Government of Western Australia
Child and Adolescent Health Service

Sample Size Calculations

Calculating simple sample sizes using PS software

2nd August 2024

Michael Dymock

Biostatistician

Telethon Kids Institute

Compassion Excellence Collaboration Accountability Equity Respect

TELETHON KIDS INSTITUTE Discover. Prevent. Cure.

Perth Children's Hospital Foundation

Acknowledgement of Country

The Child and Adolescent Health Service acknowledge Aboriginal people of the many traditional lands and language groups of Western Australia.
We acknowledge the wisdom of Aboriginal Elders both past and present and pay respect to Aboriginal communities of today.

CAHS Research Education Program

Research Skills Seminar Series

- Over 25 topics across the research process
  - 1h overview
  - Handouts are provided
- Recorded and uploaded
- Feedback
  - Back of handout
  - Emailed link
- Please hold questions to the end
  - Use provided microphone

Overview

Overview

- What is a sample size and why does it matter?
- A little theory
- Calculating sample sizes using the PS software
- Considerations for clinical trials
- Where can I get more statistical help?

What is sample size and why does it matter?

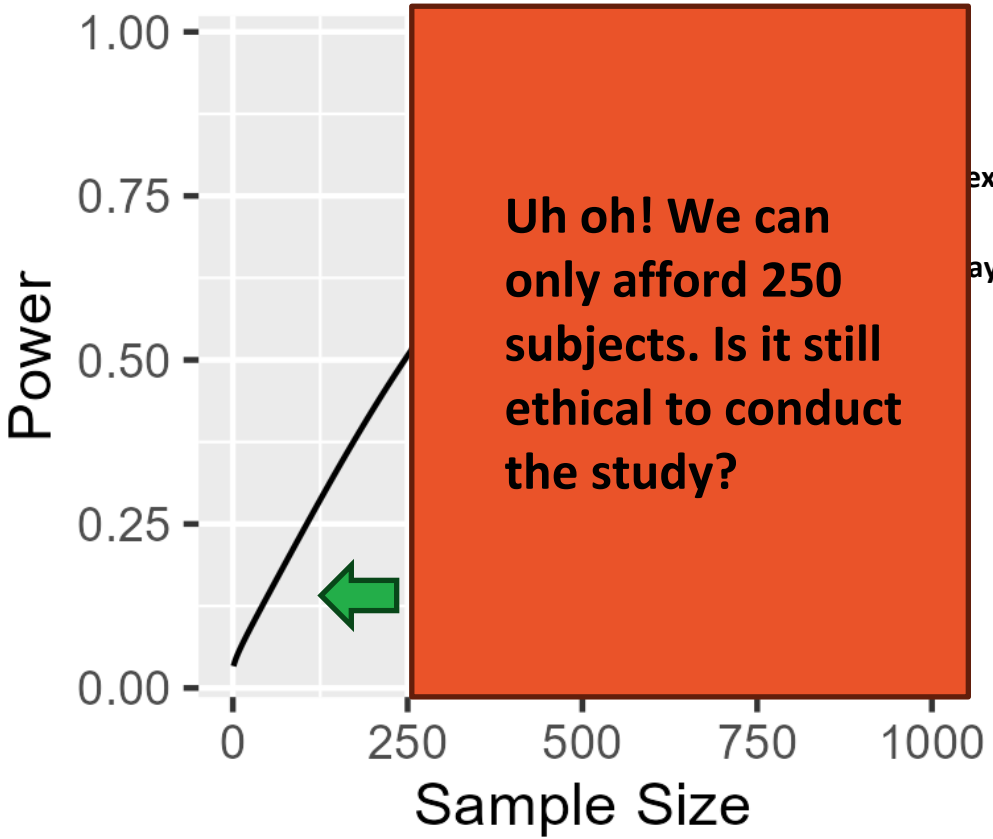
What is a sample size?

- To answer a research question effectively we should design a study carefully
- We need to decide how many subjects (participants, patients, etc.) to include and how many observations (measurements) to make on each subject
- The **sample size** is the total number of subjects*
    - *But we must consider the number of observations per subject: E.g., measuring blood pressure two times on three subjects may be considered a sample size of six

Why does everyone care so much?

- Before conducting a study, the research team must demonstrate that it will be feasible and ethical, and this requires estimating the sample size
  - Do we have the resources to conduct the study at this sample size?
  - Are we likely to be able to draw objective conclusions (i.e., power)?
  - What is the burden/risk to the subjects?
- Small studies are not always unethical (pilot studies, contribute to meta-analyses, low risk, etc.)
- Unfortunately, often the sample size is determined by the resources available

The sample size trade-off…



expose this many subjects to the study?

ay?

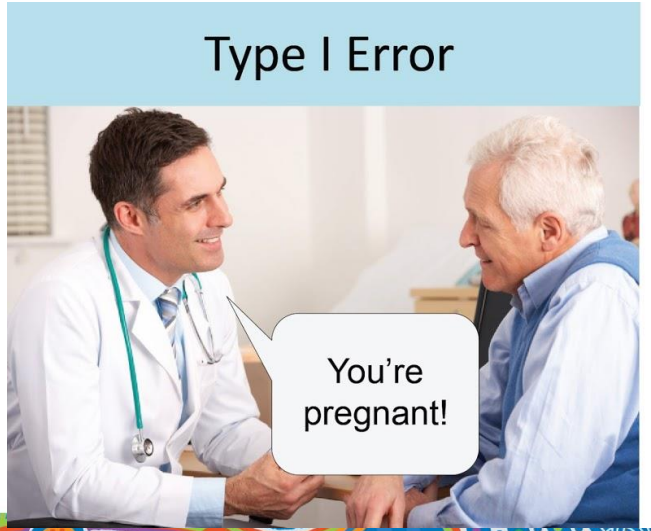
What does the statistician think?

- After contemplating the ethical considerations, in general, the larger the sample size the better
- A larger sample size means:
  - Reduced variability in our results (increased precision)
  - May be able to detect a smaller effect size
  - More likely to make objective conclusions

A little theory

Two errors, one study

- When answering a scientific question (e.g., does the treatment work?) you can be wrong in **two** different ways:
  - False positive: declaring the treatment **works** when it **doesn't**
  - False negative: declaring the treatment **doesn't work** when it **does**



You have the power!

- Probability of declaring the treatment/intervention effective (with assumptions)
- Need a (hopefully) clinically meaningful effect size
    - i.e., if the treatment has an effect size of X units, then I will declare the treatment effective Y% of the time in a series of hypothetical trials

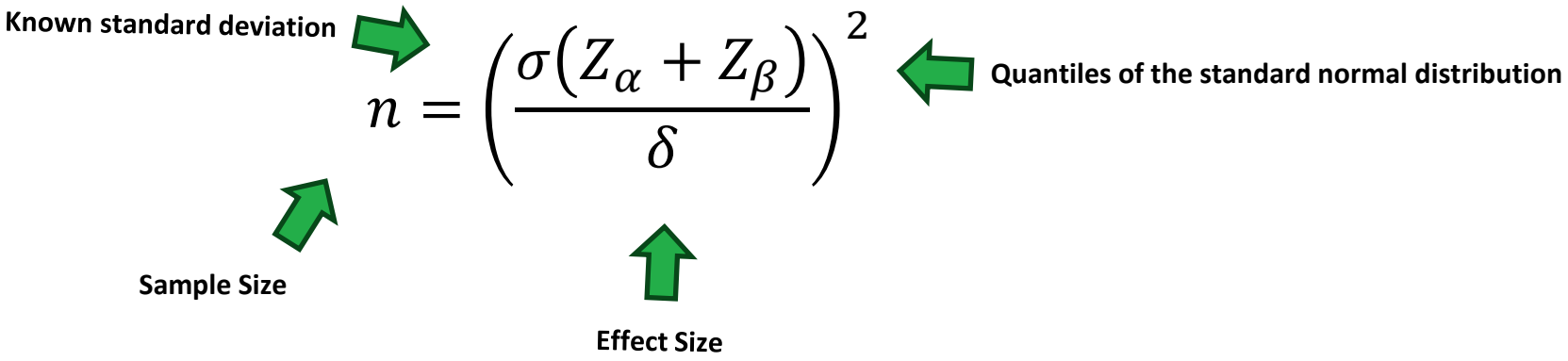
The generic problem

- We want to minimise the probability of making a Type I or Type II error
- Power = 1 – Type II error (want to maximise)
- We usually choose the Type I error and Type II errors and hold them constant and compute the required sample size
    - e.g., significance level = 5%, power = 80%

YOU CAN CHOOSE YOUR OWN VALUES

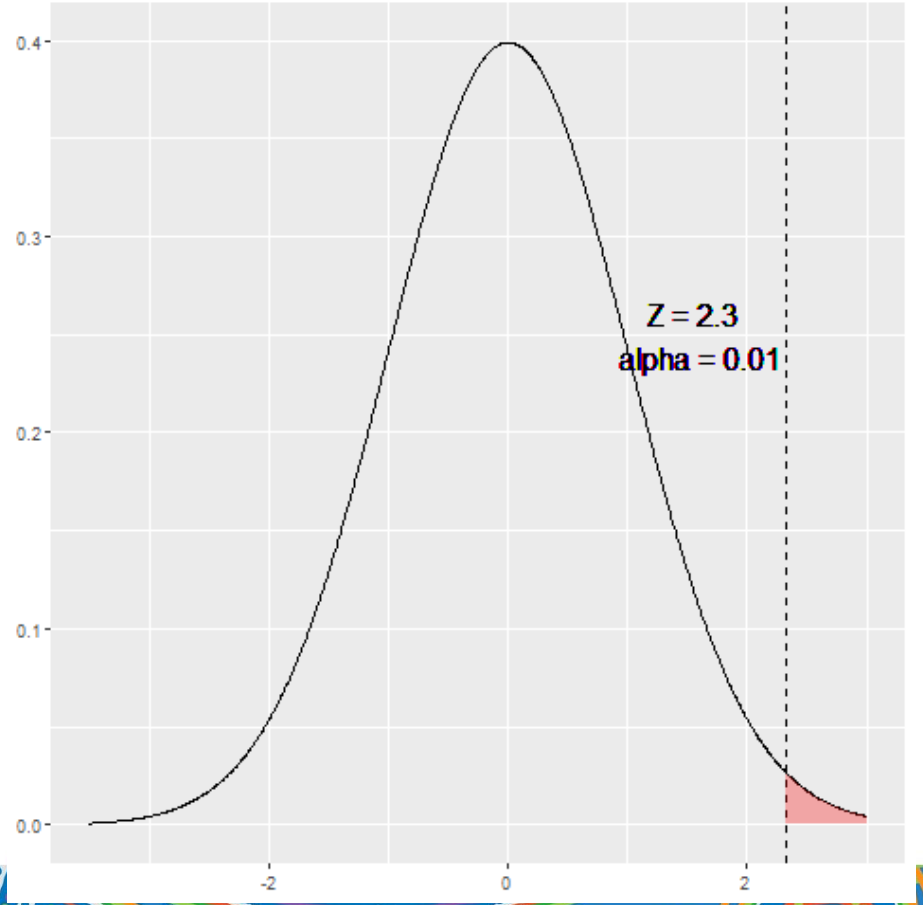
Calculating a sample size

- What sample size do I need so that Type I error = X and Type II error = Y?
- We use formulas! (Or simulations when it is too hard!)
- For example, one sample Z-test sample size formula:



A quick note on the standard normal distribution

- A **probability distribution** (something that allocates probabilities over a set of possible outcomes)
- Mean = 0 and Variance = 1
- Z_α is the point on the x-axis where there is α probability to the right (upper tail)



What factors will affect the sample size?

- Recall: $n = \left(\frac{\sigma(Z_\alpha + Z_\beta)}{\delta}\right)^2$
- Significance level (Type I error)
- Power (1 - Type II error)
- Data variability (standard deviation)
- Detectable effect size (delta)
- *Statistical method*

An example

- Recall: $n = \left(\frac{\sigma(Z_\alpha + Z_\beta)}{\delta}\right)^2$
- Suppose that we wanted to conduct a one sample Z-test on an outcome with known standard deviation of 3 units. Using a significance level of 5% we want 80% power to detect an effect size of 0.5 units.

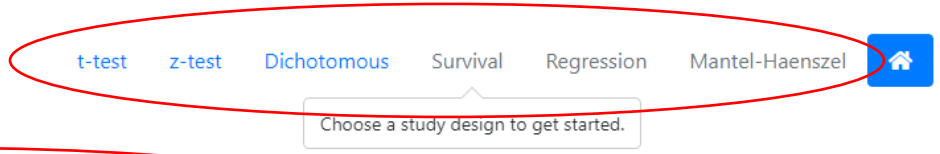
$$n = \left(\frac{3\times(Z_{0.05}+Z_{0.2})}{0.5}\right)^2 = \left(\frac{3\times(1.64+0.84)}{0.5}\right)^2 = 222.6 = 223$$

Calculating sample sizes using the PS software

Who needs formulas anyway…

- PS: Power and Sample Size Calculation v3.1.6, 2018
  - William D Dupont and Walton D Plummer, Jr.
  - http://biostat.app.vumc.org/wiki/Main/PowerSampleSize
- PS is an interactive program for performing power and sample size calculations for free
- Can be downloaded or used via a web browser (recommended)

Walkthrough



Department of Biostatistics
VANDERBILT UNIVERSITY

Build version: a57e8c3 (Feb 16, 2021)

PS is an interactive program for performing power and sample size calculations. It may be run as a web app at https://vbiostatps.app.vumc.org/ or downloaded for free. This version can be used for studies with dichotomous or continuous, response measures. An older version, which also handles other designs may be downloaded from http://biostat.app.vumc.org/wiki/Main/PowerSampleSize. Work on expanding the new version to handle all of the designs from the older version are in progress.

The alternative hypothesis of interest may be specified either in terms of differing means, or in terms of relative risks or odds ratios. Studies with dichotomous or continuous outcomes may involve either a matched or independent study design. The program can determine the sample size needed to detect a specified alternative hypothesis with the required power, the power with which a specific alternative hypothesis can be detected with a given sample size, or the specific alternative hypotheses that can be detected with a given power and sample size.

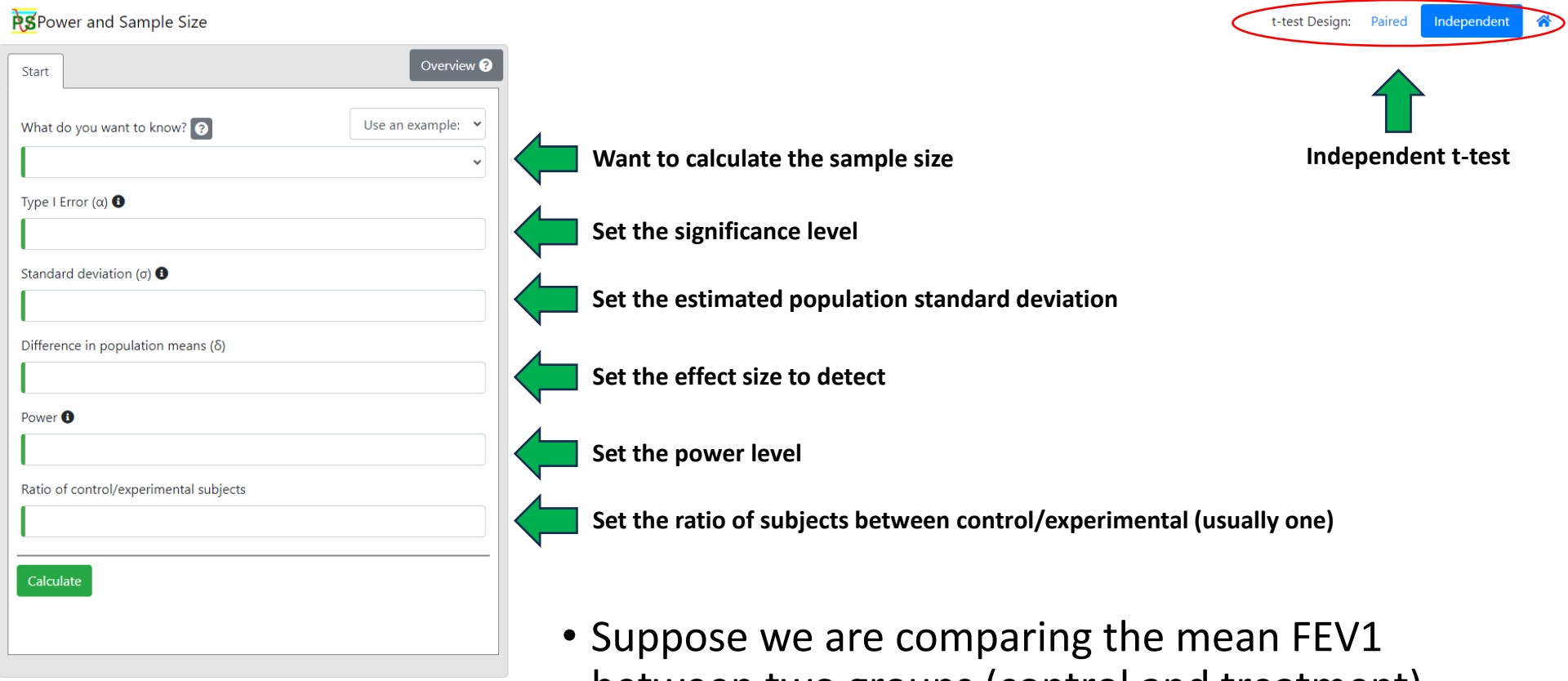
The PS program can produce graphs to explore the relationships between power, sample size and detectable alternative hypotheses. It is often helpful to hold one of these variables constant and plot the other two against each other. The program can generate graphs of sample size versus power for a specific alternative hypothesis, sample size versus detectable alternative hypotheses for a specified power, or power versus detectable alternative hypotheses for a specified sample size. Multiple curves can be plotted on a single graphic.

Choose the test (statistical method)

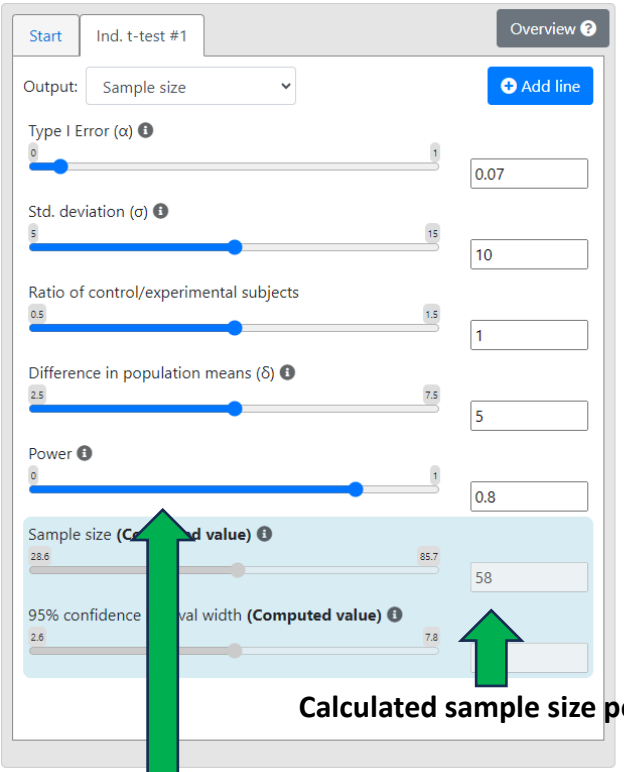
Survival, regression and Mantel-Haenszel under development

What study designs can it evaluate?

- PS can calculate the power and sample size for a range of study designs including those that require a:
  - t-test (continuous variable, two groups)
  - z-test (t-test with normality assumption)
  - binary analysis (odds ratios, matched case-control etc.)
  - survival analysis (time to event, e.g., remission, death)
  - linear regression (continuous variable, covariates)
  - Mantel-Haenszel test (2 x 2 tables, odds ratios etc.)

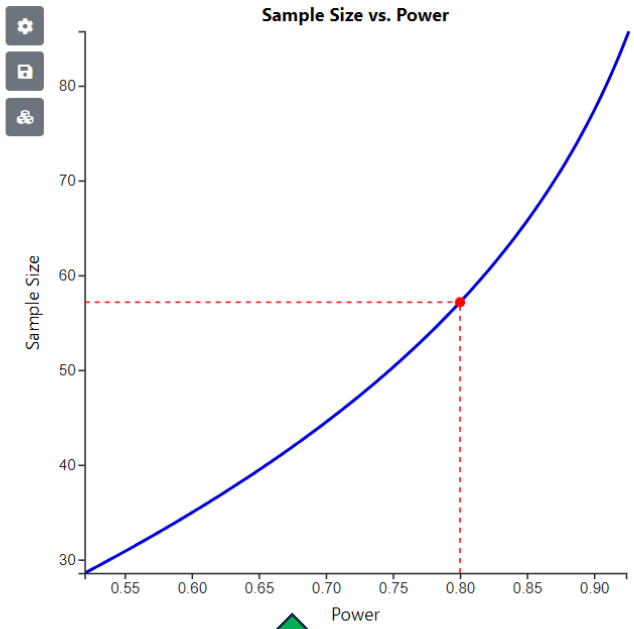


- Suppose we are comparing the mean FEV1 between two groups (control and treatment)

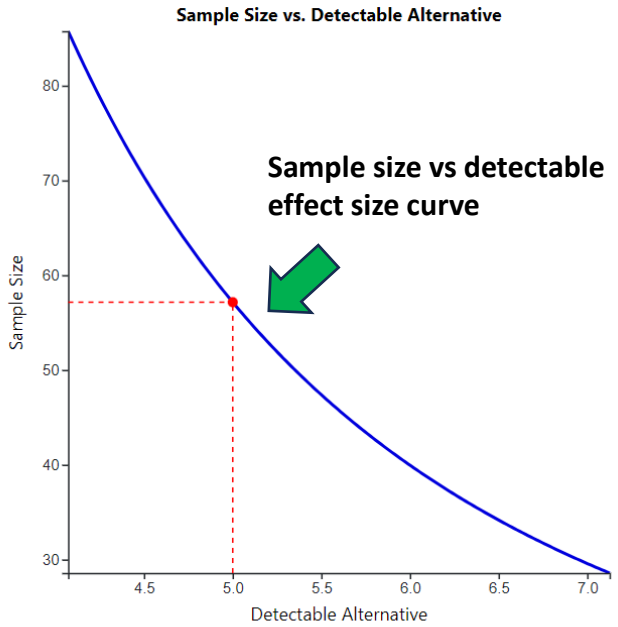


Can toggle inputs and watch the calculation change

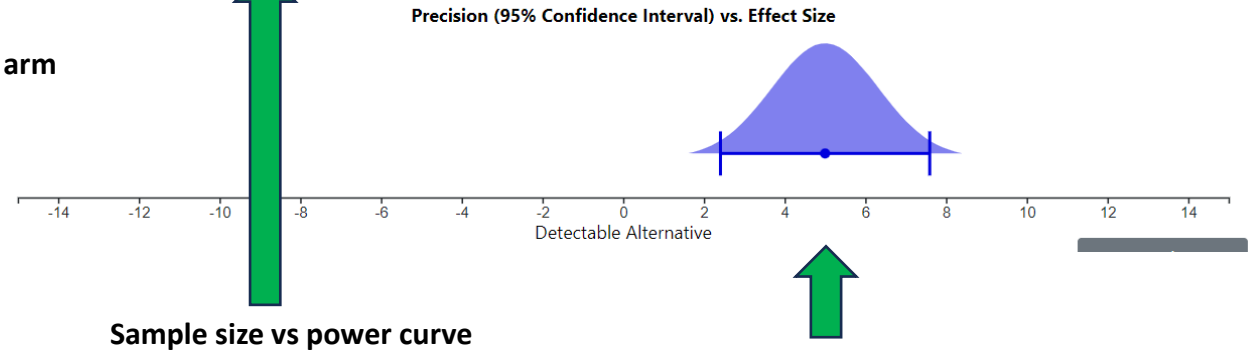
Calculated sample size per arm



Sample size vs power curve



Sample size vs detectable effect size curve



Confidence interval around effect size

- Suppose we are comparing a binary outcome between two groups (control and treatment)
  - Are the groups matched case-control?
  - What is the probability of exposure in the control group?
  - What is the correlation coefficient for exposure? (2 x 2 table)
  - What is the odds ratio of exposure?

Power and Sample Size

t-test | z-test | Dichotomous | Survival | Regression | Mantel-Haenszel

Start | Dichot #1

Overview

Matched / Case control

Output: Sample size

Add line

Type I Error (α): 0.03

Sample size **(Computed value)**: 774

Power: 0.9

Probability of exposure in controls (p_0): 0.3

Ratio of control/experimental subjects (m): 1

Correlation coefficient for exposure (φ): 0.7

Odds ratio of exposure (ψ): 2

Sample Size vs. Power

Calculated sample size per arm

Sample Size

Power

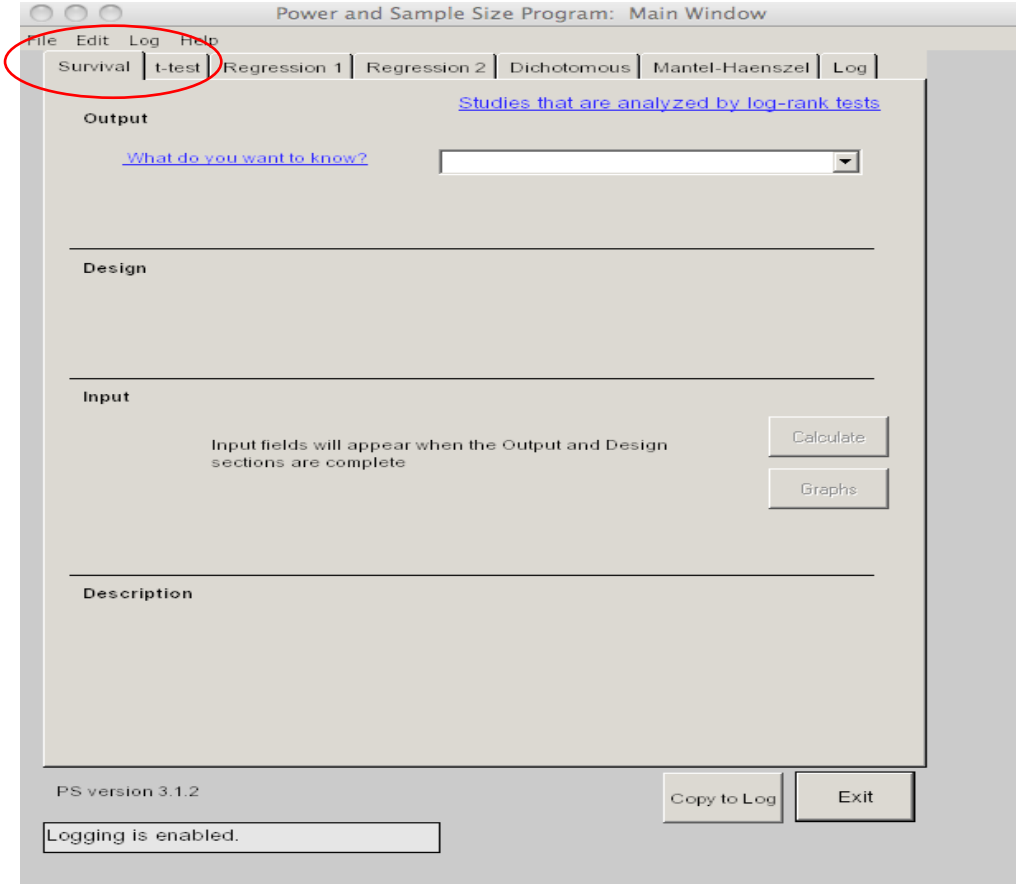
Sample Size vs. Detectable Alternative

Sample Size

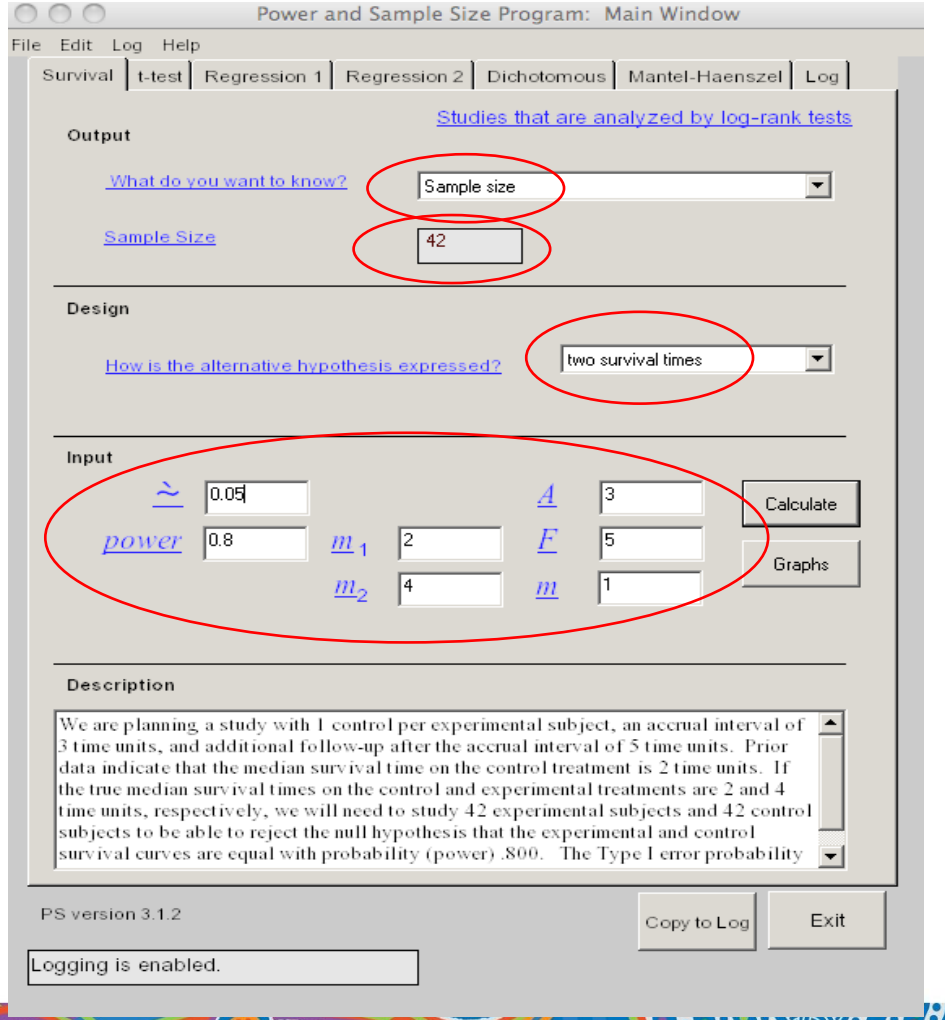
Detectable Alternative

Precision (95% Confidence Interval) vs. Effect Size

Odds ratio of exposure (ψ)

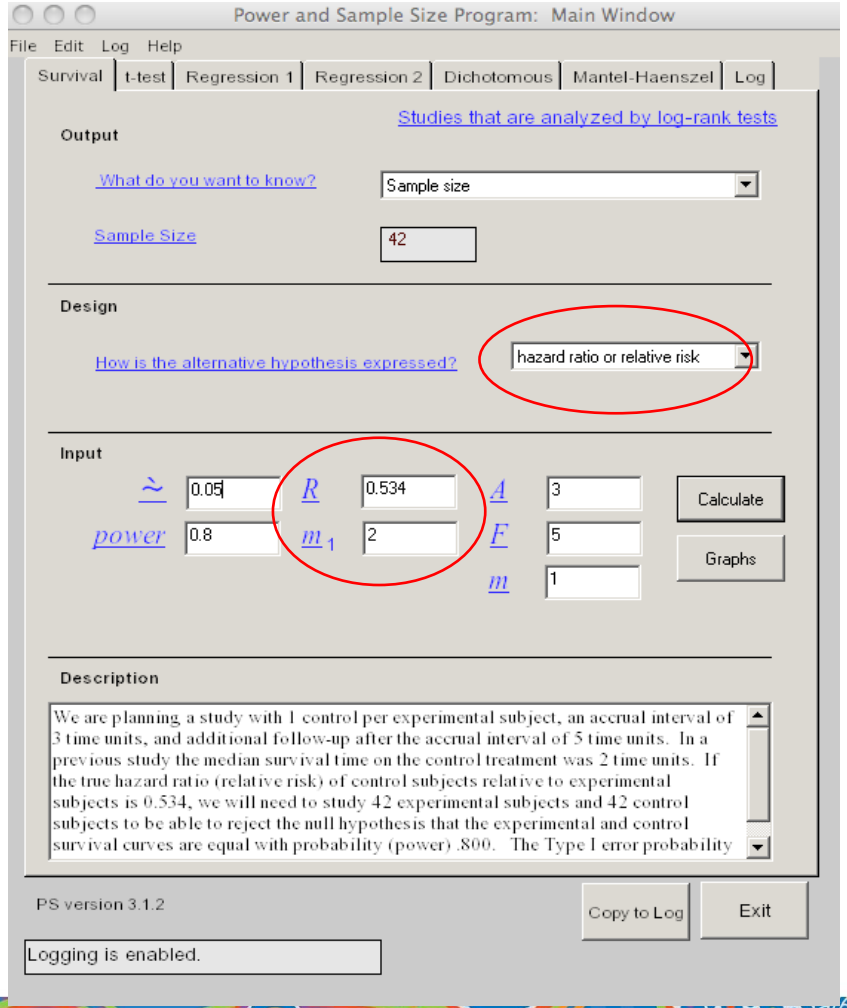


- Suppose we were comparing time to death between two groups (control and treatment)
- Each participant has two outcomes:
  - Dead/alive
  - Time to death (may be censored)



- significance = 0.05
- power = 0.8
- $\boldsymbol{m}$ = 1 (ratio of control/treatment)
- $\boldsymbol{m_1}$= 2 (median survival time - control)
- $\boldsymbol{m_2}$= 2 (median survival time - treatment)
- $\boldsymbol{A}$= 3 (accrual time)
- $\boldsymbol{F}$= 5 (follow up time)
- Make sure the time units are consistent between parameters!

- Sample size = 42 units per arm



- Alternatively specify alternative hypothesis using a hazard ratio / relative risk
- m_1= 2 (median survival time - control)
- R= 0.534 (hazard ratio)
- Sample size = 42 units per arm

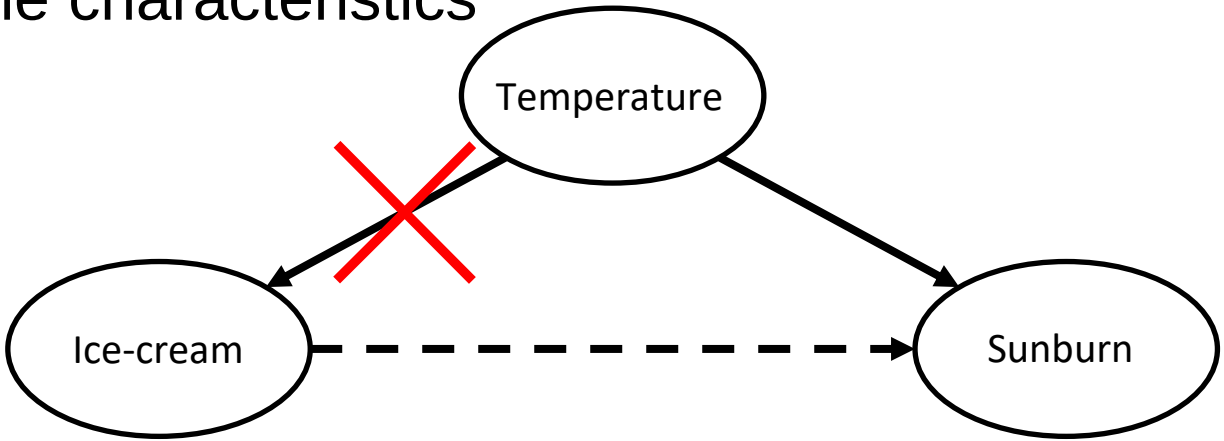
Considerations for clinical trials

Observational studies vs randomised designs

- In an **observational study** the **independent variable** (such as a treatment) is **not under the control** of the researcher
- In a **randomised design** (such as an RCT), the **independent variable** is **randomly** allocated to participants
- This “breaks” the links with any **uncontrolled** variables

Randomised controlled trials

- Randomised controlled trials (RCTs) are the gold standard in clinical research
- The goal of randomisation is to break the link between treatment assignment and confounders and to balance the baseline characteristics



How do I estimate the variability?

- Previous research work (pilot study?)
- Similar published clinical studies
- Animal studies (although humans tend to be more heterogeneous than lab animals)
- For studies restricted by resources you may want to consider a more homogeneous sample with lower variability (but beware of generalisability)

What about inclusion/exclusion criteria?

- Inclusion/exclusion criteria determines the generalisability of the results
- The standard deviation is directly affected as it is a function of how homogeneous your target population is
- Should the study be pragmatic?

Where can I get more help?

Where can I find a statistician?

Perth Children's Hospital:
Free advice through **Telethon Clinical Research Centre**

Telethon Kids Institute (consultancy service):
Biometrics@telethonkids.org.au

UWA (consultancy service):
consulting-cas@uwa.edu.au
The Centre for Applied Statistics, UWA, offers free advice to UWA postgraduate research students

More in handouts

Checklist for talking to a Statistician

- Clear hypothesis
- Proposed study design
- Primary endpoint & estimate of variability
- Clinically relevant effect size
- Estimate of feasible sample size based on budget or potential annual patient recruitment
- Important confounders & source of bias
- Similar publications or systematic reviews

How can I learn more about statistics?

- In the absence of large, randomised, well-controlled clinical trials to address every research question we all need to increase our **statistical literacy**

In person at Perth Children's Hospital:

Attend Research Skills Seminars.

In person at UWA:

The Centre for Applied Statistics provides short courses in statistics which are heavily discounted for students.

Joint Clinical-Statistical Supervision:

If one of your supervisors is a statistician, then you will have "unlimited" access to statistical knowledge/training.

How can I learn more about statistics?

Online: **Data Science Specialization**
Johns Hopkins University

FAQ: You can access the course for free via https://www.coursera.org/specializations/jhu-data-science#courses

This will allow you to explore the course, watch lectures, and participate in discussions for free. To be eligible to earn a certificate, you must either pay for enrolment or qualify for financial aid.

Links in your handouts

Questions?
Comments

Government of Western Australia
Child and Adolescent Health Service

TELETHON KIDS INSTITUTE
Discover. Prevent. Cure.

Perth Children's Hospital Foundation

Child Health Research Symposium

Empowering Futures: Advancing Child Health

4 – 7 November 2024 Perth Children's Hospital

Neonatology | Community Health | Mental Health | Perth Children's Hospital

Coming up next

9 Aug Rapid Critical Appraisal of Scientific Literature – Dr Natalie Strobel, ECU

16 Aug Media and Communications in Research – Peta O'Sullivan, CAHS

Register → **trybooking.com/eventlist/researcheducationprogram**

We love feedback

A survey is included in the back of your handout, or complete online
https://tinyurl.com/surveySampleSizeCalc

✉ ResearchEducationProgram@health.wa.gov.au cahs.health.wa.gov.au/ResearchEducationProgram

✉ ResearchEducationProgram@health.wa.gov.au
cahs.health.wa.gov.au/ResearchEducationProgram