MANCHESTER 1824

The University of Manchester

methods@manchester

7th Feb 2013

What are...

Power And Sample Size Calculations?

Sample Size, Power and Responsibility

Steve Roberts

Senior Lecturer in Medical Statistics

Centre for Biostatistics

Steve.roberts@manchester.ac.uk

What are Power Calculations?

- A method to estimate the chances of detecting the effect you are looking for before a study is conducted.
  - No role after the study has been done!
- A way of informing your choice amongst study design options
- A device to get you to talk to a statistician about your study before it is too late?

Why Bother?

When designing an experimental study...

- If you have too few participants, you may not be able to answer the question you are asking.
- If you have too many you waste resources, and expose participants to inconvenience, risk or inferior treatment unnecessarily.
- Grant awarding bodies and Ethics committees need to be convinced that you have made a sensible choice.

Take home message

- There is no right size for any study!
- You need to make a judgement, and convince others you are right.
- Balance benefits (likelihood of finding the answer) against costs (financial, resources, participant pain, ...)
- When appropriate, power calculations inform the arguments and put them on a rational basis.

Outline

- Why Bother?
- What are power calculations?
- What are power calculations for really?
- Examples
- Resources

What are Power calculations

The text-book bit

Hypothesis Testing

- We accept a (usually 5%) false positive rate – Type I error (α)
  - the significance level used in hypothesis testing or the use of 95% confidence intervals to describe the range of effect sizes consistent with the data.
- There is also a false negative rate – the probability of missing an effect that is really present – Type II error (β)

Hypothesis Testing

Statistical significance (P-value, size of confidence intervals) depends on:

- The size of the effect
- The variability in effect size
- The significance level (usually 0.05)
- The statistical test and study design
- The number of participants/samples completing the study

Power

There is always a chance that the real effect will be masked by the inherent variability in the study.

Statistical power is how we quantify this:

- The probability of detecting the effect if it is really there and if it is of the size assumed.
- Usually ask for 80-90%
- Power = 1-β, where β is the probability of a type II error (false negative)

Power and Sample Size

- This power depends on
  - Size of effect
  - Variability in effect size
  - Significance level (usually 0.05)
  - Statistical test and study design
  - Numbers completing the study

- These are not all known precisely before you undertake the study!

Why not?

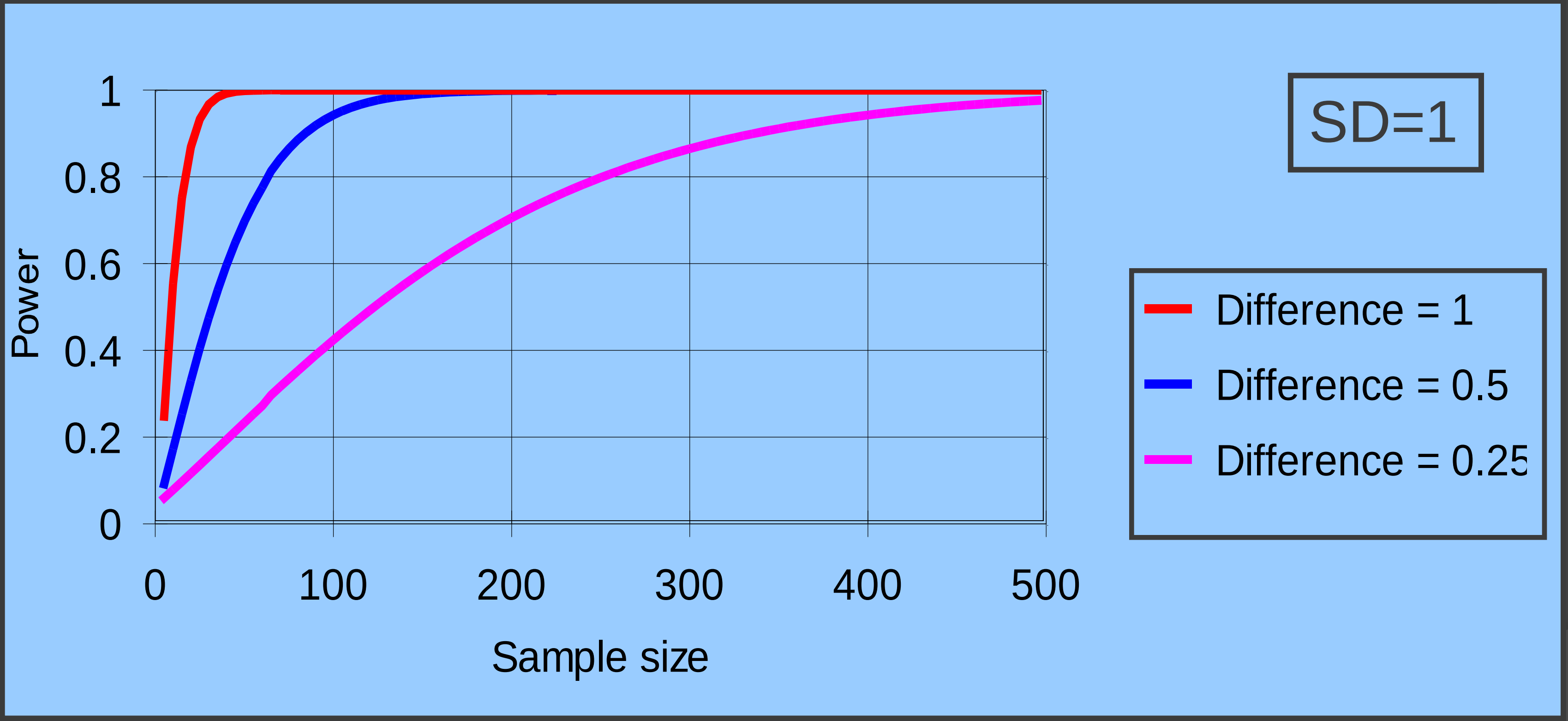
- There is no clear hypothesis
  - eg prevalence study, descriptive studies
- There is no prior data to enable a power calculation
  - Pilot studies
- There is no choice over sample size
  - Whole population studies

Power calculations are not compulsory...

- Just because there is a box on the form doesn’t mean you have to make up something inappropriate
- But you do have to say why you are studying the number you are – but be honest!

Power v sample size

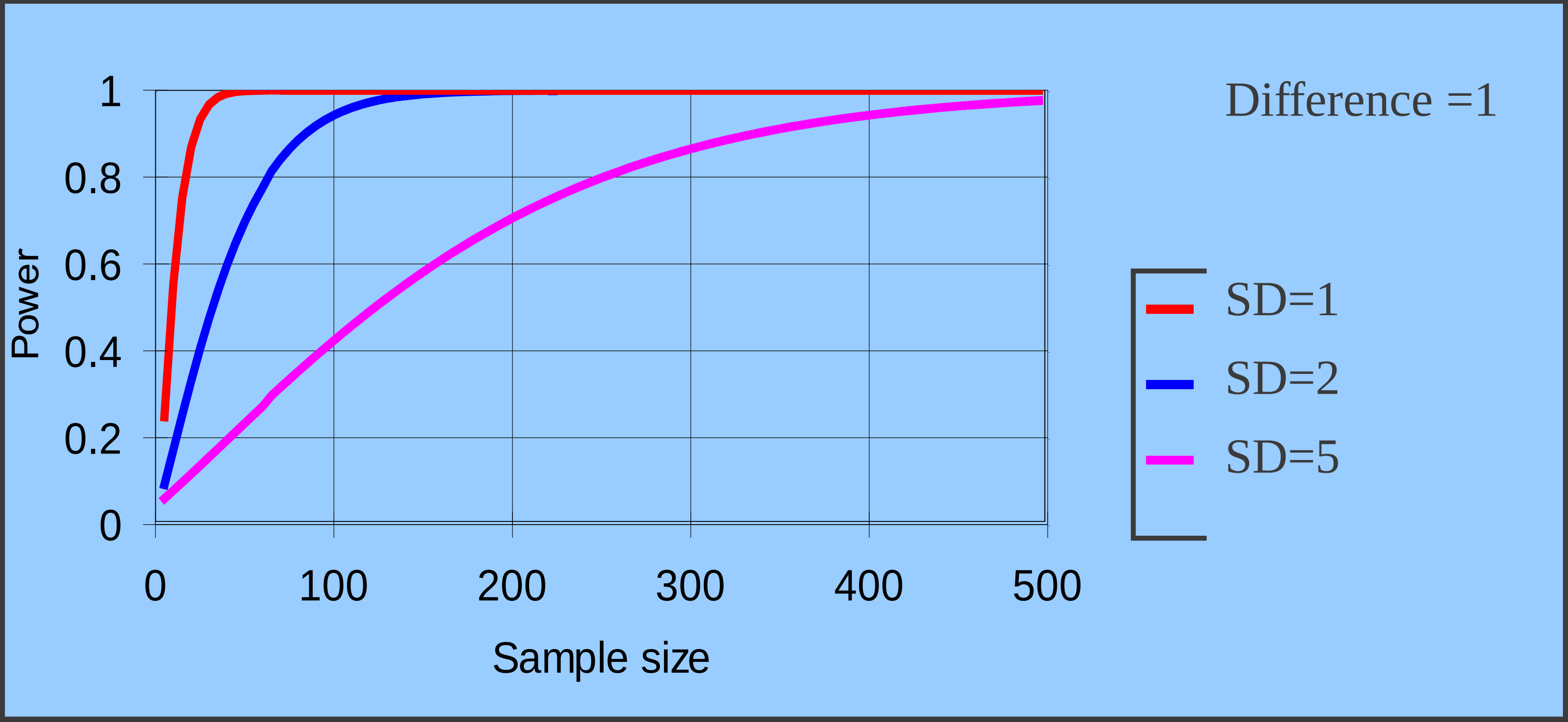
(Numerical Endpoint)



Two groups: Unpaired t-test

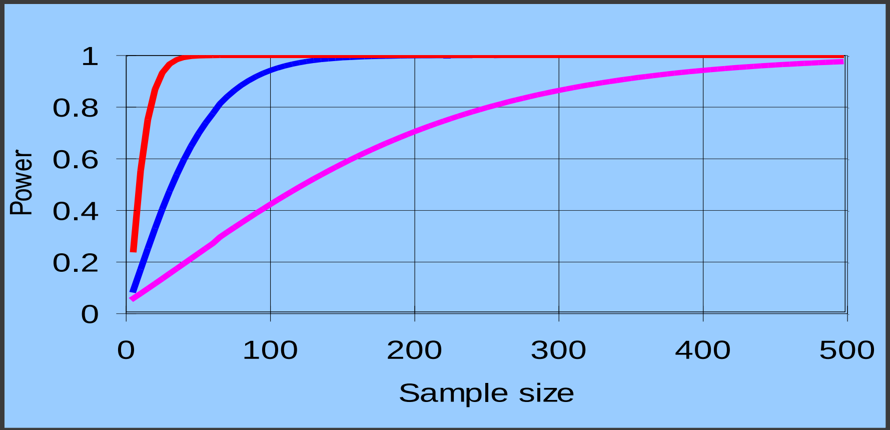
Power v sample size

(Numerical Endpoint)



Two groups: Unpaired t-test

Small differences in numbers make very little difference!



- Halve the effect size and you need 4 times the sample size
- Flat curves -> diminishing returns

Power and Sample Size

- To compute power we need to know
  - Study design, significance level and statistical test to be used
  - Variability in effect size
  - The effect size the study is planned to detect
  - Numbers of participants

Choosing the sample size
(Text book version)

- Identify the endpoint
- Identify the main comparison
- Estimate the variance of the endpoint
- Decide on the effect size
- Decide on the power you need
- Calculate the sample size
- Consider the practicalities
- And calculate

What Design and Analysis?

- Some designs are more efficient than others
- Some statistical tests are more powerful than others (e.g. matched designs)
- The method of analysis follows from, and is part of, the design
- *If your design is not simple or the analysis is complex you need to consult a statistician early in the design process*

What Variance?

- Pilot data or other published studies.
  - Look for other uses of the measurement tool on similar populations of participants.
- May need to do a need a pilot study to get the information to power a large definitive study.
- Variance in endpoint or difference?
- Be conservative and test sensitivity to assumptions

What Numbers?

- How many is reasonable?
  - Ethical considerations
  - Time scales
  - Cost
  - Availability of participants/samples
  - State of Knowledge
- Are controls easier than cases?
  - Consider 2:1 or 3:1
- Think of numbers of events.
- Allow for losses – dropouts

What Effect Size?

- The smallest effect that you want your study to be able to detect:
  - “Clinically relevant difference”.
  - How much would you need to make you change policy or practice?
  - What size effect would be biologically significant?
  - What is reasonable to expect?
  - Must be feasible!
- Somewhere between minimum useful and maximum feasible

What Effect size?

- Which effect?
- What is already known?
- May be interested in more than one effect. Or subgroups.
- Is effect the effect itself or a change in the effect?

What are Power calculations for *really*?

In practice….

Choosing the sample size
(Real life version)

- Identify the endpoint
- Identify the main comparison
- Estimate the variance of the endpoint
- Decide on the effect size
- Decide on the power you need
- Calculate the sample size
- Consider the practicalities
- And….

Negotiate

- What Effect size?
- What Power?
- What Numbers?
- Another design
  - number of groups, matching,...
- Consider alternatives and look at sensitivity to assumptions and variance estimates
- Subgroup analysis?

The Decision

- Is the design appropriate, optimal and ethical?
- Given the cost of the study (Effort, Financial, Ethical) with the number of participants you propose (sample size) and given the chance of finding a positive result (power) of the size your study can detect (effect size), is the study worth doing?

Is this all statistical waffle?

- Well it can be!
- But power calculations are good estimates based on the assumptions
- But assumptions can be questioned and negotiated
- So assumptions must be made clear and open to counter-argument.

There is no right answer

- Power calculations are correct given the assumptions but...
- There are many competing designs for any study with pros and cons
- The calculations have to make assumptions, which may well prove wrong!

In Practice….

For the purpose of sample size estimation, most hypothesis testing studies can be made to look like either:

- A comparison of two means
- A comparison of two proportions

Compare two means

- Student’s t-test
- Paired or unpaired?
- Effect size: difference between groups
- SD of measurements on each group (of difference if paired)
- Consider using log-transformed data

- Use software to do computation

Compare two proportions

- χ^2 test (Some software will do Fisher's Exact test as well - little practical difference)
- Effect size
  - Two proportions
  - Odds ratio (Case-control)
  - Relative Risk (Cohort)
- Software tends to split according to design

Philosophy

- Sample size calculations are not an exact science - the formulae might be accurate, but the data you plug into them is usually fairly tentative.
- When appropriate power calculations provide a rational, numerate basis to inform the decision on the size of study, and what (if anything!) you can expect to get out of it.

Pitfalls

- Failure to state assumptions
  - Effect size and why
  - Variances/Prevalence estimates – from where?
  - Compliance/dropout allowance
- Spurious accuracy - e.g. n=996
- Dishonesty
  - If n=100 because that is all you can do – say so! and justify that it is worth doing anyway.

Pitfalls

- This number worked last time
- We always use 3
- But some fields have good conventions (eg image analysis)

Pitfalls

- Not providing any justification for study size
- Inappropriate power calculations
  - Inventing a hypothesis to power
  - Powering the wrong design (e.g. paired v unpaired)
  - Trying to make up something without any data
- Expecting the same power in subgroup analyses

Summary

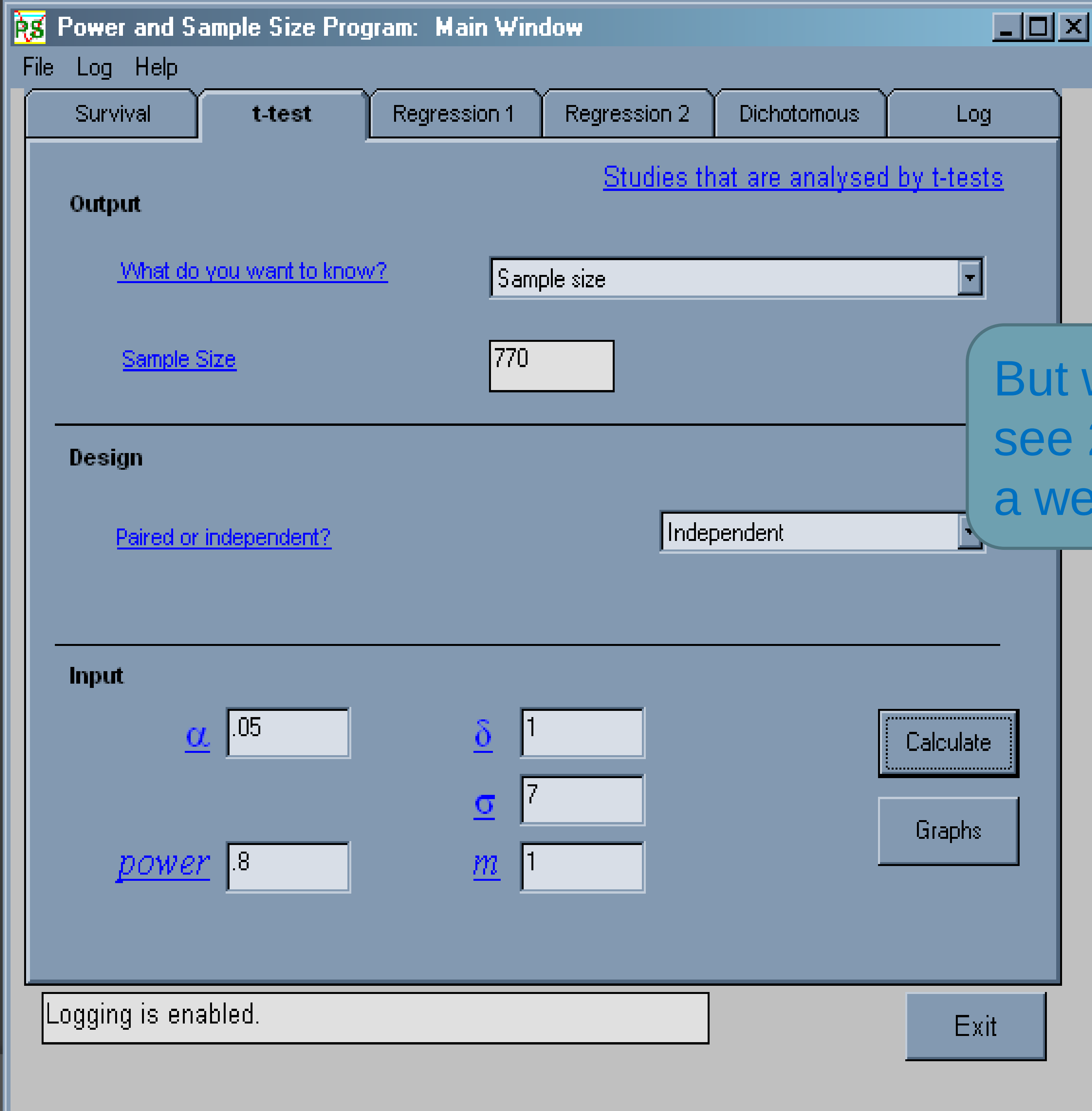
- Sample size calculations inform the argument they do not usually offer a definitive answer.
- The assumptions are central and should be clearly stated.
- Simple calculations and approximations are usually adequate.
  - Complex designs and large trials are the exception, but here you will have a statistician as part of the project team from the outset

Case Studies: Sample size

A session in the eye hospital….

Study 1

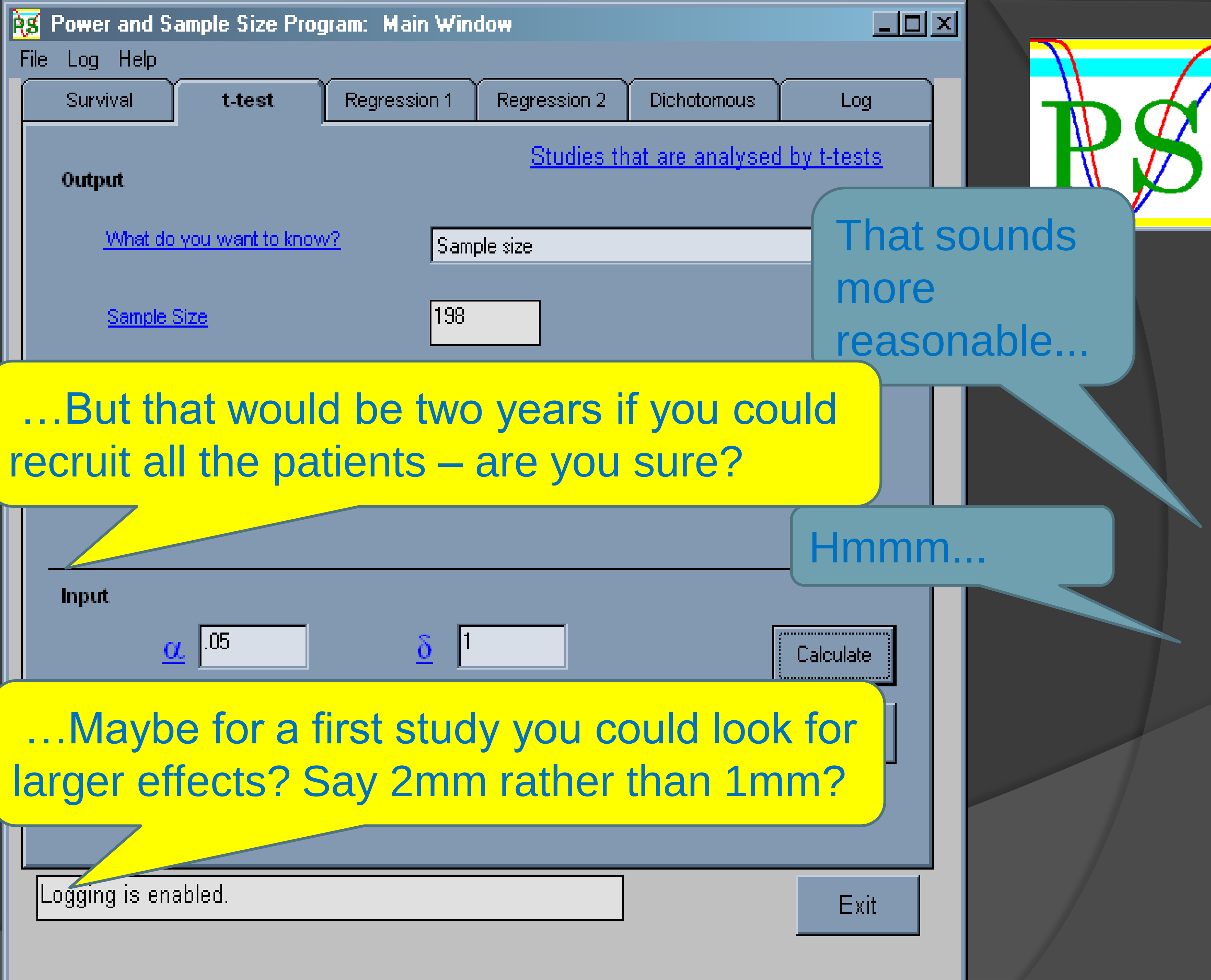
- Treatment for Ocular hypotension (low pressure inside eye)
- Compare two treatments: post-treatment
- Clinically significant difference 1mmHg
- SD between patients 7mmHg

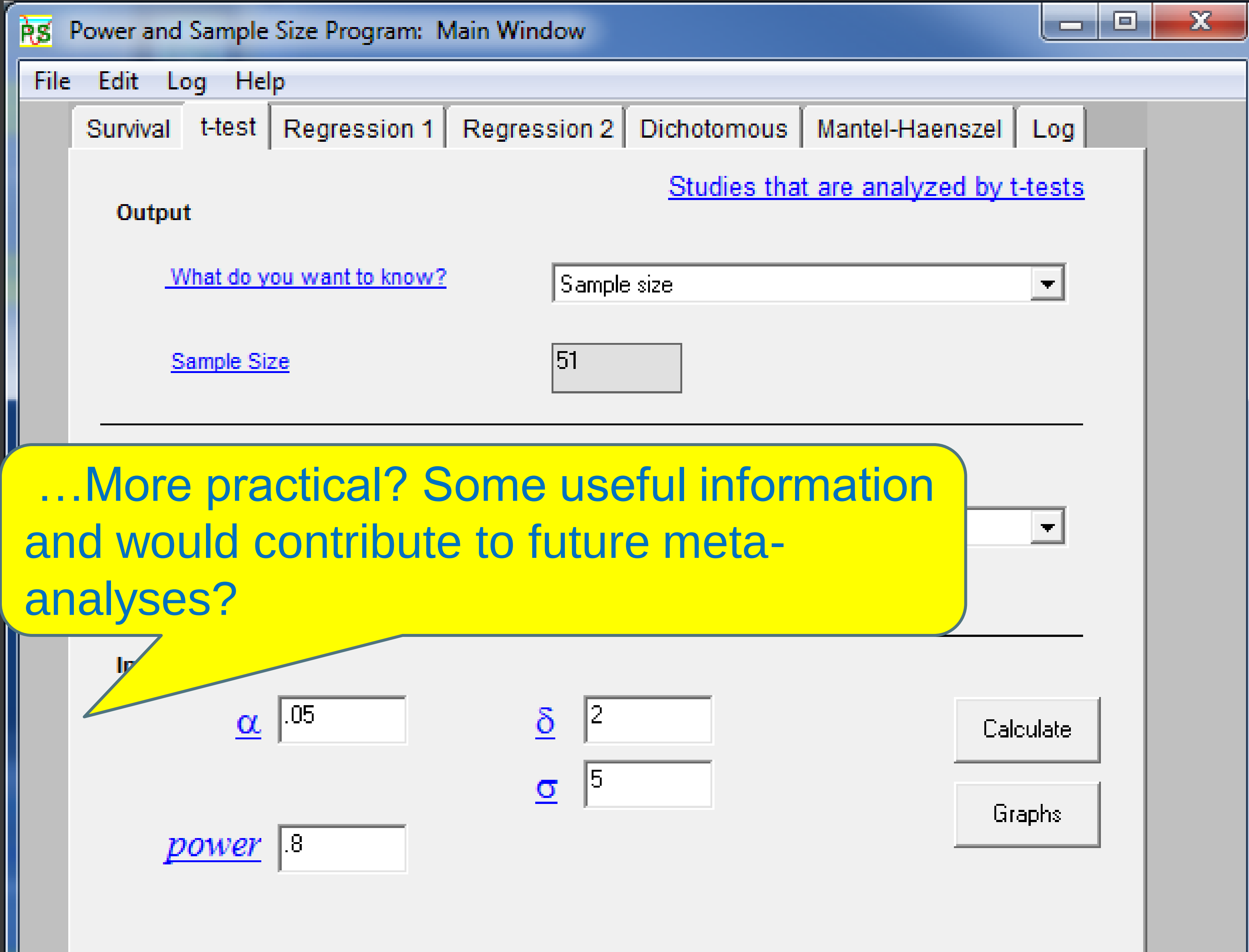


But we only see 2 patients a week!

Study 1 revised

- Treatment for *chronic, stable,* Ocular hypotension
- Compare two treatments: crossover design, *compare within patients*
- Clinically significant difference 1mmHg
- SD between patients 7mmHg
- SD of differences 5mmHg

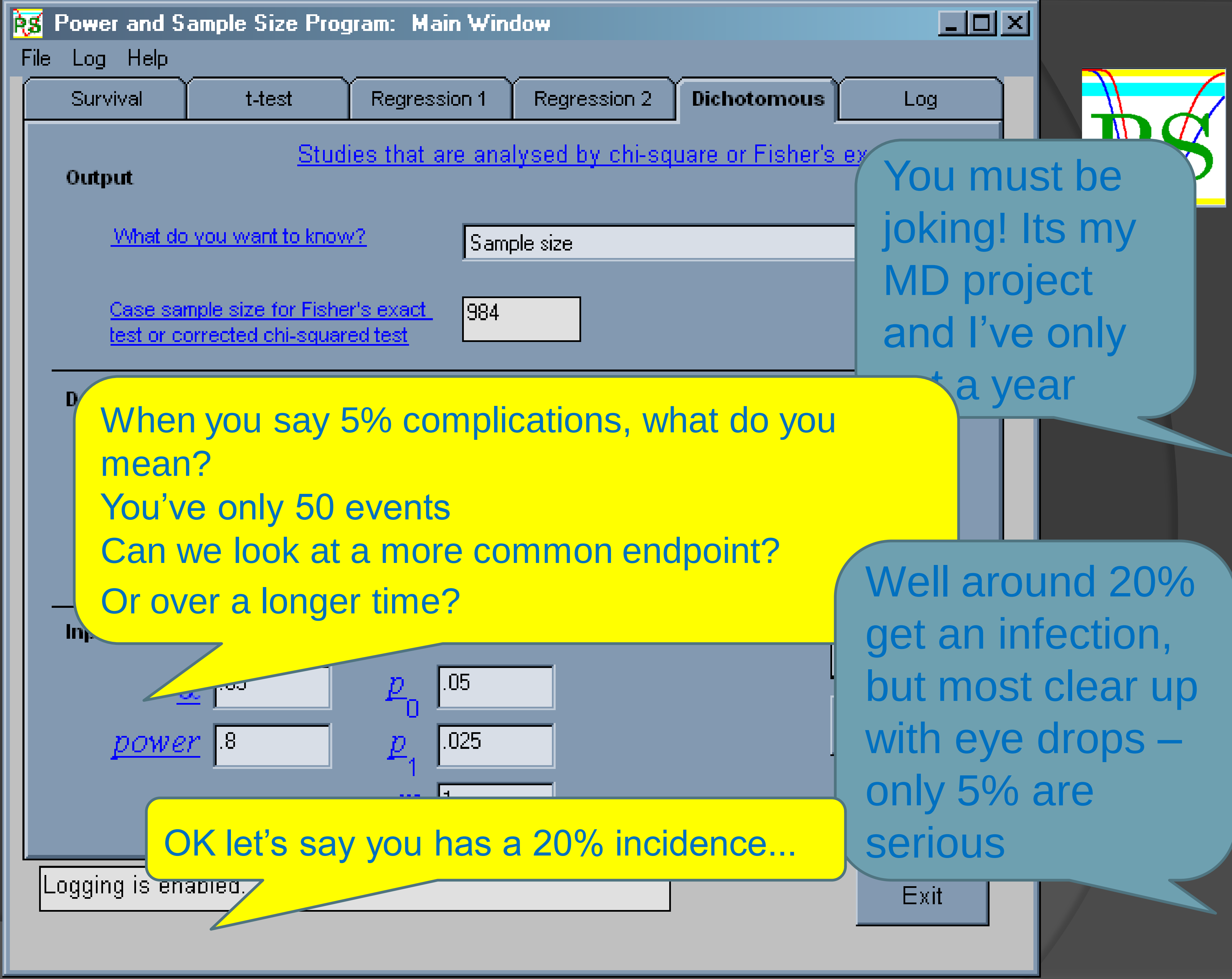

Power and Sample Size Program: Main Window
File Log Help
Survival
t-test
Regression 1
Regression 2
Dichotomous
Log
Studies that are analysed by t-tests
Output
What do you want to know?
Sample size
Sample Size
198
Input
α
.05
δ
1
Calculate
Logging is enabled.
Exit
That sounds more reasonable...
…But that would be two years if you could recruit all the patients – are you sure?
Hmmm...
…Maybe for a first study you could look for larger effects? Say 2mm rather than 1mm?



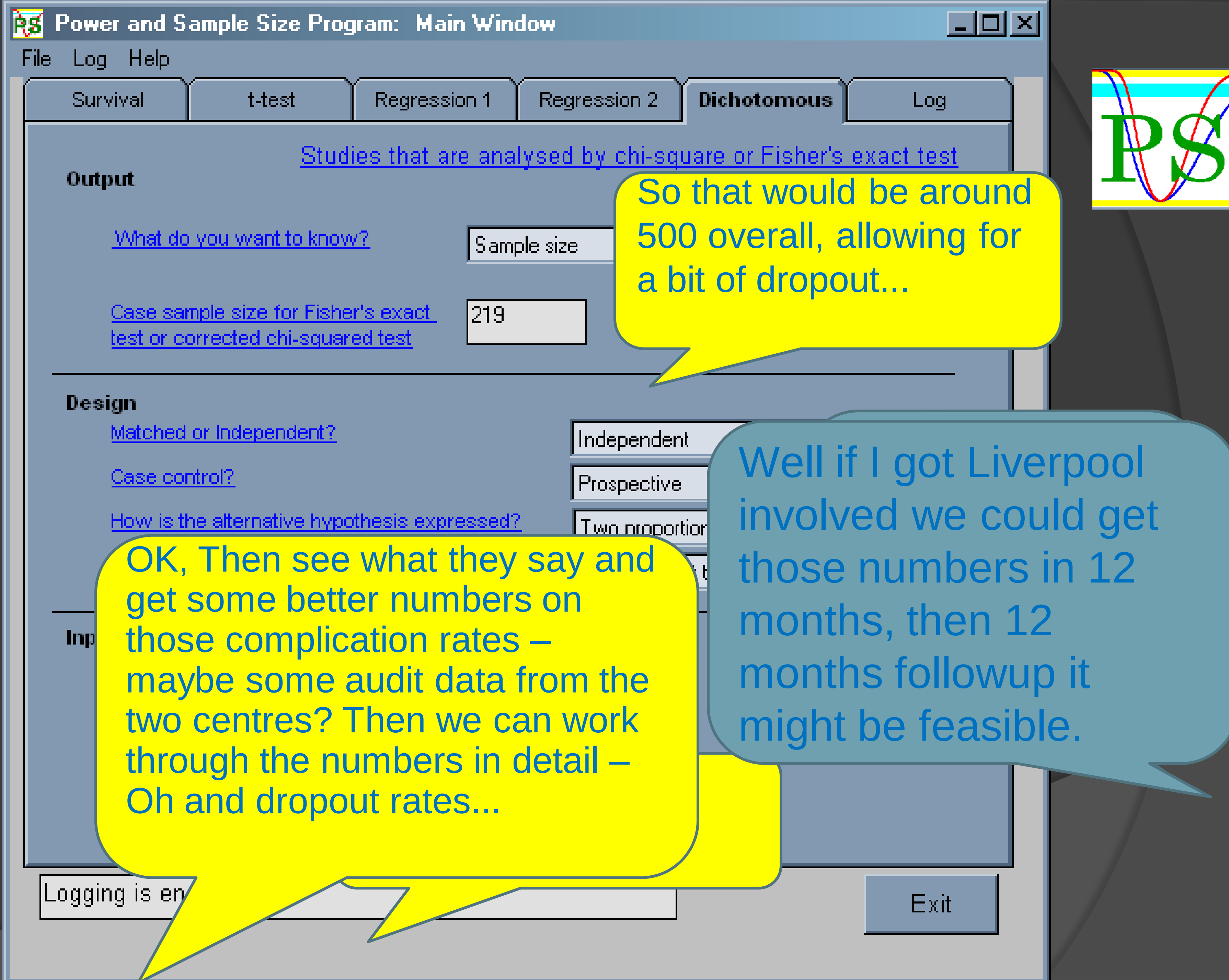
…More practical? Some useful information and would contribute to future meta-analyses?

Study 2

- Comparison of two lens implant types
- “Standard” has 5% complication rate requiring replacement
- Would use “new” if it halved that rate



Power and Sample Size Program: Main Window
File Log Help
Survival
t-test
Regression 1
Regression 2
Dichotomous
Log
Studies that are analysed by chi-square or Fisher's ex
Output
What do you want to know?
Sample size
Case sample size for Fisher's exact test or corrected chi-squared test
984
power
.8
p_0
.05
p_1
.025
Logging is enabled.
Exit
You must be joking! Its my MD project and I've only
a year
When you say 5% complications, what do you mean?
You've only 50 events
Can we look at a more common endpoint?
Or over a longer time?
Well around 20% get an infection, but most clear up with eye drops – only 5% are serious
OK let's say you has a 20% incidence...


Power and Sample Size Program: Main Window
File Log Help
Survival
t-test
Regression 1
Regression 2
Dichotomous
Log
Studies that are analysed by chi-square or Fisher's exact test
Output
What do you want to know?
Sample size
Case sample size for Fisher's exact test or corrected chi-squared test
219
Design
Matched or Independent?
Independent
Case control?
Prospective
How is the alternative hypothesis expressed?
So that would be around 500 overall, allowing for a bit of dropout...
Well if I got Liverpool involved we could get those numbers in 12 months, then 12 months followup it might be feasible.
OK, Then see what they say and get some better numbers on those complication rates – maybe some audit data from the two centres? Then we can work through the numbers in detail – Oh and dropout rates...
Exit

Resources

Software

- PS (free download) http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize
- Stats Direct does common stuff www.statsdirect.com
- R, Stata have functions to compute power
- NQuery Advisor (Expensive) www.statsol.ie/nquery/nquery.htm
- Several Web sites – e.g. Power Calculator http://calculators.stat.ucla.edu/powercalc/

See www.biostat.ucsf.edu/sampsize.html for a fuller list

Collaboration

- If its not simple then talk to your local friendly statistician….

Questions?