect MCRD = 1, then *n*=400 required to detect MCRD = 0.5
- **Note**: some software / formula define the 'effect size' as the standardized effect size = MCRD / σ

Where to get MCRD or variation values

- Biological / medical expertise
- Review the literature
- Pilot studies

- If unsure, get a the range of values and explore using sensitivity analyses

Example: continued

- From previous studies, the mean BP_{sys} of hypertensive patients is 145 mmHg (SD = 5 mmHg)
- Histograms also suggest that the distribution of BP is normally distributed in the population
- An expert says the new drug would need to lower BP_{sys} by 5 mmHg for it to be clinically significant, otherwise the side effects outweigh the benefit
- He assumes the standard deviation of BP_{sys} will be the same in the treatment group

What do we need?

Item	Definition	Specified value
Type I error (α)	The probability of falsely rejecting H_0 (false positive rate)	0.05
Power ($1 - \beta$)	The probability of correctly rejecting H_0 (true positive rate)	0.80
Minimal clinically relevant difference	The smallest (biologically plausible) difference in the outcome that is clinically relevant	5 mmHg
Variation	Variability in the outcome (SD for continuous outcomes)	5 mmHg

Sample size formula*

$$n \approx 2\frac{\left[Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}\right]^2 \sigma^2}{(\mu_1 - \mu_0)^2}$$

- $\mu_1 - \mu_0$ is the MCRD
- Z_p is the quantile from a standard normal distribution
- σ is the common standard deviation

*based on a two-sided test assuming σ is known

Sample size calculation

The friendly medical statistician

$$n \approx 2\frac{[1.96 + 0.84]^2 5^2}{5^2}$$

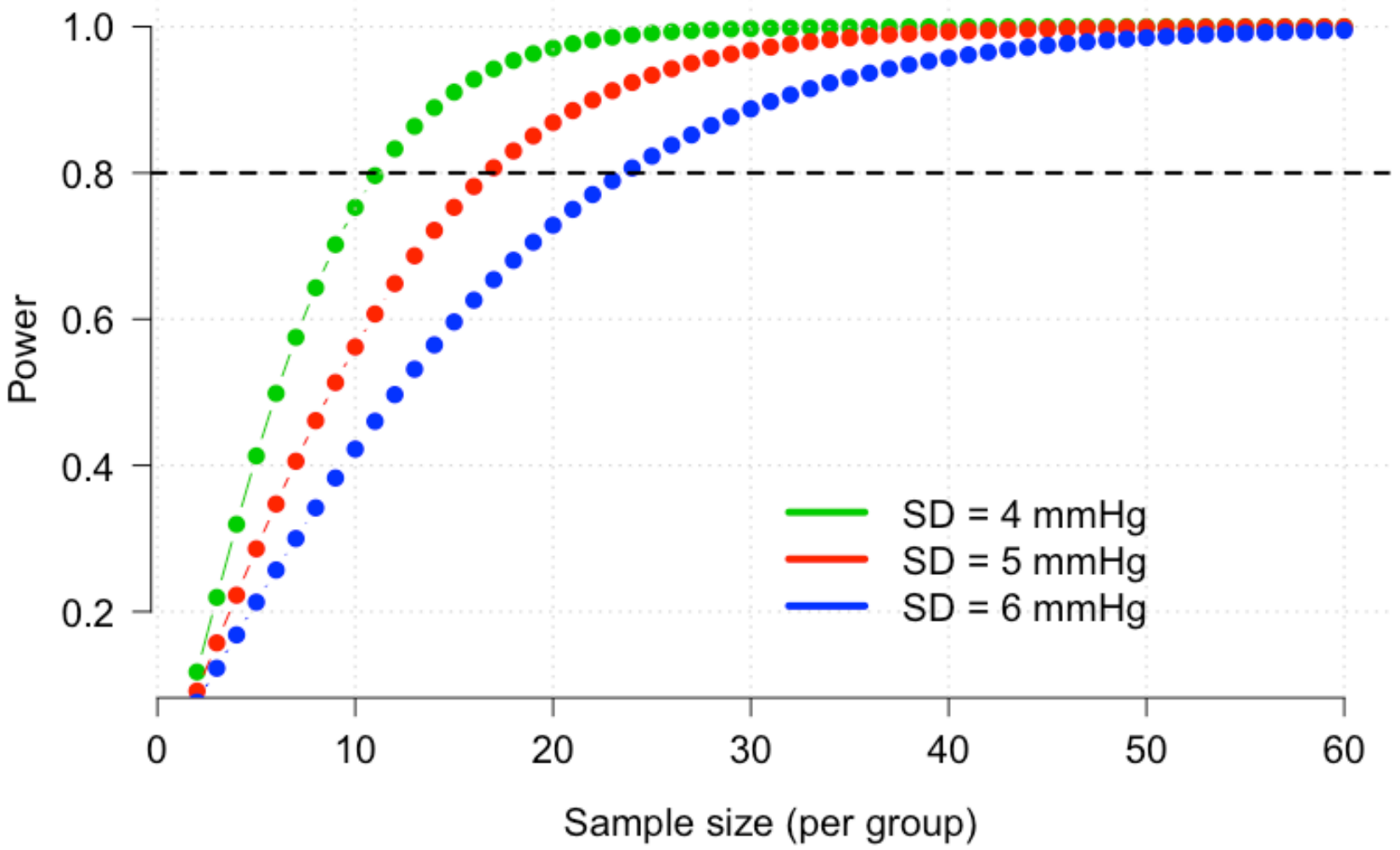
$$= 2\frac{[1.96 + 0.84]^2 5^2}{5^2} = 15.7$$

Therefore we need 16 patients per treatment group

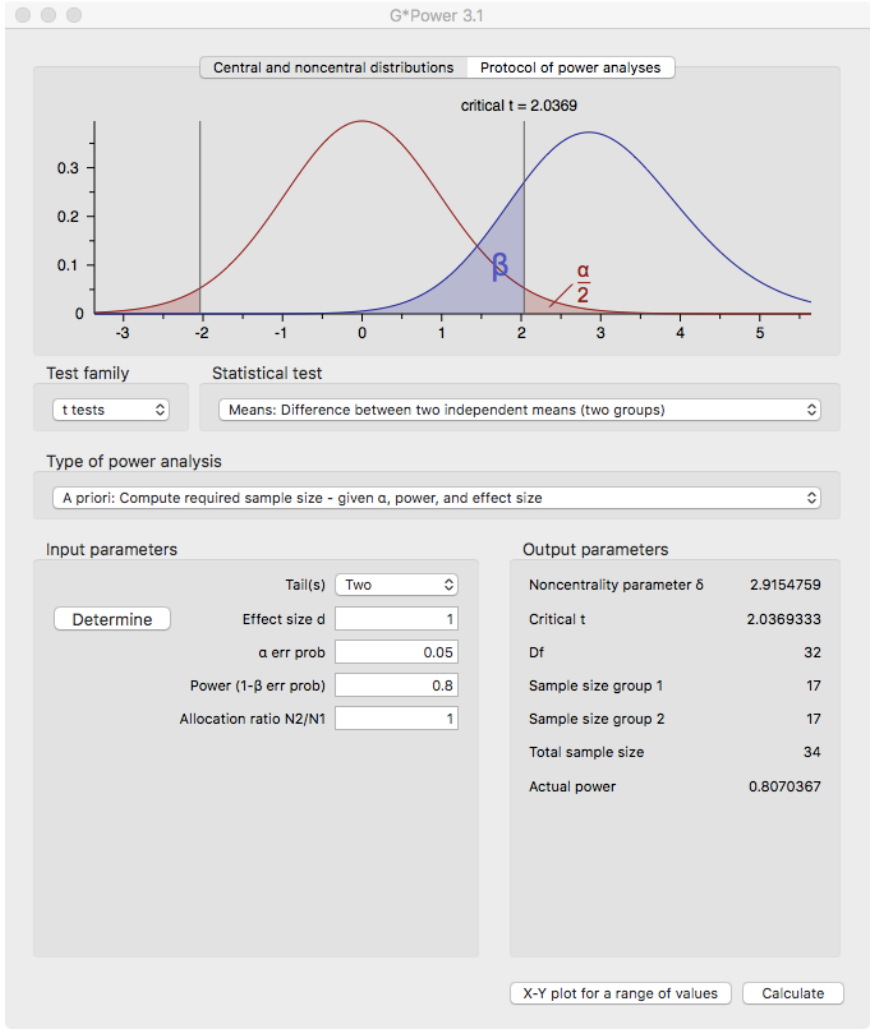
NB: we always round up, never down

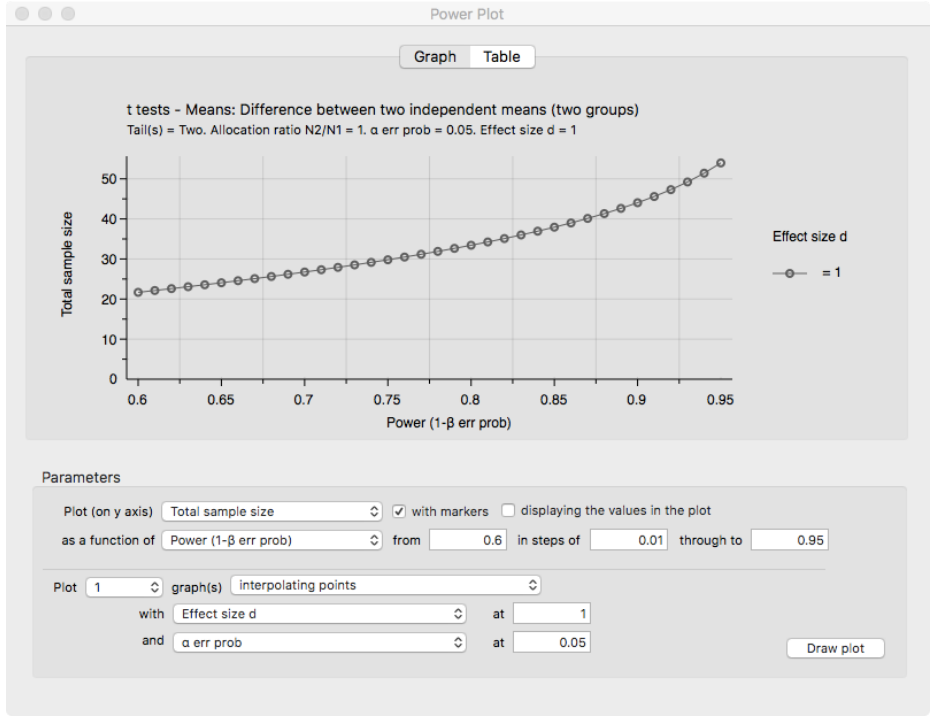
Sensitivity analyses

- Sample size sensitive to changes in α, β, MCRD, σ
- Generally a good idea to consider sensitivity of calculation to parameter choices
- If unsure, generally choose the largest sample size



Sample size calculation software





- Standalone tools: G*Power (http://www.gpower.hhu.de/)
- Many statistics software packages have built-in functions
- Lots of web-calculators available
- Lots of formulae published in (bio)statistics papers

Practical limitations

- What if the study duration is limited; the disease rare; financial resources stretched; etc.?
- Calculate the power from the maximum sample size possible (reverse calculation)
- Possible solutions:
  - change outcome (e.g. composite)
  - use as an argument for more funding
  - don’t perform the study
  - reduce variation, e.g. change scope of study
  - pool resources with other centres

Estimation problems

- Study objective may be to estimate a parameter (e.g. a prevalence) rather than perform a hypothesis test
- Sample size, *n*, chosen to control the width of the confidence interval (CI)
- E.g. if a prevalence, the approximate 95% CI is given by

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Margin of error (MOE)

where $\hat{p}$ is the estimated proportion

Example

- David and Boris want to estimate how support among cardiothoracic surgeons for the UK to leave the EU
- They want the MOE to be <3%
- SE maximized when $\hat{p} = 0.5$, so need $\frac{1.96}{2\sqrt{n}} < 0.03$
- So need to (randomly) poll *n* = 1068 members

Drop-outs / missing data

- Sample size calculation is for the number of subjects providing data
- Drop-outs / missing data are generally inevitable

- If we anticipate losing *x*% of subjects to drop-out / missing data, then inflate the calculated sample size, *n*, to be:

$$n^{\star} = \frac{n}{\left(1 - \frac{x}{100}\right)}$$

Sample size formula and software available for other…

- Effects:
  - Comparing two proportions
  - Hazard ratios
  - Odds ratios
  - …
- Study designs:
  - Cluster RCTs
  - Cross-over studies
  - Repeated measures (ANCOVA)
  - …
- Hypotheses:
  - Non-inferiority
  - Superiority
  - …

Observational studies

Issues

- Study design features:
  - Non-randomized $\Rightarrow$ bias
  - Missing data
  - Assignment proportions unbalanced
- Far fewer ‘closed-form’ formulae

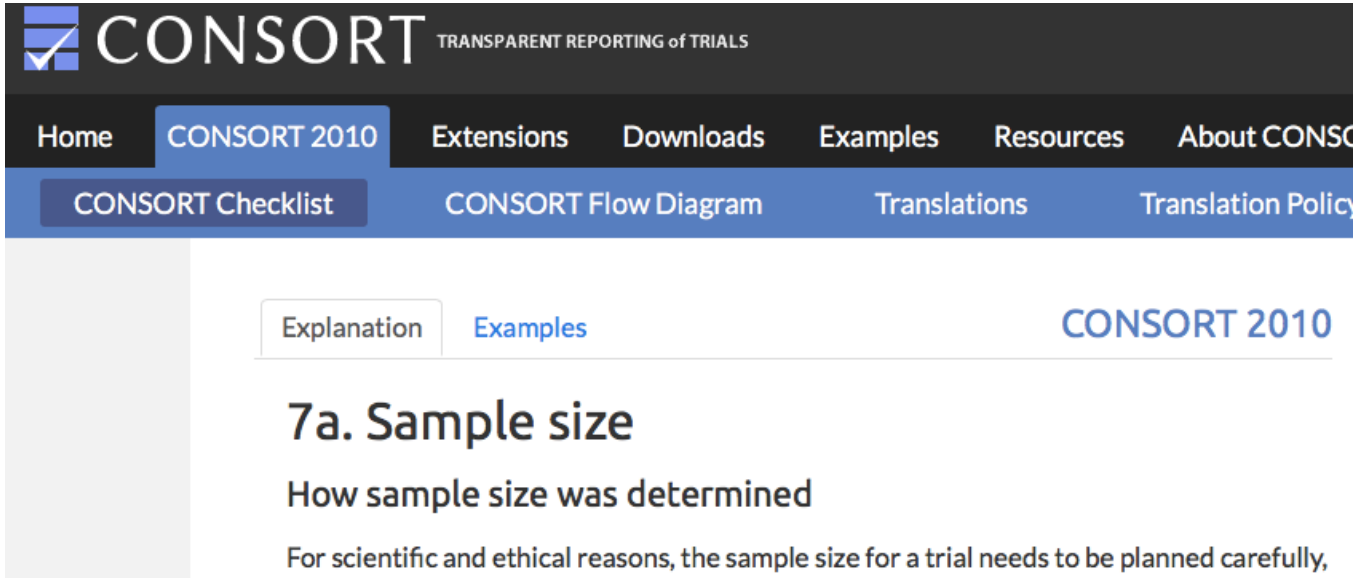
How to approach (depending on study objective)

- Start from assuming randomization as a reference
- Correction factors (e.g. [1,2])
- Inflate sample size for PSM to account for potential unmatched subjects
- …

[1] Hsieh FY *et al*. *Stat Med*. 1998; 17: 1623–34.
[2] Lipsitz SR & Parzen M. *The Statistician*. 1995; 1: 81-90.

Reporting

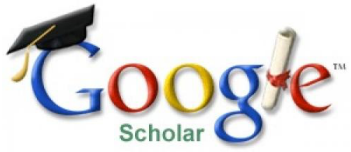
- Six high-impact journals in 2005-06*:
  - 5% reported no calculation details
  - 43% did not report all required parameters
  - Similar reporting inadequacies in papers submitted to EJCTS/ICVTS
- Information provided should (in most cases) allow the statistical reviewer to reproduce the calculation
- CONSORT Statement requirement



* Charles et al. *BMJ* 2009;338:b1732

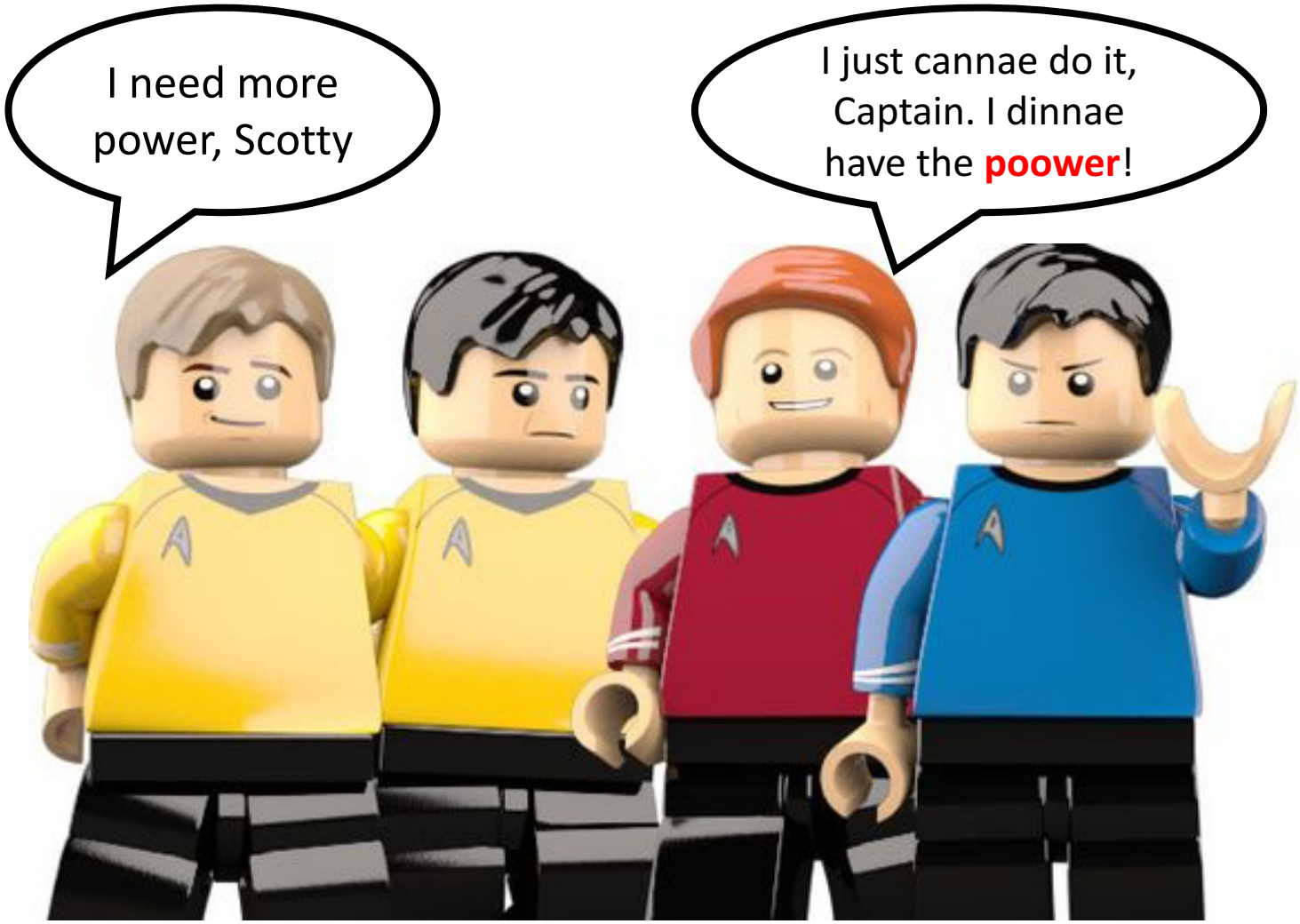
Final comments

- All sample size formulae depend on significance, power, MCRD, variability (+ possible additional assumptions / parameters, e.g. number of events, correlations, …) no matter how complex

- Lots of published formula (search 

), books, software, and of course… statisticians – need to find the one right for your study

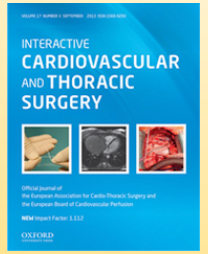
- A *post hoc* power calculation is worthless
  - Instead report effect size + 95% CI

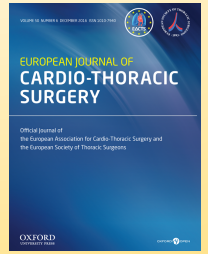


Thanks for listening

Any questions?

Statistical Primer article to be published soon!





Slides available (shortly) from: www.glhickey.com