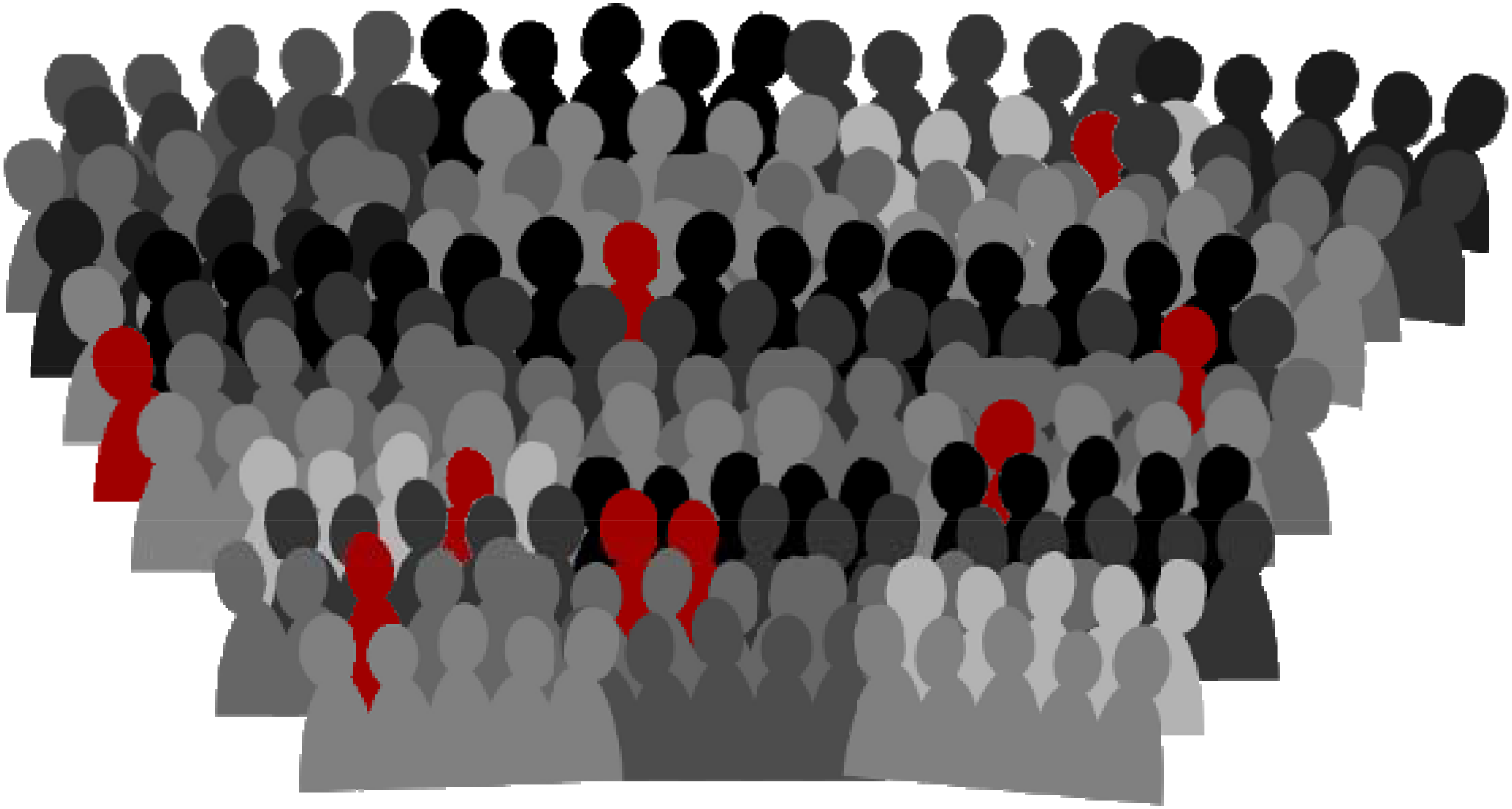
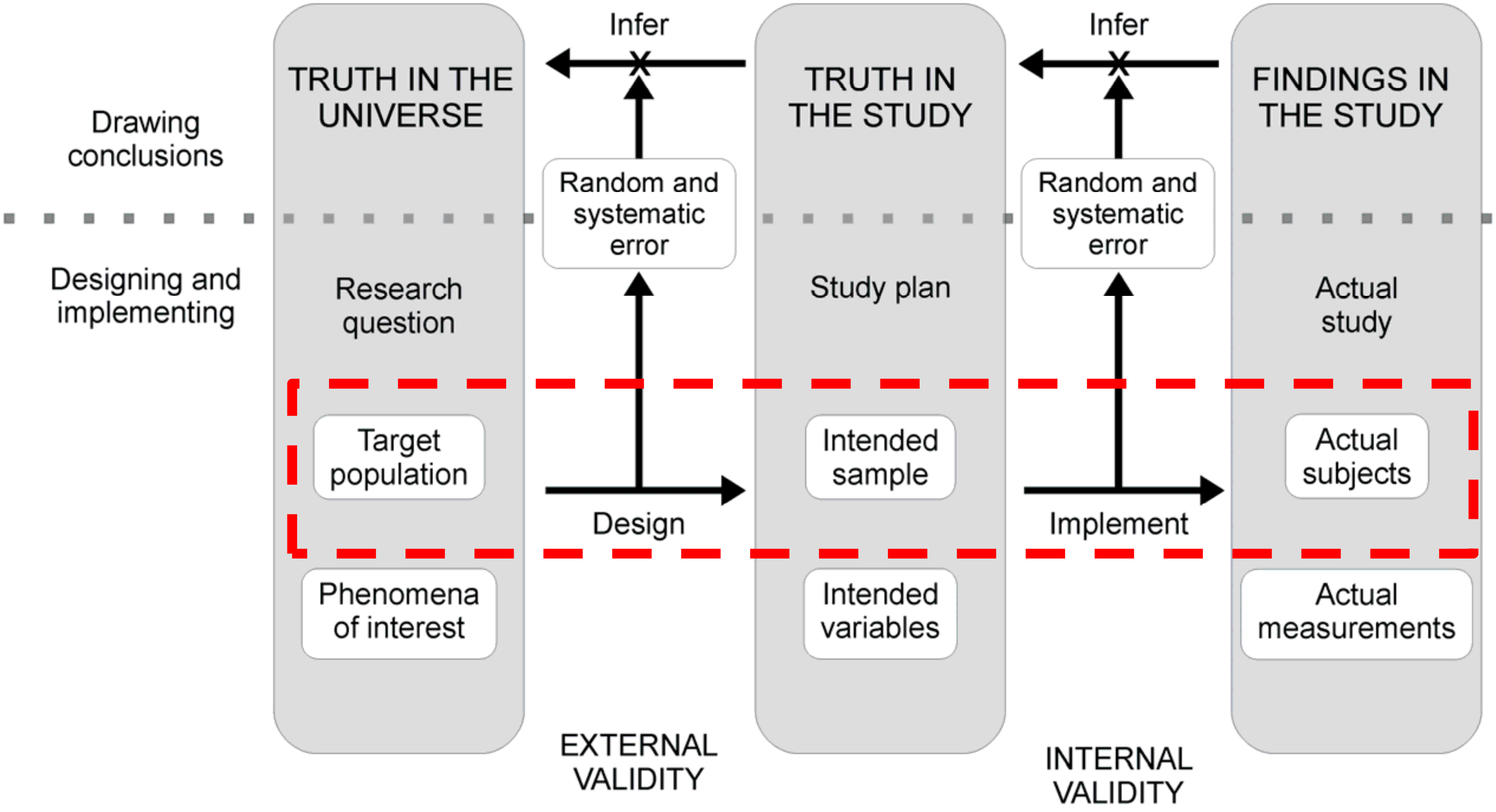
Sample Size Determination



Meridith Blevins, MS
Assistant in Biostatistics
Vanderbilt University Medical Center

Physiology of Research



Hulley, Stephen B., et al. *Designing clinical research*. LWW, 2013.

Importance

- Competition for grant funding
- Minimal exposure
- Expedite results
- Conserve funds
- Determine feasibility (“prohibitively large”)

Introduction

- Goal: estimate an *appropriate number* of ‘subjects’ for a given study design.
- Sample size calculations are only as accurate as the data and estimates on which they are based, which are often just informed guesses.
- Often reveals that the research design is not feasible or that different predictor or outcome variables are needed.

Ingredients for Sample Size Calculation

- Research Hypothesis
- Hypothesis Test
- Type I and II error rates
- Effect size
- Study design

Research hypothesis

- Should be *simple* (ie, contain one predictor and one outcome variable); *specific* (ie, leave no ambiguity about the subjects and variables or about how the statistical hypothesis will be applied); and *stated in advance*.
- *Example*: Adherence to antiretroviral medication, assessed with pharmacy records, is better in HIV patients with an enhanced counseling intervention compared with standard of care during the first year of treatment.

Hypothesis testing

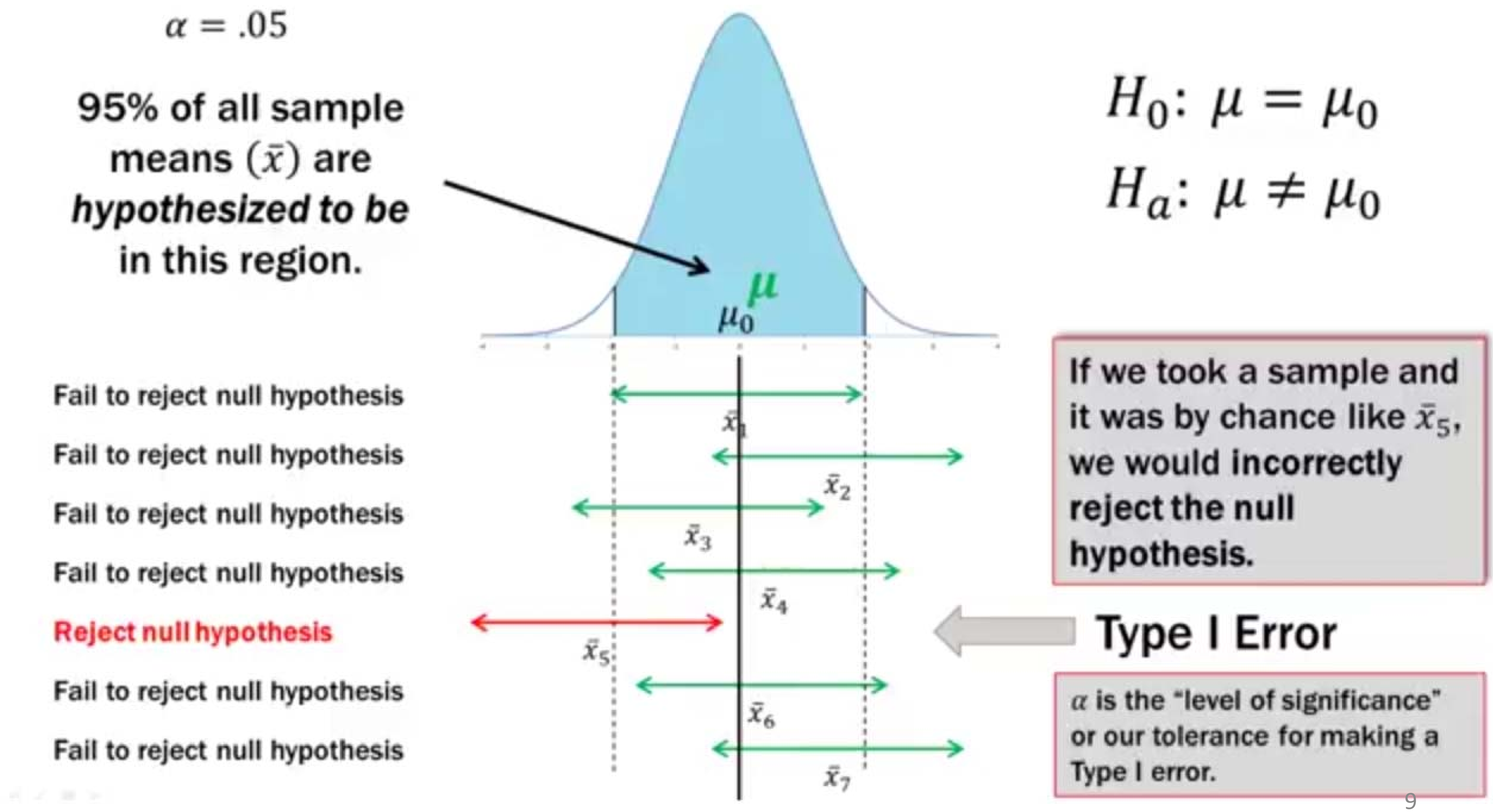
- Presume the null hypothesis (eg, no association between the predictor and outcome variables in the population).
- Based on the data collected in the sample, use statistical tests to determine whether there is sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis (eg, there is an association in the population).
- Example tests: chi-squared test (dichotomous) and t-test (continuous)

Help selecting test: http://www.ats.ucla.edu/stat/mult_pkg/whatstat/

Two possible errors

	Null Hypothesis is true	**Null hypothesis is false**
Reject null hypothesis	Type I Error False Positive	Correct Outcome True Positive
Fail to reject null hypothesis	Correct outcome True Negative	Type II Error False Negative

When null hypothesis is true…



$\alpha = .05$

95% of all sample means $(\bar{x})$ are ***hypothesized to be*** in this region.

$H_0\colon \mu = \mu_0$

$H_a\colon \mu \neq \mu_0$

If we took a sample and it was by chance like $\bar{x}_5$, we would incorrectly reject the null hypothesis.

Type I Error

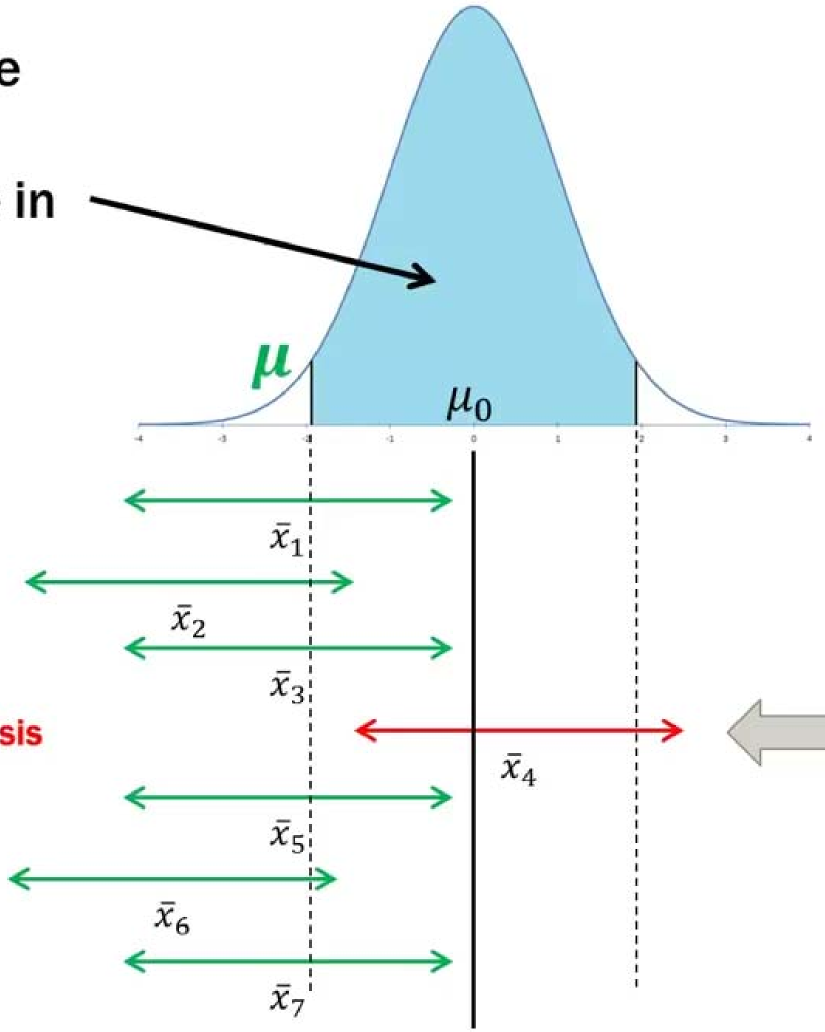
α is the "level of significance" or our tolerance for making a Type I error.

μ
μ_0
Fail to reject null hypothesis
$\bar{x}_1$
Fail to reject null hypothesis
$\bar{x}_2$
Fail to reject null hypothesis
$\bar{x}_3$
Fail to reject null hypothesis
$\bar{x}_4$
Reject null hypothesis
$\bar{x}_5$
Fail to reject null hypothesis
$\bar{x}_6$
Fail to reject null hypothesis
$\bar{x}_7$

Reference: https://youtu.be/k80pME7mWRM

When null hypothesis is false…

95% of all sample means ($\bar{x}$) are ***hypothesized to be*** in this region.
$\alpha = .05$



Reject null hypothesis

Reject null hypothesis

Reject null hypothesis

Fail to reject null hypothesis

Reject null hypothesis

Reject null hypothesis

Reject null hypothesis

$$H_0\colon \mu = \mu_0$$

$$H_a\colon \mu \neq \mu_0$$

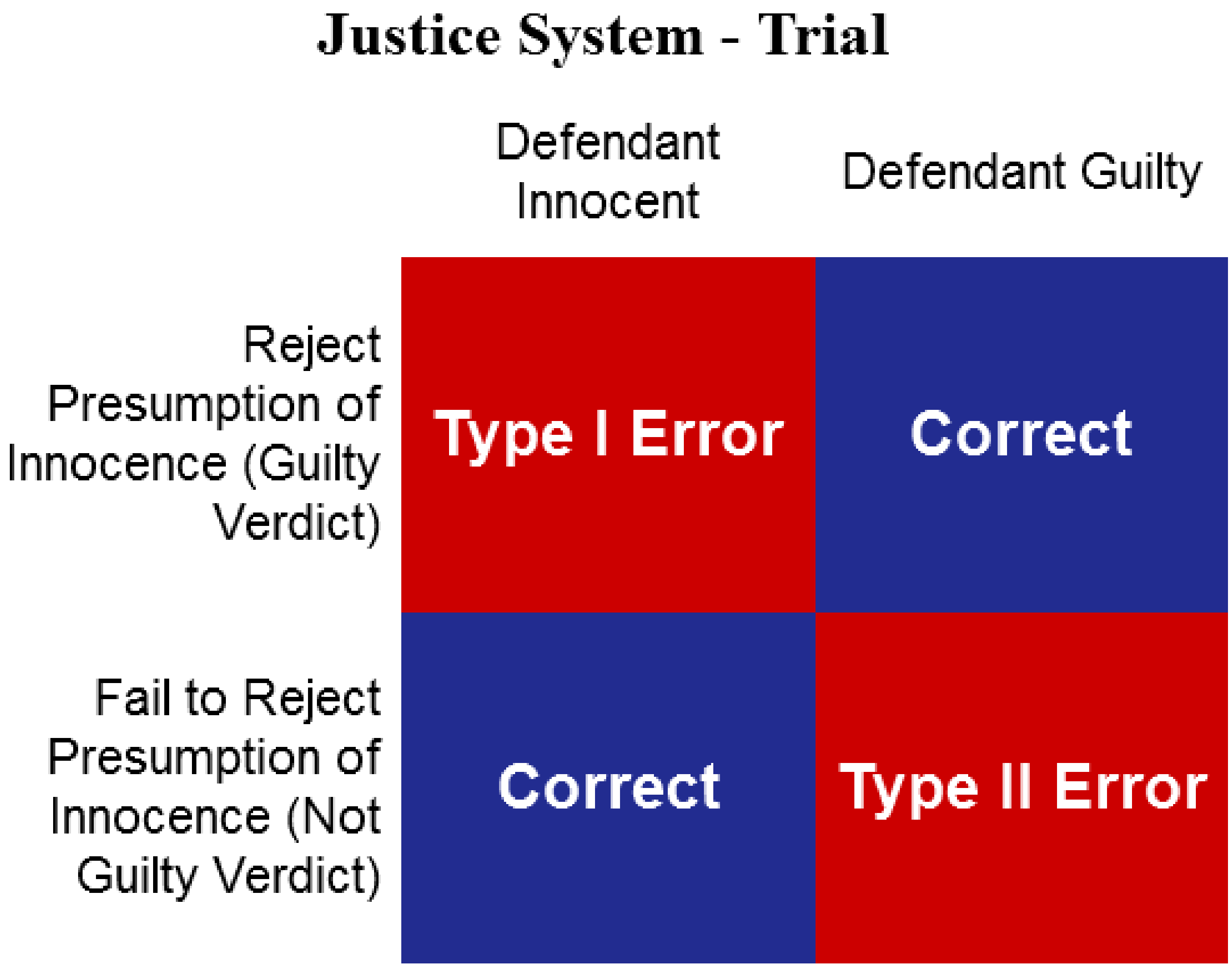
If we took a sample and it was by chance like $\bar{x}_4$, we would incorrectly "accept" the null hypothesis.

Type II Error

Beta (β) is the probability of committing a Type II error. The value of β varies with certain experimental factors.

Reference: https://youtu.be/k80pME7mWRM

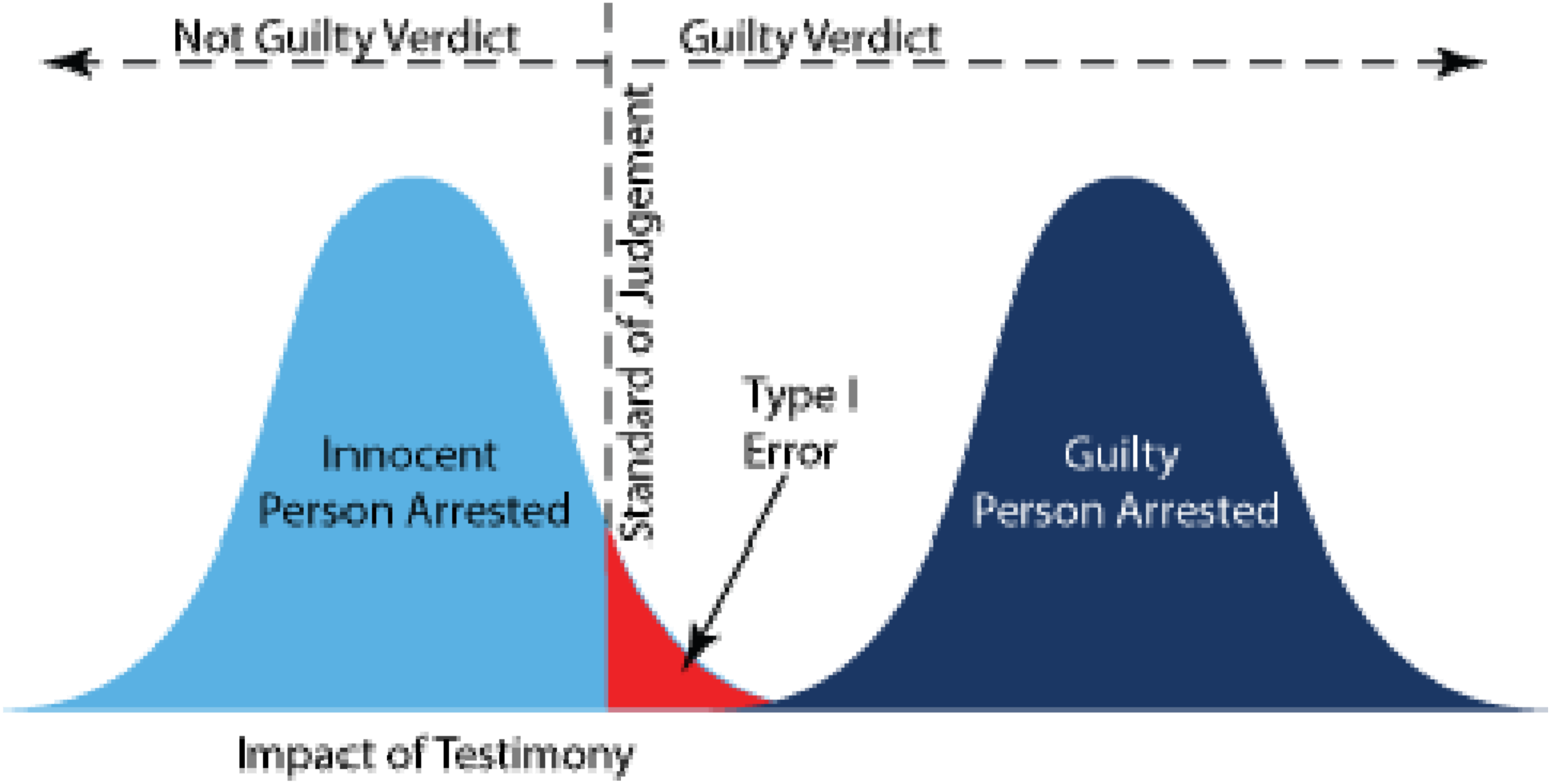
http://intuitor.com/statistics/T1T2Errors.html



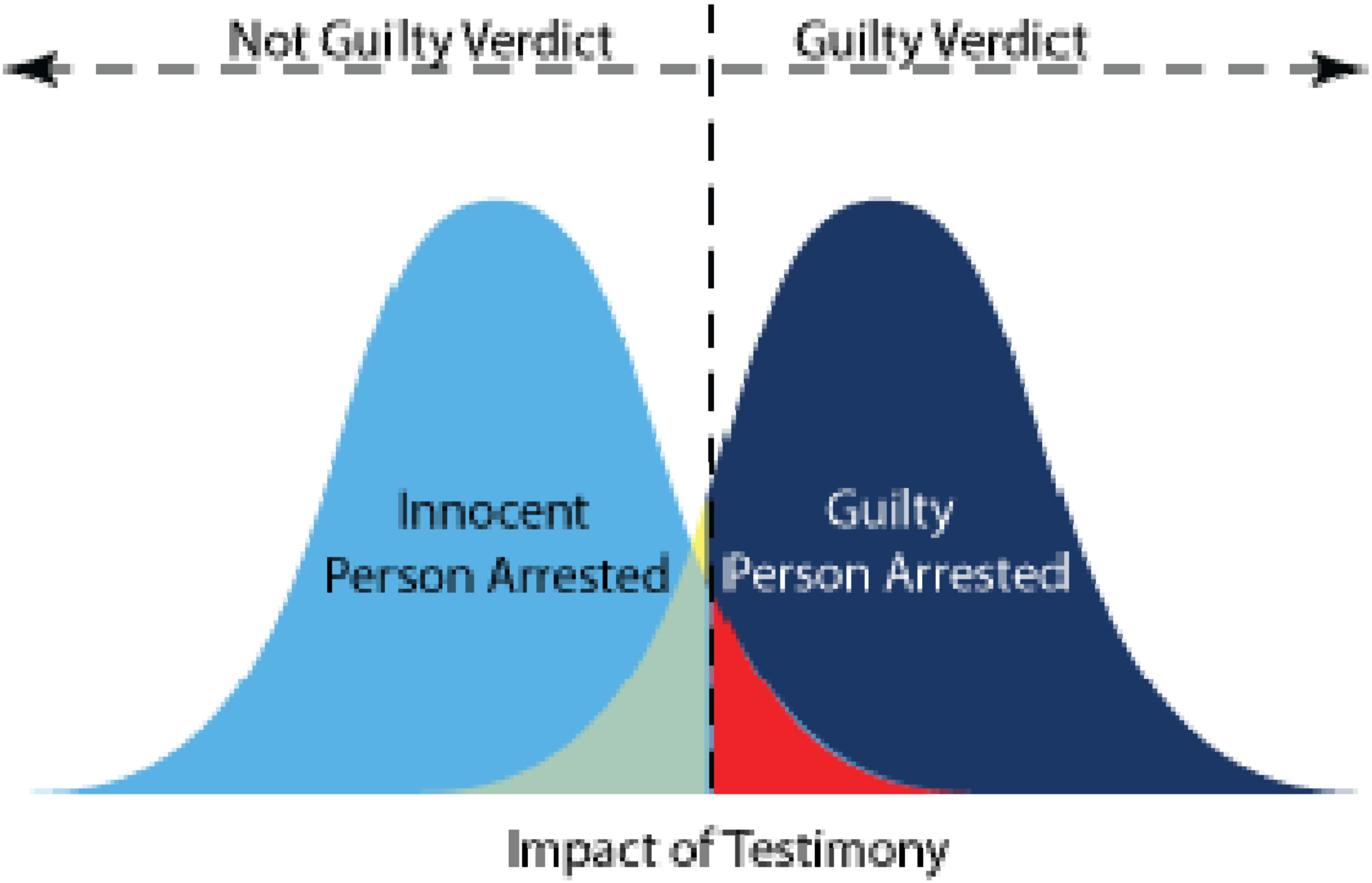
Justice System - Trial

	Defendant Innocent	Defendant Guilty
Reject Presumption of Innocence (Guilty Verdict)	Type I Error	Correct
Fail to Reject Presumption of Innocence (Not Guilty Verdict)	Correct	Type II Error

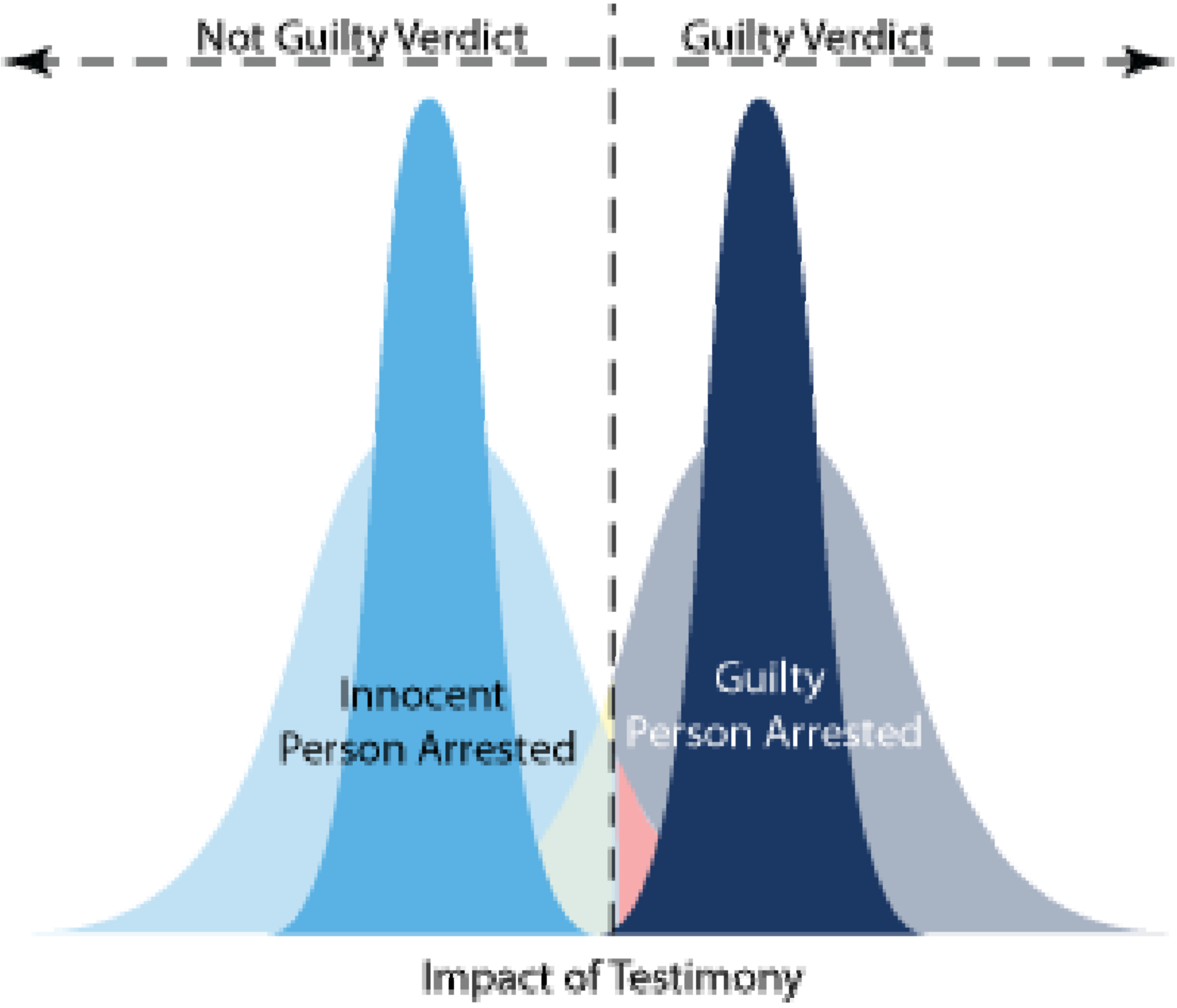
http://intuitor.com/statistics/T1T2Errors.html



http://intuitor.com/statistics/T1T2Errors.html



http://intuitor.com/statistics/T1T2Errors.html



Effect size

- Size of the association/difference/effect you expect to be present in the sample.
  - Mean difference (need standard deviation)
  - Two proportions
  - Odds Ratio
  - Hazard Ratio
  - Correlation coefficient
- *Good rule of thumb:* choose the smallest effect size that would be *clinically* meaningful (and you would hate to miss).
  - Will be okay if true effect size ends up being larger.

Study Design

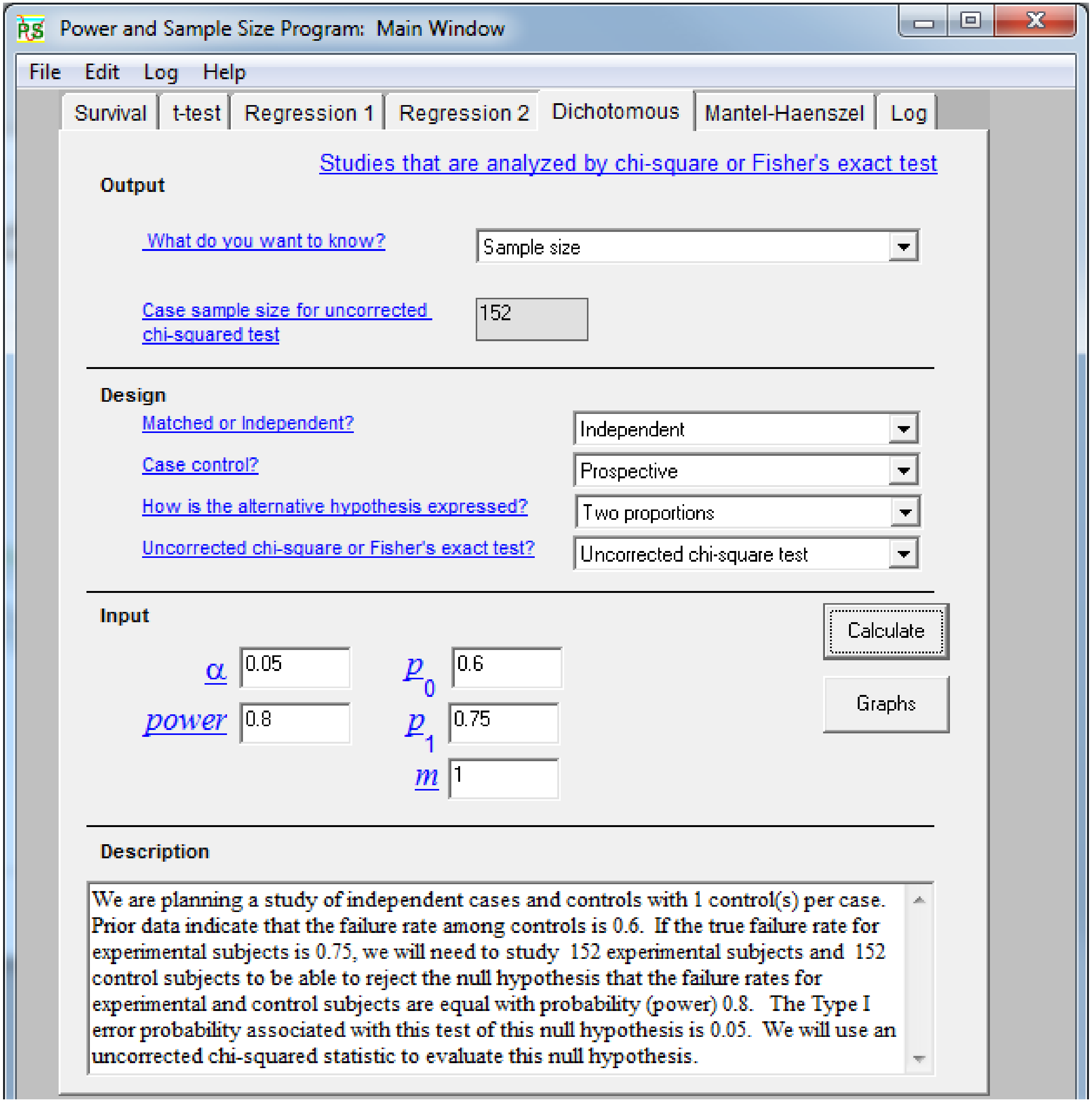
- Will subjects be allocated 1 to 1?
- Independent versus Case-Control (matching)
- Other considerations
  - Accrual/Enrollment (response rate)
  - Drop-outs (ie, loss to follow-up) and missing data
  - Budgetary constraints
  - Correlation (longitudinal or multi-site)

Recipe

1. State the null and alternative hypothesis.
2. Select the appropriate statistical test based on the type of predictor and outcome variables.
3. Choose a reasonable effect size (and variability, if necessary).
4. Specify α and power.
5. Use an appropriate table, formula, or software program to estimate the sample size.

Example using the Chi-square-test:

- *Research question*: Is there a difference in the adherence to antiretroviral medication between HIV patients enrolled to standard of care (SOC) and intervention groups?
- *Previous data*: the 1-year adherence to ART is about 0.60 in HIV patients on SOC.
- *Wish*: to determine that the 1-year adherence is 0.75 in HIV patients on intervention.
- *Assumptions*: (two-sided) = 0.05; power = 0.80; P1 (1-year adherence in SOC) = 0.60; P2 (1-year adherence in intervention) = 0.75.
- *Calculation*: If the true 1-year adherence rate following intervention is 0.75, a sample size of 152 HIV patients on SOC and 152 HIV patients on intervention is needed to reject the null hypothesis that 1-year adherence rates are equal for intervention and control groups with 80% power, using a Chi-square test and assuming a (two-sided) alpha of 0.05 and a 1-year adherence to ART of 0.60 in SOC group.



Power and Sample Size Program: Main Window
File
Edit
Log
Help
Survival
t-test
Regression 1
Regression 2
Dichotomous
Mantel-Haenszel
Log
Studies that are analyzed by chi-square or Fisher's exact test
Calculate
Graphs

Output

What do you want to know? Sample size

Case sample size for uncorrected chi-squared test 152

Design

Matched or Independent? Independent

Case control? Prospective

How is the alternative hypothesis expressed? Two proportions

Uncorrected chi-square or Fisher's exact test? Uncorrected chi-square test

Input

α 0.05

power 0.8

p_0 0.6

p_1 0.75

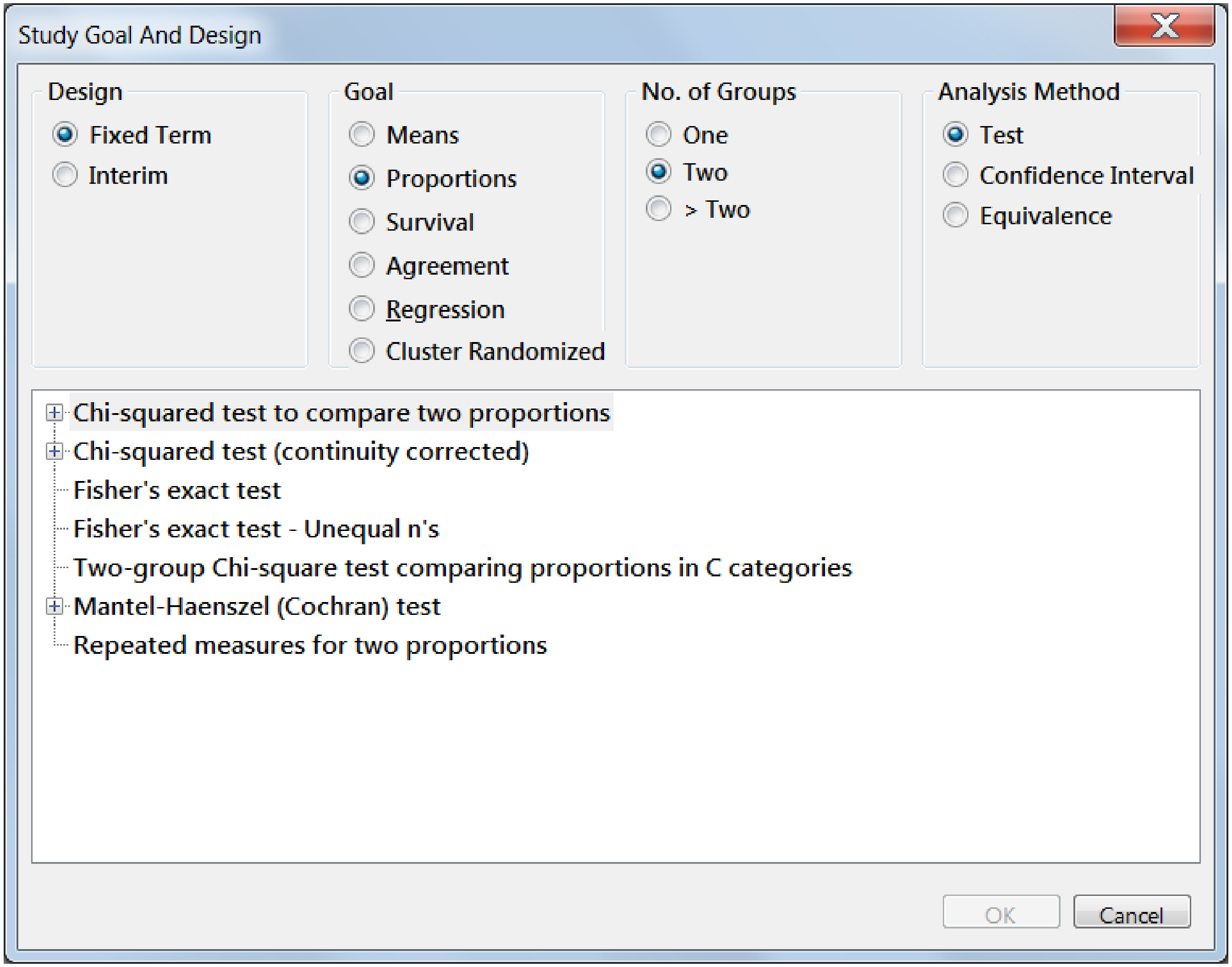
m 1

Description

We are planning a study of independent cases and controls with 1 control(s) per case. Prior data indicate that the failure rate among controls is 0.6. If the true failure rate for experimental subjects is 0.75, we will need to study 152 experimental subjects and 152 control subjects to be able to reject the null hypothesis that the failure rates for experimental and control subjects are equal with probability (power) 0.8. The Type I error probability associated with this test of this null hypothesis is 0.05. We will use an uncorrected chi-squared statistic to evaluate this null hypothesis.

http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize

Using nQuery



Study Goal And Design

Design
● Fixed Term
○ Interim

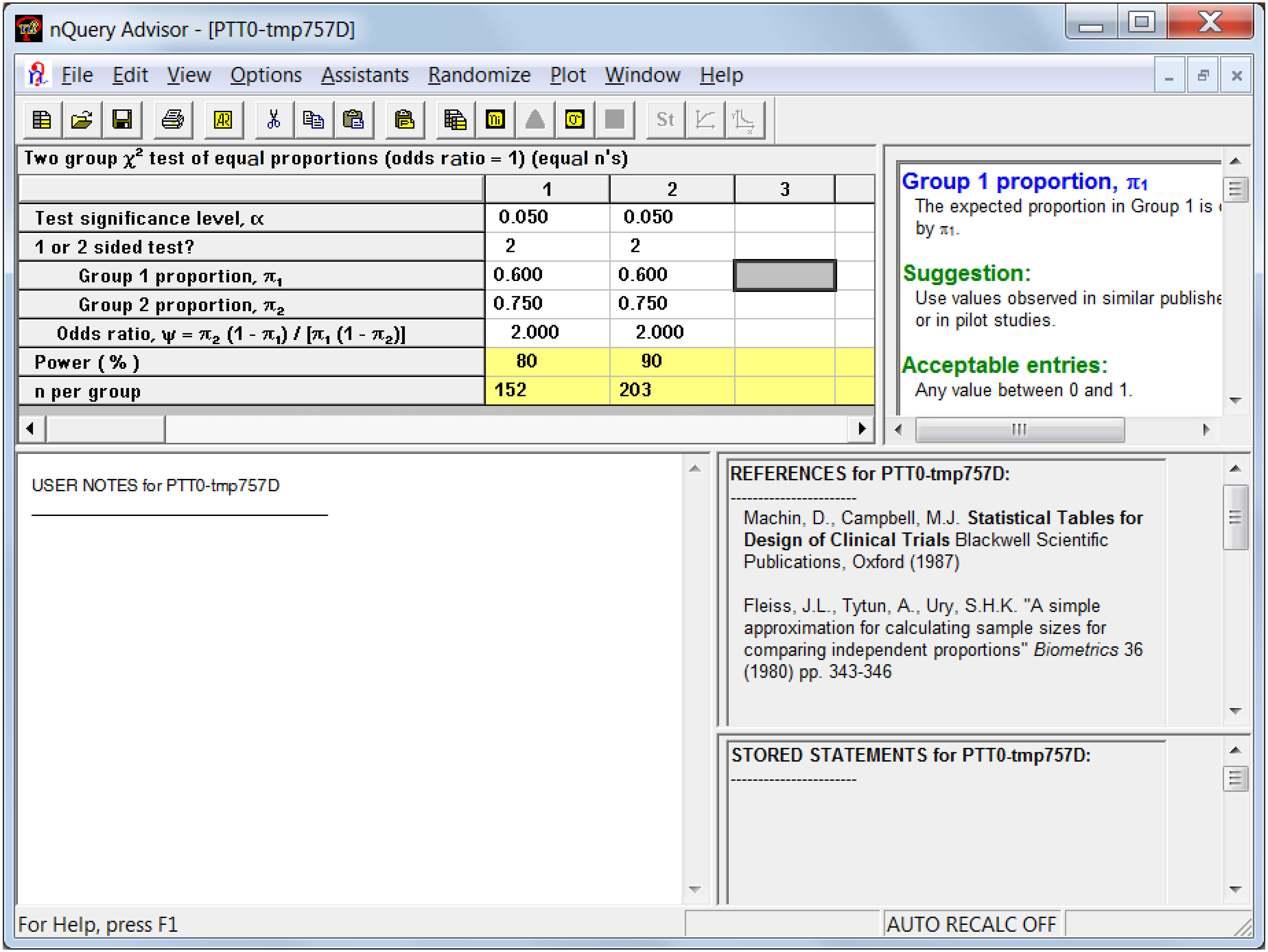
Goal
○ Means
● Proportions
○ Survival
○ Agreement
○ Regression
○ Cluster Randomized

No. of Groups
○ One
● Two
○ > Two

Analysis Method
● Test
○ Confidence Interval
○ Equivalence

- Chi-squared test to compare two proportions
- Chi-squared test (continuity corrected)
- Fisher's exact test
- Fisher's exact test - Unequal n's
- Two-group Chi-square test comparing proportions in C categories
- Mantel-Haenszel (Cochran) test
- Repeated measures for two proportions

Available to VUMC faculty/staff at https://it.vanderbilt.edu/software-store/


nQuery Advisor - [PTT0-tmp757D]
File Edit View Options Assistants Randomize Plot Window Help
St
For Help, press F1
AUTO RECALC OFF

Two group χ^2 test of equal proportions (odds ratio = 1) (equal n's)

	1	2	3
Test significance level, α	0.050	0.050	
1 or 2 sided test?	2	2	
Group 1 proportion, π_1	0.600	0.600	
Group 2 proportion, π_2	0.750	0.750	
Odds ratio, $\psi = \pi_2 (1 - \pi_1) / [\pi_1 (1 - \pi_2)]$	2.000	2.000	
Power (%)	80	90	
n per group	152	203	

Group 1 proportion, π_1

The expected proportion in Group 1 is by π_1.

Suggestion:

Use values observed in similar publishe or in pilot studies.

Acceptable entries:

Any value between 0 and 1.

USER NOTES for PTT0-tmp757D

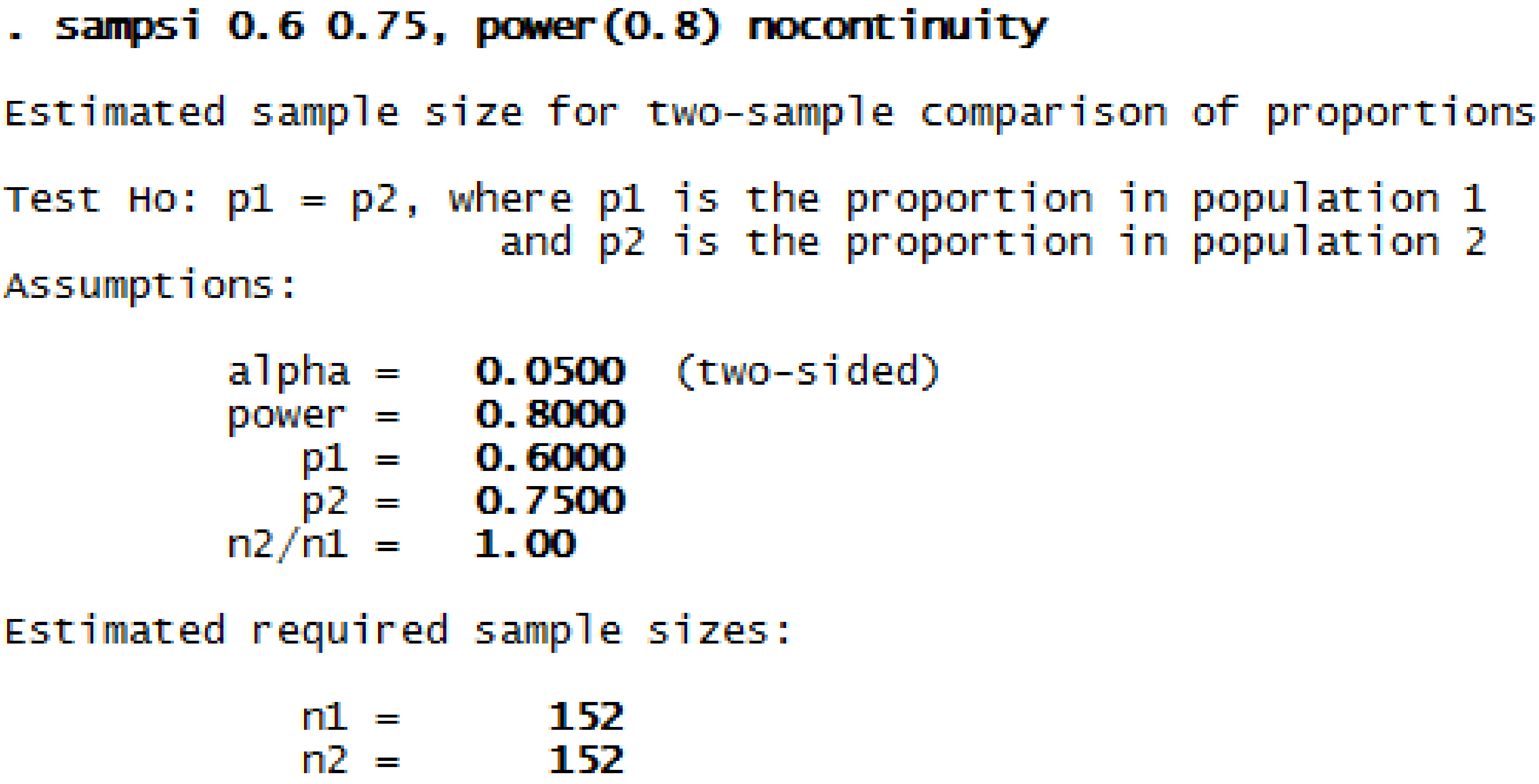
REFERENCES for PTT0-tmp757D:

Machin, D., Campbell, M.J. **Statistical Tables for Design of Clinical Trials** Blackwell Scientific Publications, Oxford (1987)

Fleiss, J.L., Tytun, A., Ury, S.H.K. "A simple approximation for calculating sample sizes for comparing independent proportions" *Biometrics* 36 (1980) pp. 343-346

STORED STATEMENTS for PTT0-tmp757D:

Using Stata

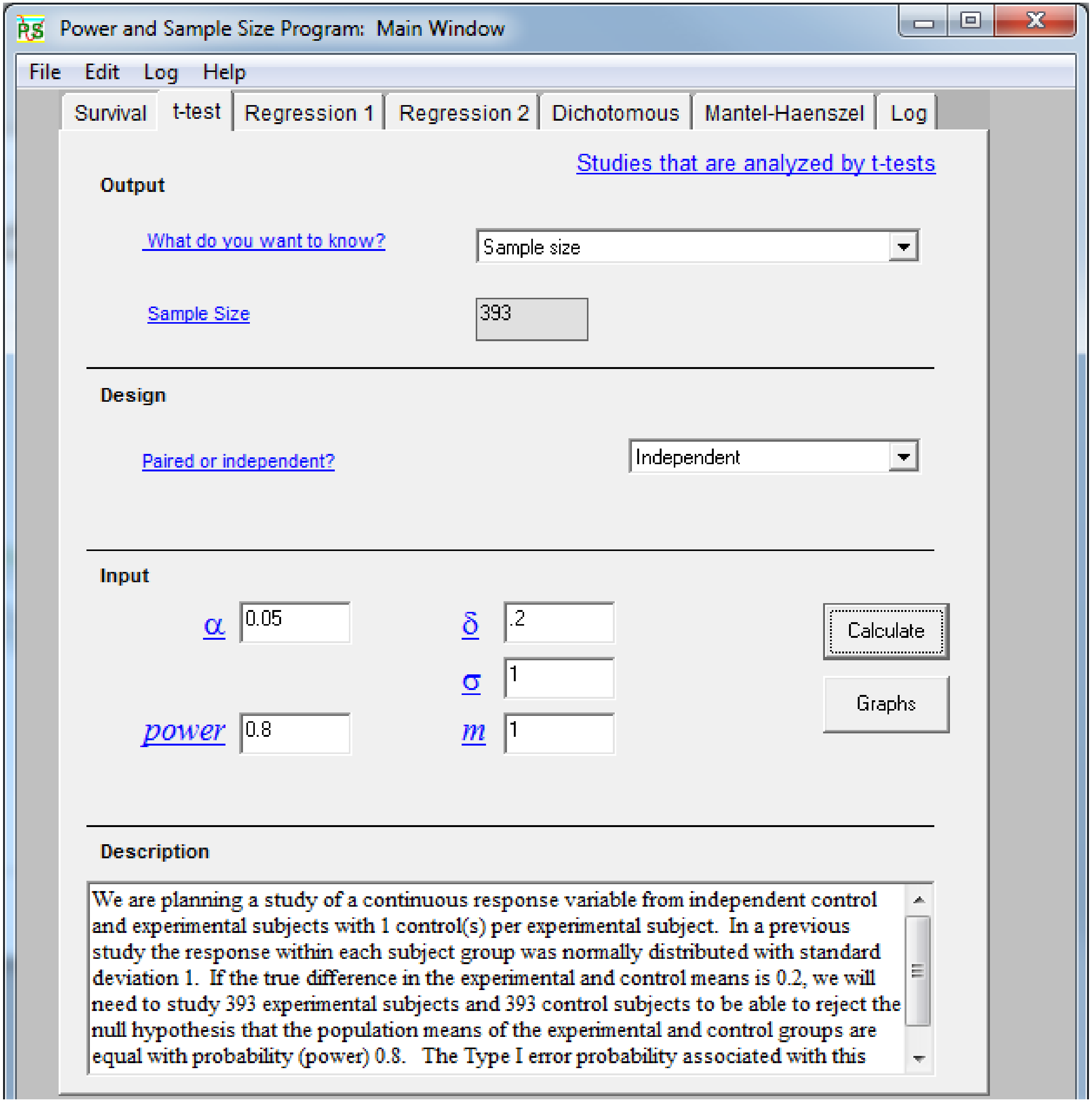


```
. sampsi 0.6 0.75, power(0.8) nocontinuity

Estimated sample size for two-sample comparison of proportions

Test Ho: p1 = p2, where p1 is the proportion in population 1
                    and p2 is the proportion in population 2
Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.8000
            p1 =   0.6000
            p2 =   0.7500
         n2/n1 =   1.00

Estimated required sample sizes:

            n1 =      152
            n2 =      152
```

See http://www.ats.ucla.edu/stat/stata/dae/proportionpow.htm

Example using the t-test:

- *Research question*: Is there a difference in the efficacy of two drugs (salbutamol and ipratropium bromide) for the treatment of asthma?
- *Planned study*: randomized trial of the effect of these drugs on FEV1 (forced expiratory volume in 1 second) after 2 weeks of treatment.
- *Previous data*: mean FEV1 in persons with asthma treated with ipratropium was 2.0 liters, with a SD of 1.0 liter.
- *Wish*: to be able to detect a difference of 10% in mean FEV1 between the 2 treatment groups.
- *Assumptions*: (two-sided) = 0.05; power = 0.80; effect size = 0.2 liters (10% X 2.0 liters); SD = 1.0 liter.
- *Calculation*: A sample size of 393 patients per group is needed to detect a difference of 10% in mean FEV1 between the 2 (independent) treatment groups with 80% power, using a two-sample t-test and assuming a (two-sided) alpha of 0.05, a mean FEV1 of 2.0 liters in the ipratropium group, and a SD of 1.0 liter.



Output

What do you want to know? Sample size

Sample Size 393

Design

Paired or independent? Independent

Input

α 0.05

δ .2

σ 1

power 0.8

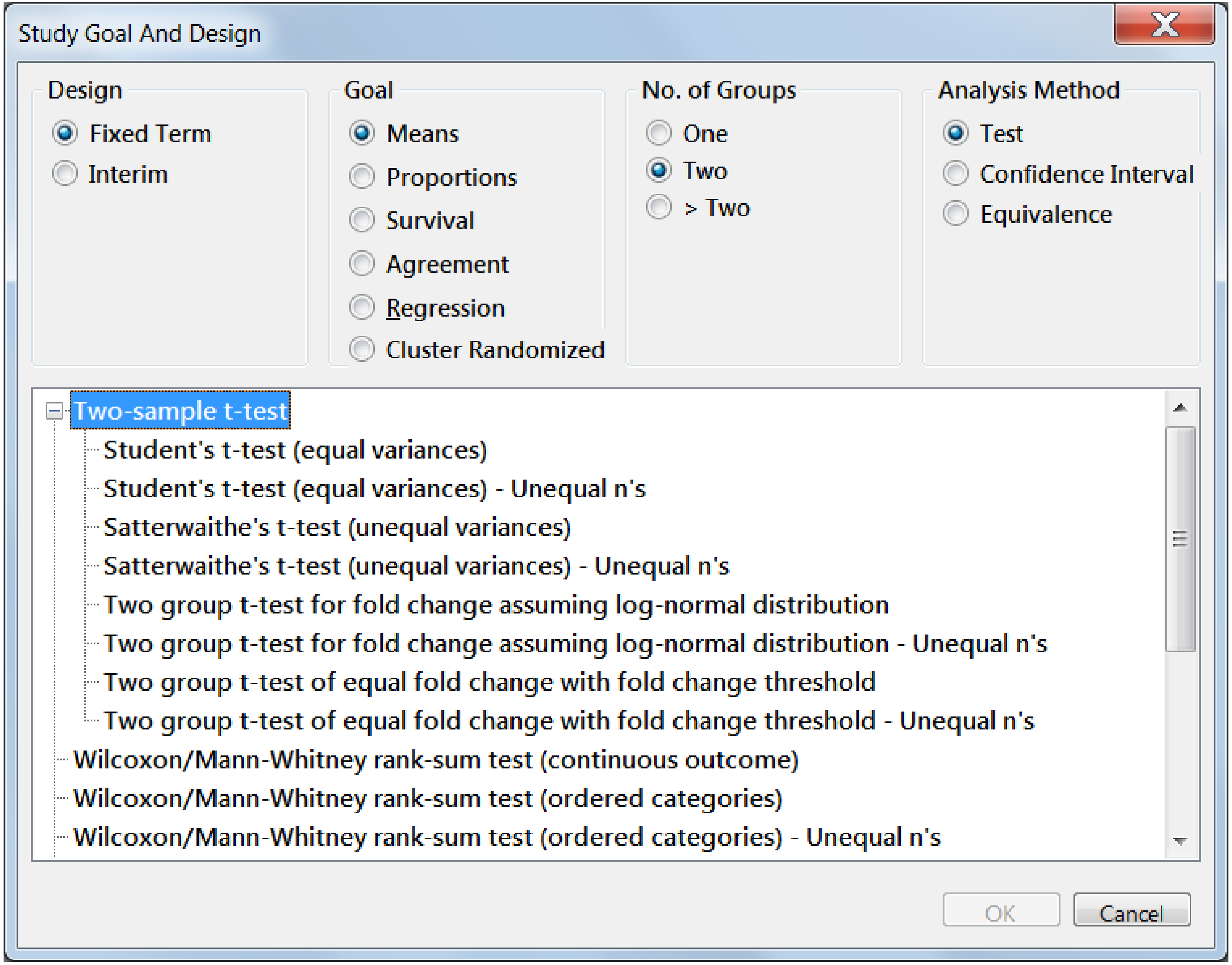
m 1

Description

We are planning a study of a continuous response variable from independent control and experimental subjects with 1 control(s) per experimental subject. In a previous study the response within each subject group was normally distributed with standard deviation 1. If the true difference in the experimental and control means is 0.2, we will need to study 393 experimental subjects and 393 control subjects to be able to reject the null hypothesis that the population means of the experimental and control groups are equal with probability (power) 0.8. The Type I error probability associated with this

http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize

Using nQuery



Study Goal And Design

Design

- ● Fixed Term
- ○ Interim

Goal

- ● Means
- ○ Proportions
- ○ Survival
- ○ Agreement
- ○ Regression
- ○ Cluster Randomized

No. of Groups

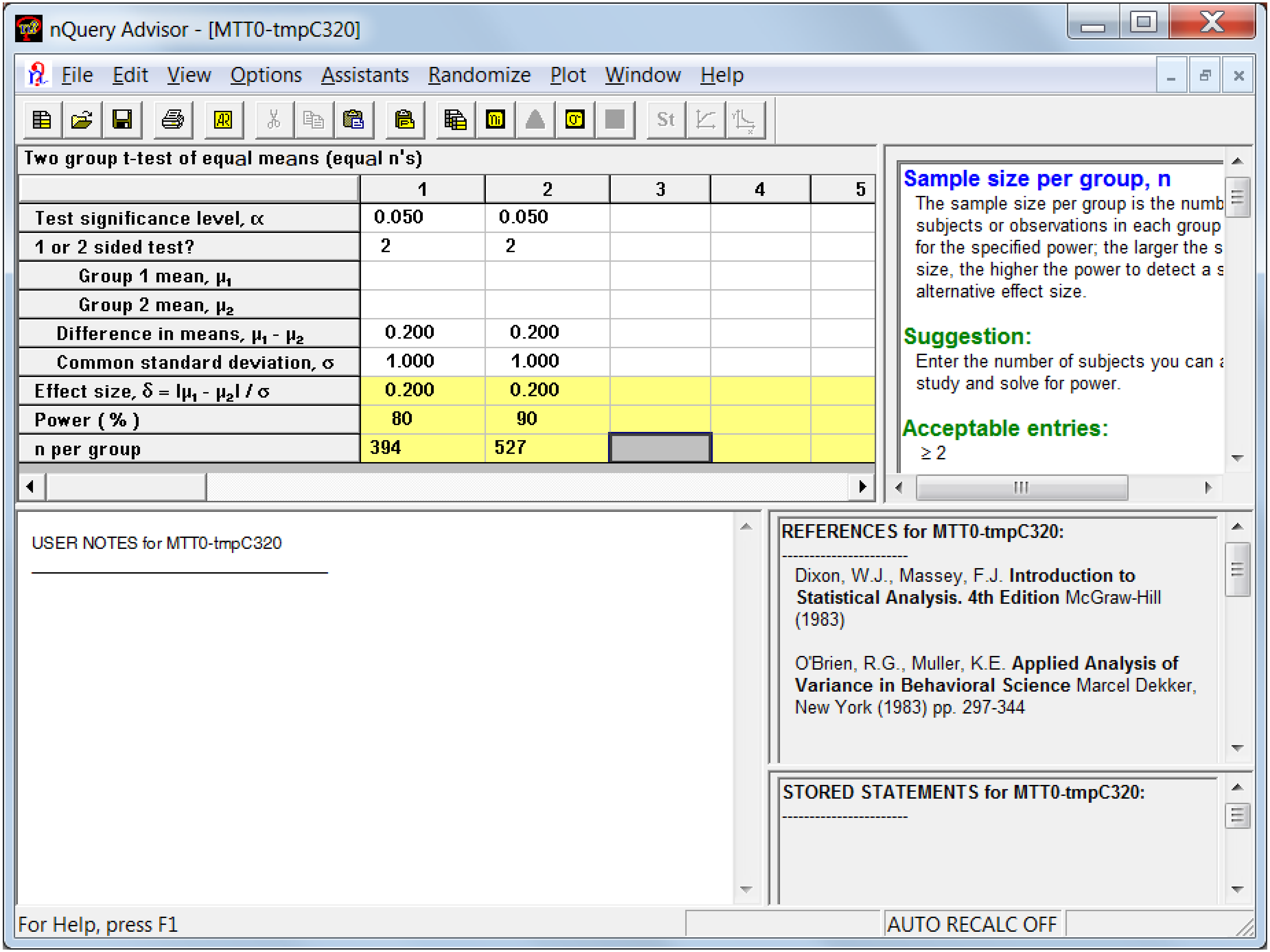
- ○ One
- ● Two
- ○ > Two

Analysis Method

- ● Test
- ○ Confidence Interval
- ○ Equivalence

- Two-sample t-test
  - Student's t-test (equal variances)
  - Student's t-test (equal variances) - Unequal n's
  - Satterwaithe's t-test (unequal variances)
  - Satterwaithe's t-test (unequal variances) - Unequal n's
  - Two group t-test for fold change assuming log-normal distribution
  - Two group t-test for fold change assuming log-normal distribution - Unequal n's
  - Two group t-test of equal fold change with fold change threshold
  - Two group t-test of equal fold change with fold change threshold - Unequal n's
- Wilcoxon/Mann-Whitney rank-sum test (continuous outcome)
- Wilcoxon/Mann-Whitney rank-sum test (ordered categories)
- Wilcoxon/Mann-Whitney rank-sum test (ordered categories) - Unequal n's

Available to VUMC faculty/staff at https://it.vanderbilt.edu/software-store/



Two group t-test of equal means (equal n's)

	1	2	3	4	5
Test significance level, α	0.050	0.050			
1 or 2 sided test?	2	2			
Group 1 mean, μ_1					
Group 2 mean, μ_2					
Difference in means, $\mu_1 - \mu_2$	0.200	0.200			
Common standard deviation, σ	1.000	1.000			
Effect size, $\delta = \lvert\mu_1 - \mu_2\rvert / \sigma$	0.200	0.200			
Power (%)	80	90			
n per group	394	527			

Sample size per group, n

The sample size per group is the numb subjects or observations in each group for the specified power; the larger the s size, the higher the power to detect a s alternative effect size.

Suggestion:

Enter the number of subjects you can a study and solve for power.

Acceptable entries:

≥ 2

USER NOTES for MTT0-tmpC320

REFERENCES for MTT0-tmpC320:

Dixon, W.J., Massey, F.J. **Introduction to Statistical Analysis. 4th Edition** McGraw-Hill (1983)

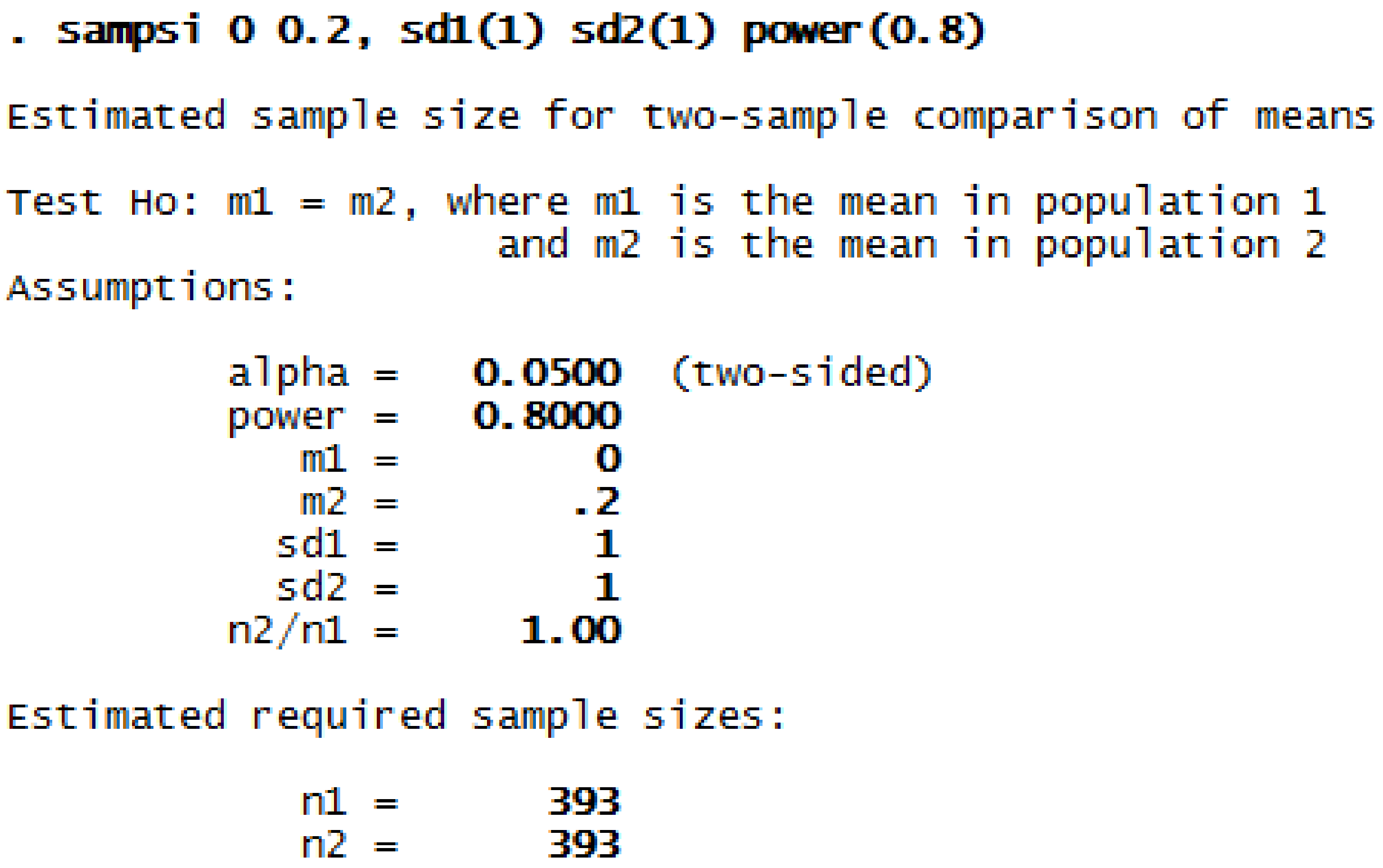
O'Brien, R.G., Muller, K.E. **Applied Analysis of Variance in Behavioral Science** Marcel Dekker, New York (1983) pp. 297-344

STORED STATEMENTS for MTT0-tmpC320:

For Help, press F1

AUTO RECALC OFF

Using Stata



```
. sampsi 0 0.2, sd1(1) sd2(1) power(0.8)

Estimated sample size for two-sample comparison of means

Test Ho: m1 = m2, where m1 is the mean in population 1
                    and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.8000
            m1 =        0
            m2 =       .2
           sd1 =        1
           sd2 =        1
         n2/n1 =     1.00

Estimated required sample sizes:

            n1 =      393
            n2 =      393
```

See http://www.ats.ucla.edu/stat/stata/dae/t_test_power2.htm

TIME FOR A BREAK

Sample Size Calculator

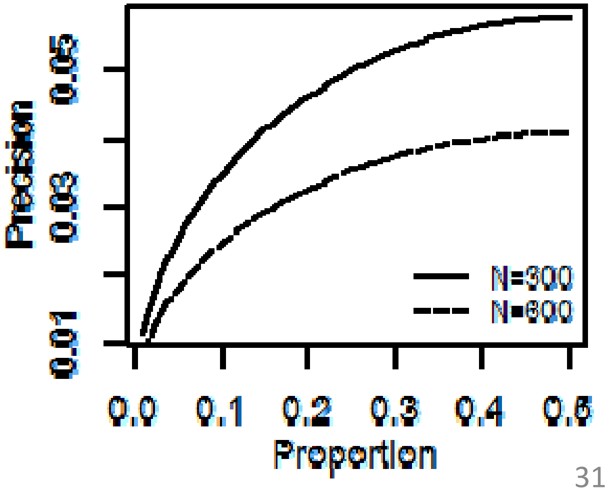
- PS is available freely at http://biostat.mc.vanderbilt.edu/PowerSampleSize
- Other online software is available that is free
- Some software are expensive (e.g. PASS) but they are very good too
- nQuery is currently free to VUMC faculty/staff
- Sometimes you just need a sample size formula from a text book

Strategies for minimizing sample size

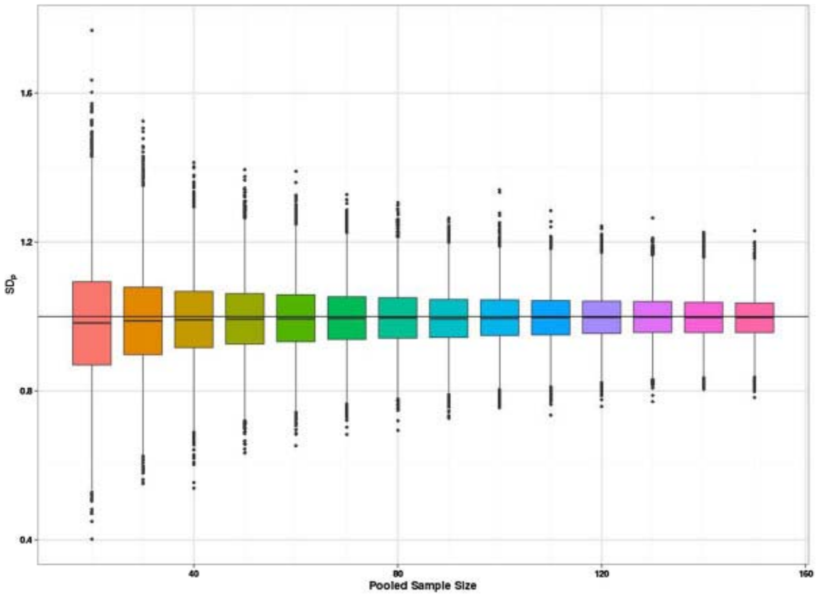
- Continuous variables,
- Paired measurements,
- Unequal group sizes, and
- A more common (ie, prevalent) binary outcome.
- Stricter inclusion criteria (eg, enroll people at risk for HIV acquisition as opposed to the general population)

Power vs. Precision

- Are you testing a hypothesis?
- Do you need to exclude a certain value for your study to be conclusive?
- Do you want to summarize a number of effects (e.g. survey items)?



Collecting Pilot Data



- External pilot or feasibility study
- Little consensus on size of pilot studies

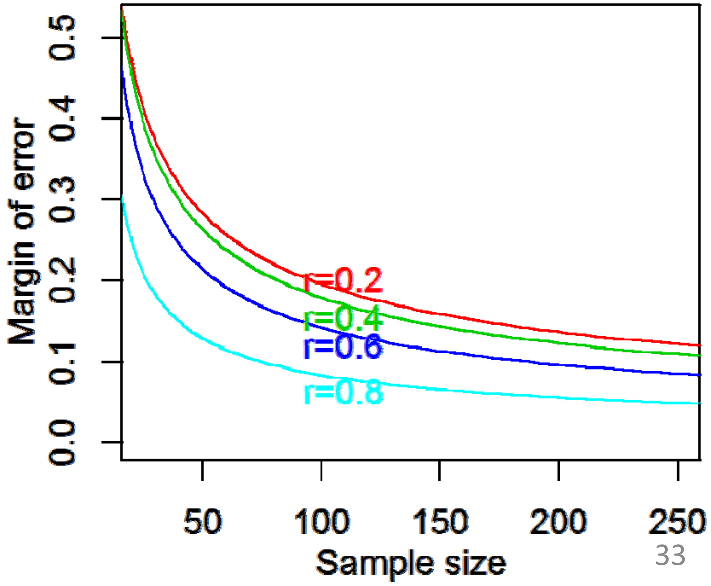
We recommend that an external pilot study has at least 70 measured subjects (35 per group) when estimating the SD_p for a continuous outcome. If the event rate in an intervention group needs to be estimated by the pilot then a total of 60 to 100 subjects is required. Hence if the primary outcome is binary a total of at least 120 subjects (60 in each group) may be required in the pilot trial.

Teare MD, Dimairo M, Shephard N, Hayman A, Whitehead A, Walters SJ. Sample size requirements to estimate key design parameters from external pilot randomised controlled trials: a simulation study. Trials. 2014 Jul 3;15(1):1.

No Pilot Data

- It would be a mistake to guess at distributions for continuous variables.
- Can the problem be reframed to use sample size approaches that require fewer assumptions?
  - Dichotomous sample size for continuous or time to event outcomes
  - Correlation coefficient
- For major grant applications, unacceptable.

Precision of Correlation Coefficient

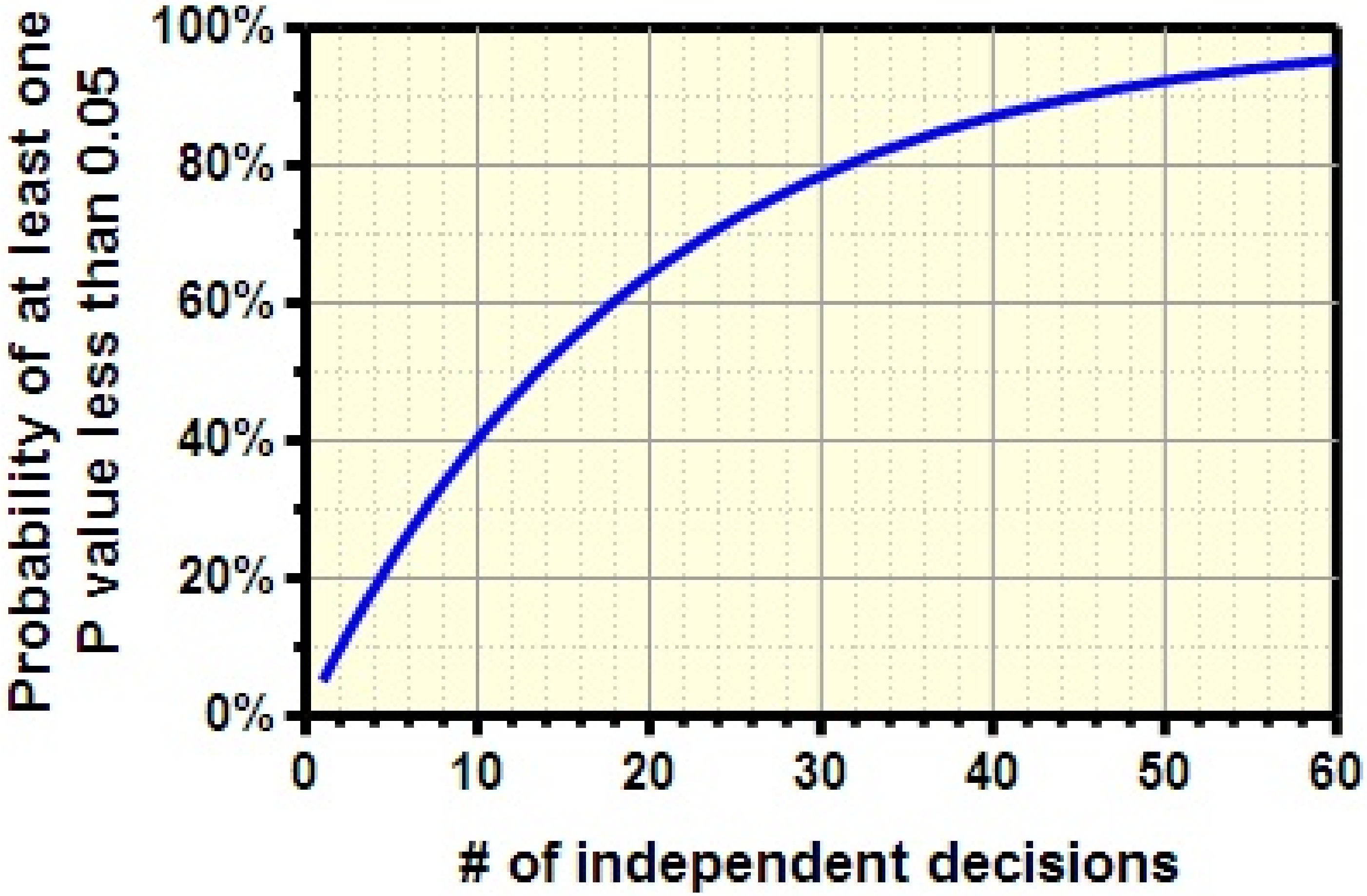


15:1 rule for regression modeling

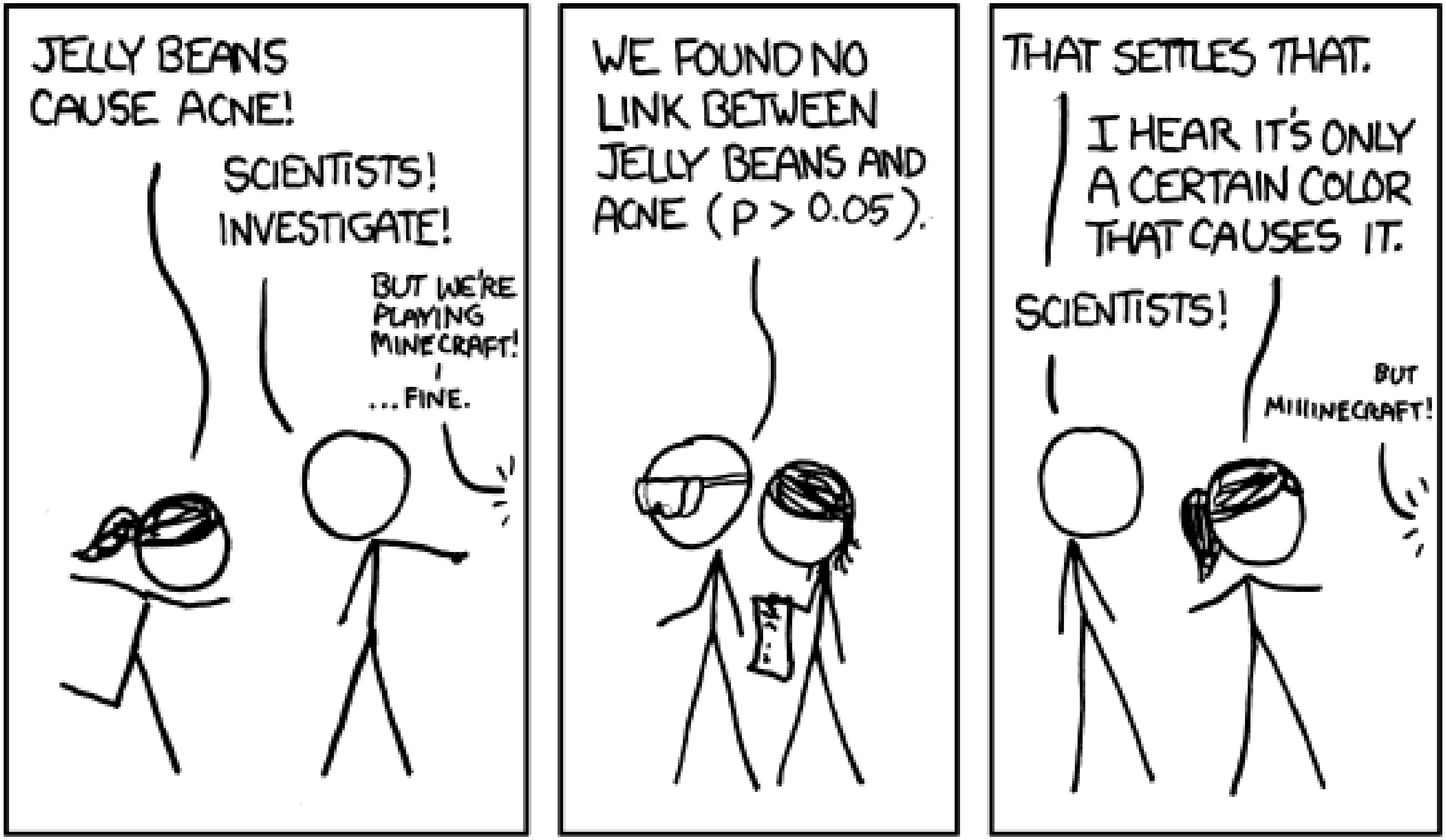
- Sometimes the required sample size for a difference in means is not very large.
- For multivariable regression, it is helpful to think of the 15:1 rule

1. Harrell Jr, F.E., 2015. Regression modeling strategies.
2. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR (1996) A simulation study of the number of events per variable in logistic regression analysis. Journal of Clinical Epidemiology 49:1373-1379.

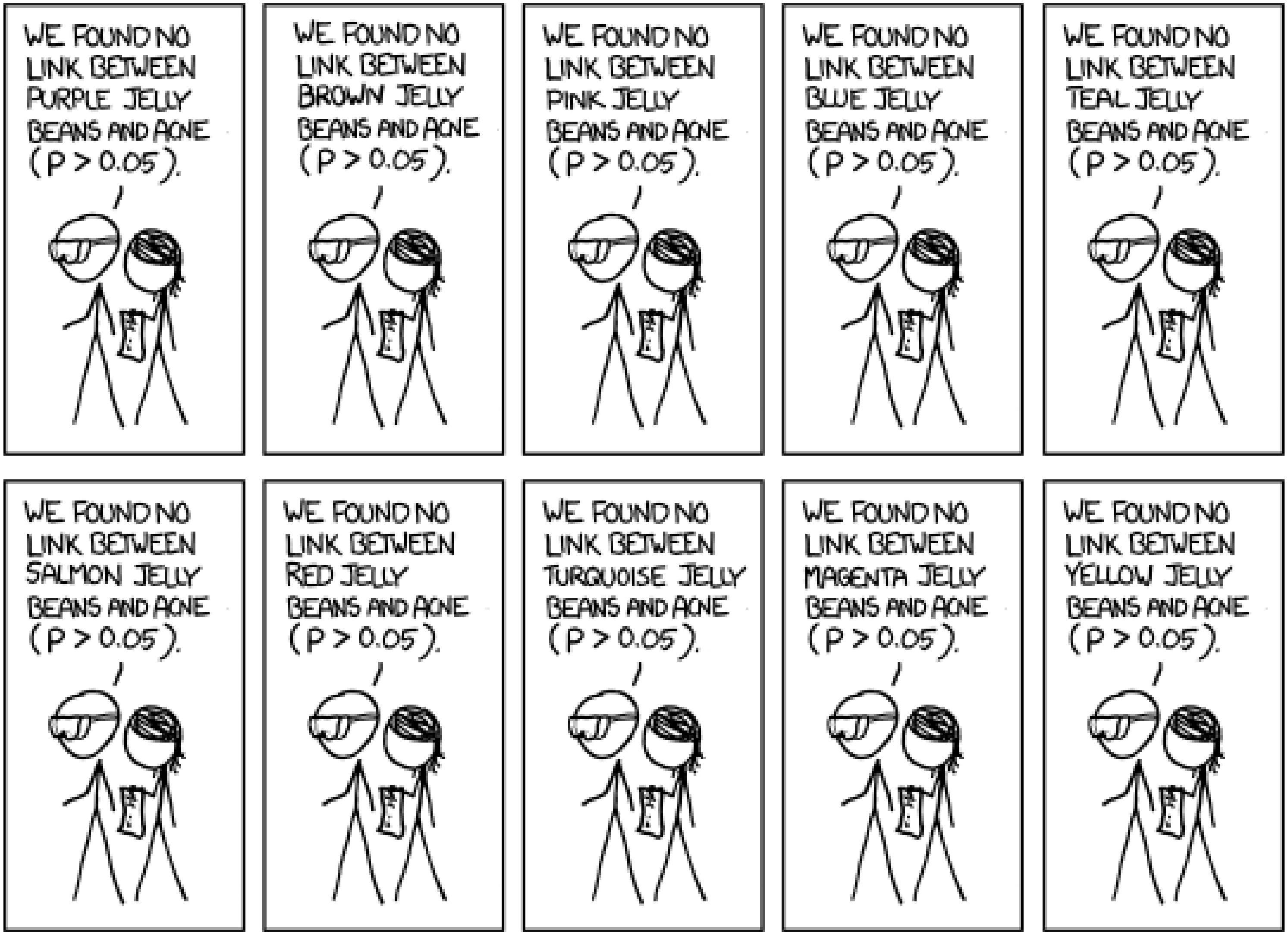
Controlling Type I error



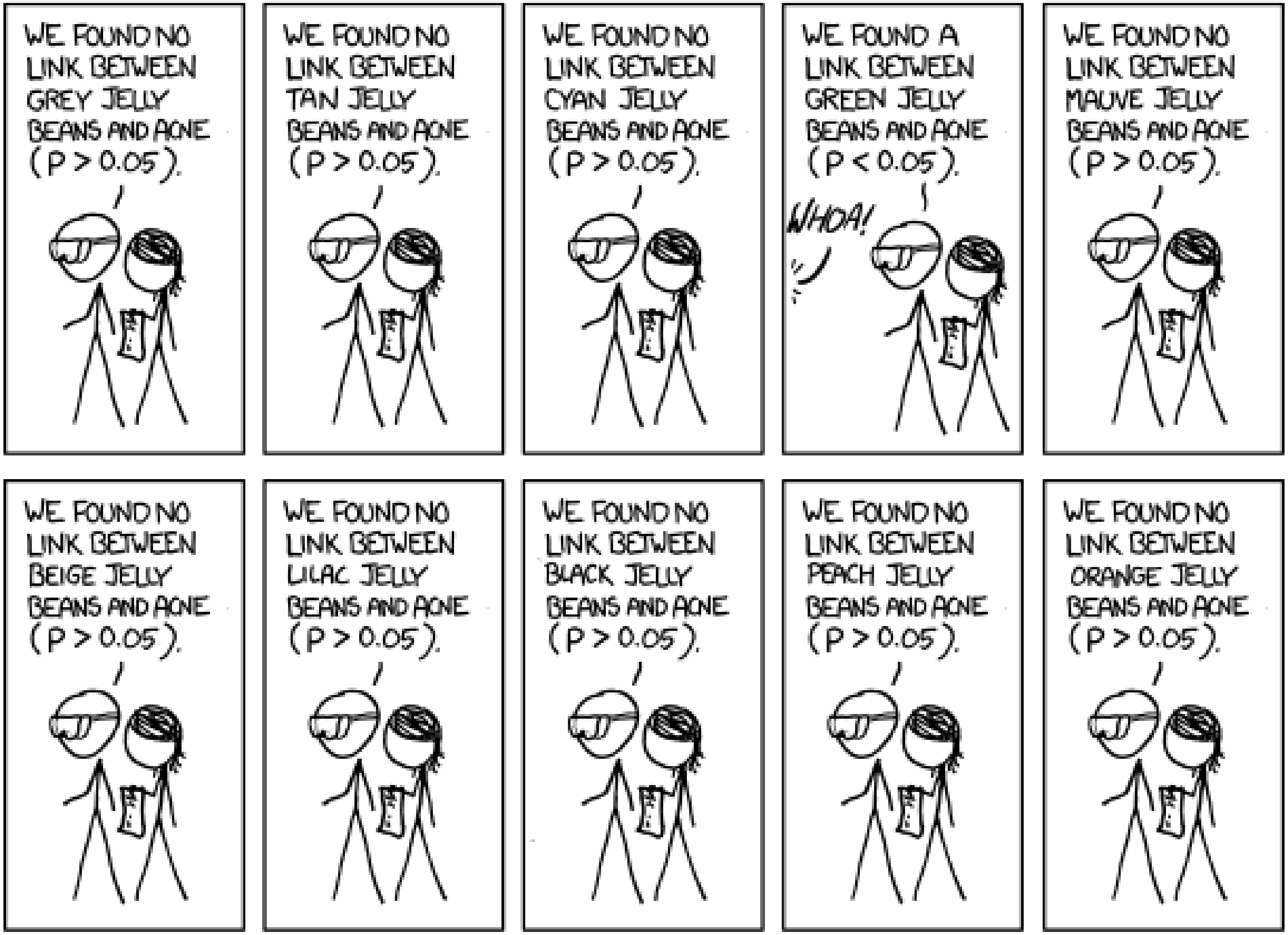
Multiplicity Considerations


JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
... FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MIIIINECRAFT!

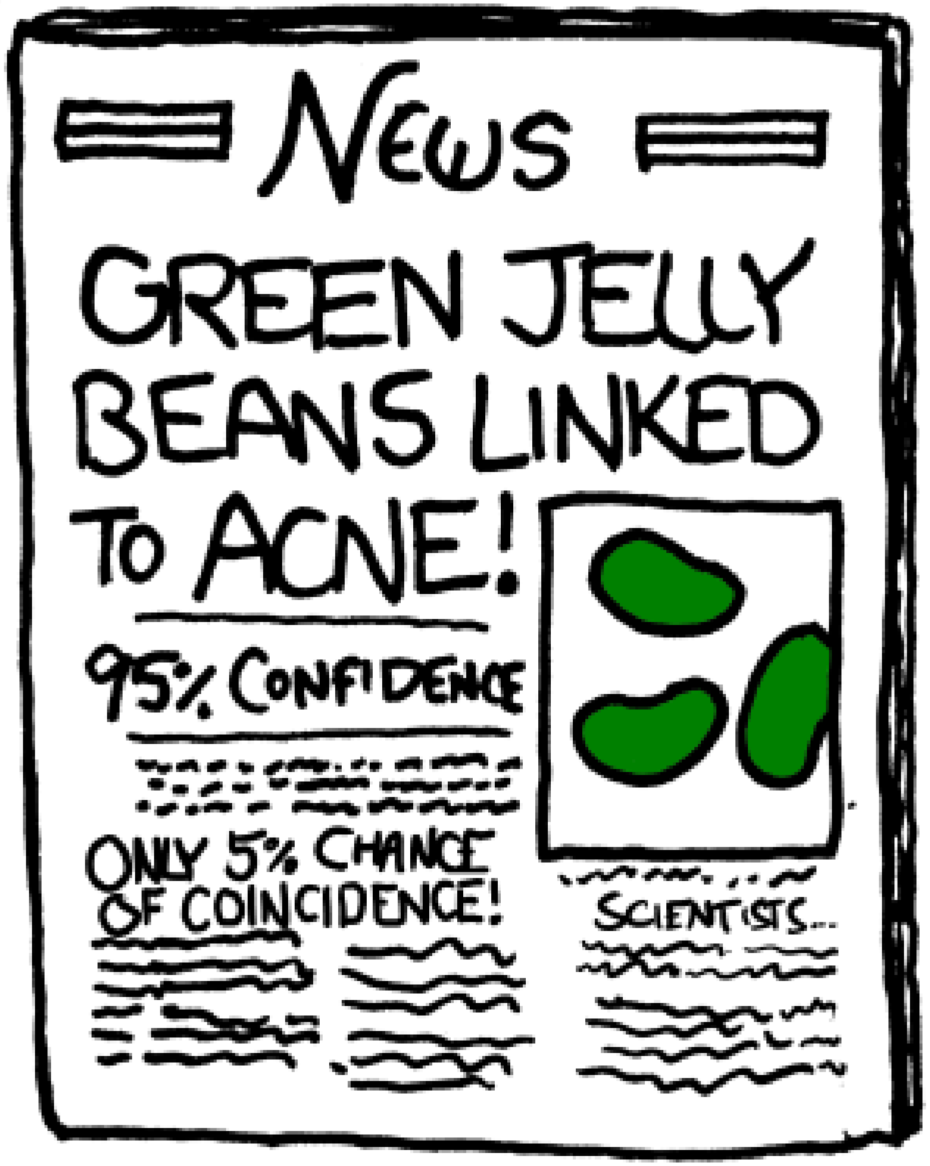
https://xkcd.com/882/


WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).

https://xkcd.com/882/


WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).

https://xkcd.com/882/



NEWS
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

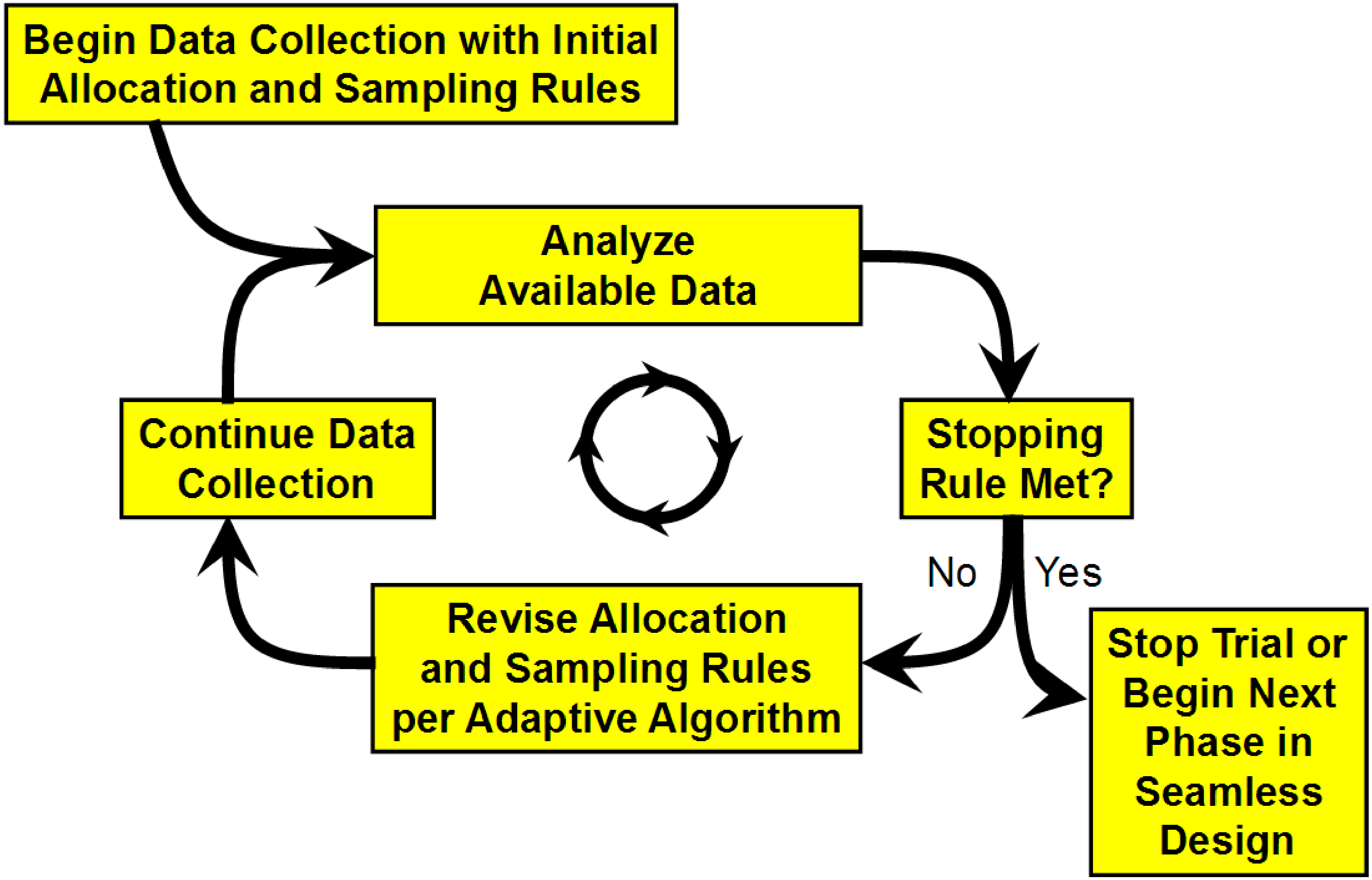
https://xkcd.com/882/

Multiplicity Considerations

Due to the potential for false positive results when many comparisons are performed, we will report test results in all abstracts, manuscripts, and presentations in a pre-determined order as ascertained by the study investigators. For example, the order of reporting results could be <<1>>, <<2>>, <<3>>, <<4>>, etc.; the final order will be selected *a priori*. Because we are interested in answering multiple questions but will report all analyses in the context of the aforementioned ordering, no adjustment for multiplicity will be performed, an approach consistent with recommendations for clinical trials.

Cook RJ, Farewell VT. Multiplicity considerations in the design and analysis of clinical trials. *J Roy Statist Assoc A* 1996;159:93-110.

Adaptive Trial Design


Begin Data Collection with Initial Allocation and Sampling Rules
Analyze Available Data
Stopping Rule Met?
No
Yes
Stop Trial or Begin Next Phase in Seamless Design
Revise Allocation and Sampling Rules per Adaptive Algorithm
Continue Data Collection

http://www.berryconsultants.com/

To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.

—Ronald Fisher (1938)

References

1. Hulley, Stephen B., et al. *Designing clinical research*. LWW, 2013.

2. PS Software

Common Values for Critical Regions

$Z_{1-\alpha/2} \approx 1.96$, when $\alpha = 0.05$
$Z_{1-\alpha/2} \approx 2.58$, when $\alpha = 0.01$
$Z_{1-\beta} \approx 0.84$, when $\beta = 0.20$
$Z_{1-\beta} \approx 1.28$, when $\beta = 0.10$

Precision of Estimate

Estimating a Proportion:

$$N = \frac{(Z^2_{1-\alpha/2})p(1-p)}{D^2}$$

Estimating a Mean:

$$N = \frac{(Z^2_{1-\alpha/2})(\sigma^2/n)}{D^2}$$

D, half-length of confidence interval (est ± D)

Detecting a Difference

Difference of (Independent) Proportions:

$$N = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2 (p_1(1-p_1) + p_2(1-p_2))}{(p_1 - p_2)^2}$$

Difference of (Independent) Means:

$$N = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2 (\sigma_1^2 + \sigma_2^2)}{(\mu_1 - \mu_2)^2}$$

Diagnostic Tests

Accuracy of One Test (same as estimating a proportion):
$N = \frac{(Z^2{}_{1-\alpha/2}) Se(1-Se)}{D^2}$
Accuracy of Two Tests[1]: The null and alternative hypotheses are:

$$H_0 : \vartheta_1 = \vartheta_2$$
$$H_a : \vartheta_1 \neq \vartheta_2$$

where ϑ_1 is the diagnostic acccuracy of test 1 and ϑ_2 is the diagnostic accuracy of test 2. The presumed value of the difference in sensitivity is denoted as Δ_1.

$$N = \frac{[Z_{1-\alpha/2}\sqrt{V_0(\hat{\vartheta}_1 - \hat{\vartheta}_2)} + Z_{1-\beta}\sqrt{V_A(\hat{\vartheta}_1 - \hat{\vartheta}_2)}]^2}{(\Delta_1)^2}$$

[1] Zhou X-H, Obuchowski NA, McClish DK. *Statistical Methods in Diagnostic Medicine*. New York, NY: Wiley; 2002.

Diagnostic Tests, *cont'd*

With a paired-study design, the variance functions under the null and alternative hypotheses are given by:

$$V_0(\hat{Se_1} - \hat{Se_2}) = \psi \qquad V_A(\hat{Se_1} - \hat{Se_2}) = \psi - {\Delta_1}^2$$

where

$$\psi = Se_1 + Se_2 - 2 \times Se_2 \times P(T_1 = 1 | T_2 = 1)$$

Se_1 and Se_2 are the presumed values of sensitivity from the alternative hypothesis and $P(T_1 = 1 | T_2 = 1)$ is the probability that the test 1 is positive given that test 2 is positive. The value of ψ ranges from Δ_1 (perfect correlation of test results) to $Se_1 \times (1 - Se_2) + (1 - Se_1) \times Se_2$ (zero correlation).

Cluster Randomized Trials

Our objective is to compare the population proportions for intervention and control groups of randomized clusters. Suppose there are n study subjects in each cluster, c.

$$c = 1 + (z_{\alpha/2} + z_{\beta})^2 [\pi_0(1-\pi_0)/n + \pi_1(1-\pi_1)/n + {k_m}^2({\pi_0}^2 + {\pi_1}^2)]/(\pi_0-\pi_1)^2$$

where π_1 and π_0 are the true proportion for intervention and control groups, and k_m is the coefficient of variation.
Clusters (eg, communities) are matched on the basis of factors that are expected to be correlated with the main study outcomes, with the aim of minimizing the degree of between-cluster variation within matched pairs. Then c, the number of clusters required, is given by:

$$c = 2 + (z_{\alpha/2} + z_{\beta})^2 [\pi_0(1-\pi_0)/n + \pi_1(1-\pi_1)/n + {k_m}^2({\pi_0}^2 + {\pi_1}^2)]/(\pi_0-\pi_1)^2$$

[1] Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. *Int J Epidemiol* 1998;**28**:319-326.

Cohort and Case-control

Depends on the outcome, see reference.

[1] Schlesselman, JJ. Sample size requirements in cohort and case-control studies of disease. *Am J Epidemiol* 1974;**9**:6:381-384.

The 10-20 Rule

A fitted regression model is likely to be reliable when the number of predictors is less than $m/10$ or $m/20$ where m is the 'limiting sample size'.

Table: Limiting Sample Sizes for Various Response Variables

Type of Response Variable	Limiting Sample Size m
Continuous	n (total sample size)
Binary	$\min(n_1, n_2)$
Ordinal (k categories)	$n - \frac{1}{n^2}\sum_{i=1}^{k} n_i^3$
Failure (survival) time	number of failures

[1] Harrell, FE. *Regression Modeling Strategies*. New York, NY: Springer; 2001.