How big is your trial?

Tony Brady

The first clinical trial?

"I took 12 patients in the scurvy aboard the Salisbury at sea. Their cases were as similar as I could have them.

2 were ordered a quart of cider a day
2 others 25 gutts of elixir vitriol
2 others 2 spoonfulls of vinegar
2 others a course of sea water
2 others the bigness of a nutmeg
2 others 2 oranges and 1 lemon a day

The consequence was the most sudden and visible good perceived from the use of oranges and lemons."

James Lind, 1753

Different perspectives

- Scientific
  - How many patients are needed:
    - to get firm evidence of a treatment difference? (power-based)
    - to estimate treatment difference well? (precision-based)
- Sponsor
  - How many patients do we need to carry out a trial at minimum time, cost and effort?
- Ethics committee
  - How soon can the trial be stopped to avoid some patients getting inferior treatment?

- In general want smallest trial possible

BUT
- Should be large enough to have a realistic chance of detecting the treatment effect
- Should be large enough to convince clinicians that it's results are reliable
- Small trials with little hope of detecting a realistic treatment effect are unethical
- Unfortunately undersized trials are very common

Scientific approach to trial size

- What is the main purpose of the trial?
  - Treatment A is better than treatment B
  - Treatment A is no better or worse than treatment B (equivalence trial)
- What is the principal measure of patient outcome?
  - Continuous
  - Categorical
  - Survival time

- How will the data be analysed to detect a treatment difference?
  - Two-sample t-test of difference in means
  - 95% CI for difference in proportion
  - survival analysis

Power-based (A is better than B):
- How small a treatment effect is it important to 'detect' and with 'what degree of certainty'?
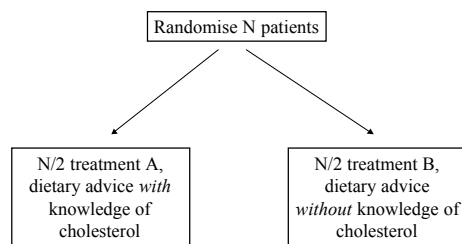
Degree of certainty

Error	Description	Pr(error)	aka
Type I	You think the treatments are different but they're not really	α	Significance level
Type II	You think the treatments are the same but in fact they're different	β	Power $= 1 - \beta$

Size of treatment effect

- Smallest effect which is clinically important to detect
- Will depend on treatment - highly invasive or toxic treatments may need to show larger effect to be useful
- Most treatment effects are small or moderate in size

Example: cholesterol reduction

Randomise N patients

N/2 treatment A, dietary advice *with* knowledge of cholesterol

N/2 treatment B, dietary advice *without* knowledge of cholesterol

Q: by how much does treatment A reduce cholesterol more than treatment B?

Example

- Main outcome: difference between groups in mean cholesterol at end of study
  $(d = \text{mean}_B - \text{mean}_A)$
- Important to detect difference, d of 0.4 mmol/l
- Other studies indicate SD(cholesterol)=2.0 mmol/l
- Standardised difference = d/SD = 0.2
- Set α=0.05 and β=0.1
- Lookup N from nomogram

Nomogram for total sample size, N
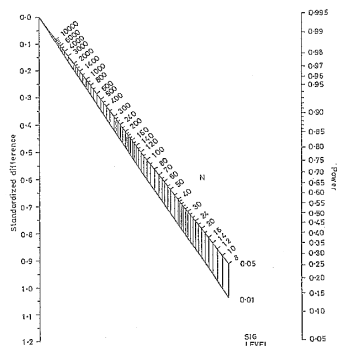
Example

- d=0.4, SD=2.0, d/SD=0.2, α=0.05 and β=0.1
- Nomogram $\Rightarrow$ N = 1000
- 500 to treatment A, 500 to treatment B
- 'We need 1000 patients to have a 90% chance of detecting, as statistically significant at the 5% level, a true treatment difference of 0.4 mmol/l'

Paired design

- Reduce sample size by measuring cholesterol once at beginning and once at end of trial
- Look at within-person differences
- SD(individual differences) = 1.28
- Standardised difference d/SD = 0.4/1.28 = 0.31 , α=0.05 and β=0.1
- N= 420 (210 per group)

Categorical data

- Main outcome is difference in proportions for dichotomous response
- p_A = proportion responding on treatment A
  p_B = proportion responding on treatment B
- Standardised difference = $(p_A - p_B)$/SD

$$SD = \sqrt{\bar{p}(1-\bar{p})} \quad \text{where } \bar{p} = \frac{p_A + p_B}{2}$$

Example: beta-blocker trial

- Trial to investigate whether preoperative beta-blockade prevents myocardial ischaemia
- Expect 20% of high risk patients to experience ischaemia
- Reduction to 15% on beta-blockers would be clinically important to detect
- d = (0.2 - 0.15) = 0.05
- SD=$\sqrt{(0.175 \times 0.825)}$ = 0.38
- Standardised difference, d/SD = 0.13
- α=0.05, β=0.2
- From nomogram, N=1800 (900 per group)

Possible alternatives

- p_A=20%, p_B=18%, α=0.01, β=0.05
  - So N=28000
  - Scientifically reasonable, but too costly

- p_A=20%, p_B=5%, α=0.05, β=0.5
  - So N=70
  - Over-optimistic, very unlikely to be conclusive

Survival data

- Usually base power calculation on proportion surviving at single time point (e.g. 5 years)
- Can base calculation on median survival time but this isn't always observed in a trial
- Power calculation based on logrank test should be used with care - may underestimate N

Compliance / withdrawals

- Lack of compliance or withdrawals from randomised treatment may lead to smaller observed treatment difference
- Larger sample is need to maintain power
- Let Q_A = proportion withdrawing or non-compliant from group A
- Let Q_B = proportion withdrawing or non-compliant from group B
- Suppose withdrawals and non-compliers from each group have a similar outcome to the other group

Compliance cont.

- Difference observed reduced by a factor $(1 - Q_A - Q_B)$
- Need to increase total sample size by a factor $1/(1 - Q_A - Q_B)^2$
- If $Q_A = Q_B = 10\%$, need to increase sample size by a factor of $1/0.8^2 = 1.56$

Equivalence trials

- Approach to sample size is usually <u>precision based</u> rather than power based
- Need to define the largest difference, e, in outcome between the 2 treatments we would accept as equivalence
- Then we need to recruit enough patients so that there is a $1-\beta$ chance that the confidence interval for the treatment difference excludes values larger than e

Equivalence trials

equivalence zone

Observed difference with 95% CI

e

0

Difference between treatments
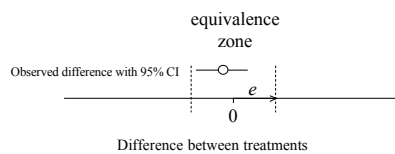
We want to make the trial big enough so that, if the two treatments are equivalent, there is a $1-\beta$ probability that the CI for the observed difference is within the equivalence zone

Example

- New beta-blocker drug being compared to well established beta-blocker for lowering blood pressure
- Prepared to accept disparity of ± 2 mmHg in mean blood pressure at end of trial, $e = 2$
- SD diastolic blood pressure is 14 mmHg
- $N = 52 \times (SD^2/e^2)$ for $\alpha=0.05$, $\beta=0.1$
- So $N = 2550$
- 'If the two beta-blocker drugs are truly equivalent a trial of 2550 patients will have a 90% chance of excluding a difference larger than 2 mmHg from the 95% CI'

Is the trial practicable?

- Assess patient accrual rate:
  - Initial estimates may often be optimistic (e.g. don't take into account exclusion criteria)
  - Accrual period should ideally be kept short
  - Consider a multi-centre trial to accrue more
- Relax the criteria, e.g:
  - Increase the difference $p_A - p_B$ to be detected
  - Increase the type II error
  - Increase the type I error

Not reasonable to make extreme changes, but there may be some room for compromise

If not practicable, abandon the trial

Pilot studies

- Check feasibility / organisation
- Check anticipated accrual rate
- Provide estimates of parameters needed for sample size calculations (e.g. SD, placebo response rate)
- Assess compliance / withdrawal rates

Monitor accrual
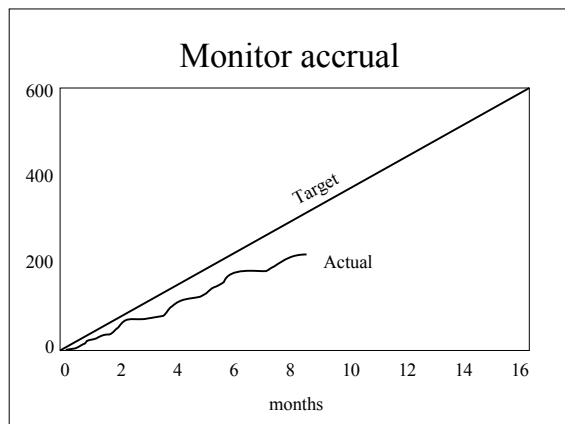
Practical issues

- Choice of p_A - p_B or mean_A - mean_B
- Compromise between scientific goals and feasible accrual
- Trialists often over-optimistic about availability of patients
- Abandon a trial if power is inadequate
- Trialists often underestimate the required size of trial
- Avoid having too many treatments
- Avoid being too restrictive about patient entry