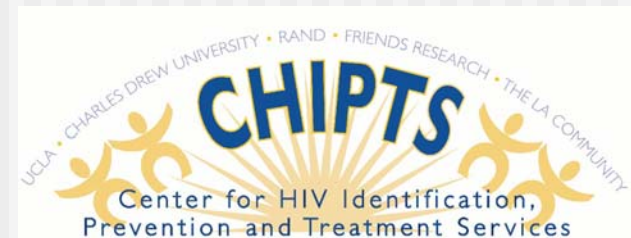




Longitudinal Sample Size Determination

Abdelmonem A. Afifi, Ph.D.
Dean Emeritus and Professor of Biostatistics
UCLA School of Public Health
afifi@ucla.edu

W. Scott Comulada, Dr.P.H.
Senior statistician, Center for Community Health
scomulad@ucla.edu



What Will I Talk About?

- Sample-size and power
- Types of study designs
- Power for one-group before-after design
- Power for two-group follow-up design
- Power for longitudinal & Cluster-randomized designs

Sample size and power

Basic principle

- Primary emphasis should be on formulating the main question...
- ...not on computations as such.

You can...

Estimate the sample size based on:

1. The *power* of a statistical test, or
2. The width of a *confidence interval*.

I will emphasize (1)

Power and Sample Size

Power = the probability of correctly rejecting the null hypothesis
= the probability that the test will detect a specified true difference (*effect size*)

Power increases as the effect size or the sample size increases

Sample Size Estimation

- Choose one of the intended analyses.
- Simplify it for the purpose of estimating the sample size.

Example: Binary Outcome

- Two-group randomized trial:
drug vs. placebo
 - Outcome: the patient is cured or not
 - Intended analysis is logistic regression with several covariates
- For power analysis, simplify to:
 - Null hypothesis: the probability of being cured is the same for the two groups
 - Alternative hypothesis: they are not

Example: Continuous Outcome

- Two-group randomized trial – new drug vs. standard
 - Reduction in SBP
 - Intended analysis is multiple linear regression with several covariates
- For power analysis, simplify to:
 - Null hypothesis: the mean SBP reduction is the same for the two groups
 - Alternative hypothesis: they are not

Factors involved in computation

- Significance level (α)
- Power ($1 - \beta$)
- Sample size
- Variability (SD = standard deviation)
- Expected mean difference
- Effect size (ES) combines variability & mean difference:
 - $ES = |mean\ difference| / SD$

Effect size issues

- Using pilot data to establish
 - Kraemer et al. (2006) – “Not recommended.”
 - Too unreliable.
- How big an effect size? – Cohen’s rules of thumb (Cohen 1988):
 - .2 = Small
 - .5 = Medium
 - .8 = Large

Example of Effect Size (ES)

- Child CBCL (Achenbach & Rescorla, 2001)
 - Normative mean = 5.2, SD = 5.8
 - ES = .50 (medium according to Cohen)
 - Equivalent to = $.50 * 5.8 = 2.6$ points on the CBCL scale

Study Designs

Three Design Types

| Manipulation of Exposure/Treatment? | Randomization? | |
|--|-----------------------|----------------------------|
| | Yes | No |
| Yes | EXPERIMENT | QUASI EXPERIMENT |
| No | — | OBSERVATIONAL STUDY |

One-group: before-after design

- A group of subjects is followed for change in outcome status;
- the outcome variable, usually continuous, is compared:
 - before versus after the intervention.
- Each subject serves as his or her own control;
 - there is no separate control group.
- Comparison of treatments is represented simply by contrasting times (before vs. after the intervention begins).

Experiments - Two-group Design: Follow up

- Subjects are followed after randomization for changes in the outcome variable, such as
 - disease occurrence, recurrence, death, clinical improvement, or complications.
- Distribution of the outcome variable is compared between assigned treatment groups (*intention-to-treat analysis*).
- If the intervention has an effect, we expect the average changes in outcome observed for the experimental and control groups to differ from each other.

Power for one-group before-after design

Continuous Outcome Example

- Y = Brief Symptom Inventory
 - Change between baseline and follow-up
- Intend to use multiple linear regression of change on several covariates
- Simplify to: paired t -test
- Power analysis in PASS (Hintze et al. 2004)

Specify values in PASS

- Use $\alpha = 0.05$, power = $1 - \beta = 0.80$
- Effect Size = 0.1 to 0.8
- Find N

Choose program

The screenshot shows the PASS 2002 Home software interface. The title bar reads "PASS 2002 Home - Power Analysis and Sample Size". The menu bar includes "Outline", "Means", "Proportions", "Regression", "Survival", "Other", "Window", and "Help". The toolbar contains various icons, including a "View Procedure" button. The main content area displays a tree view of statistical tests. The "Means" category is expanded, showing sub-categories: "One mean", "Hotelling's T-squared", and "Confidence interval of a mean". The "One mean" category is further expanded, showing four sub-items: "TT1", "E-1", "HT", and "CI-M". A blue tooltip box is overlaid on the "TT1" item, containing the text: "Test the mean of a normal distribution (t-test or nonparametric)".

Outline **Means** **Proportions** **Regression** **Survival** **Other** **Window** **Help**

B **View Procedure** [Icons]

- + Correlation
- + Diagnostic Tests (ROC Curves)
- + Equivalence and Non-Inferiority Tests
- + Group-Sequential Tests
- + Incidence Rates
- Means
 - One mean
 - TT1** Test the mean of a normal distribution (t-test or nonparametric)
 - E-1** Test the mean of an exponential distribution
 - HT** Hotelling's T-squared
 - CI-M** Confidence interval of a mean

Choose Values

The screenshot shows the PASS software interface for a 'Mean: 1 or 2 Correlated (Paired)' design. The 'Data' tab is active, displaying the following settings:

- Find (Solve For): N (Sample Size)
- Mean0 (Null or Baseline): 0
- Mean1 (Alternative): 1 to 8 by 1
- N (Sample Size): 600
- S (Std Deviation): 10
- SD: Known Standard Deviation
- Population Size: Infinite
- Alternative Hypothesis: Ha: Mean0 <> Mean1
- Nonparametric Adjustment: Ignore
- Alpha (Significance Level): 0.05
- Beta (1-Power): .2

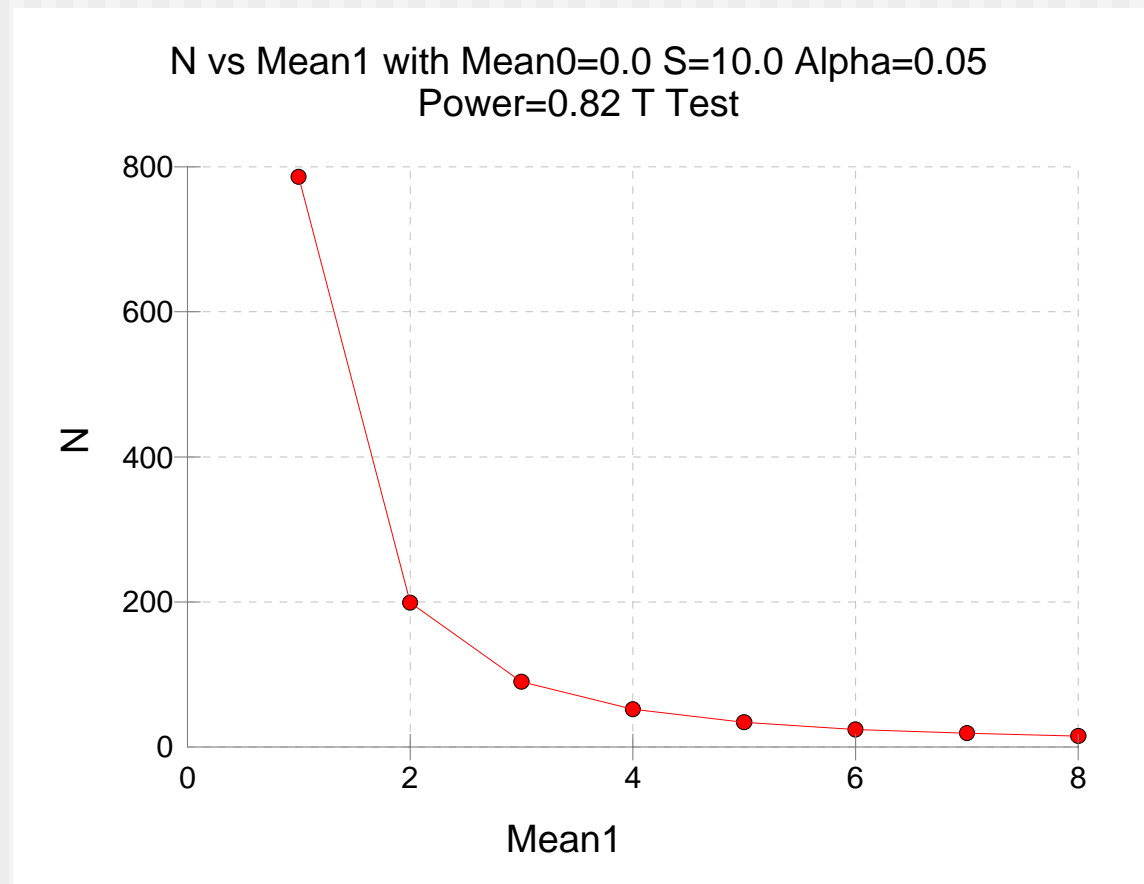
For paired designs, the data are the differences between the items of the pair (such as $X = \text{Post} - \text{Pre}$).

POPULATION SIZE: This is the number of subjects in the population. Usually, you assume that samples are drawn from a very large (infinite) population. Occasionally, however, situations arise in which the population of interest is of limited size. In these cases, appropriate adjustments must be made. Finite Population Correction Factor: When a finite population size is specified, the standard deviation is reduced according to

Template Id: _____

Buttons: Reset, Guide Me

Output



Adjusting for loss to followup

- Estimate the proportion of subjects you expect to lose contact with, say a proportion p .
- Adjust by dividing by $(1 - p)$.
- Ex: you need 400 subjects and expect 30% loss to followup ($p = 0.3$). You need to recruit
$$400/(0.7) = 571.4 \text{ subjects,}$$

→ round up to 572 subjects.

Categorical Outcome Example

- Y = proportion of protected sex acts
 - Change between baseline and follow-up
- Intend to use multiple logistic regression of change on several covariates
- Simplify to : McNemar's test
 - Assume outcome is dichotomous
 - 100% protected sex (Yes/No)
- Power analysis in PASS

Data setup

Engaged in 100% protected sex

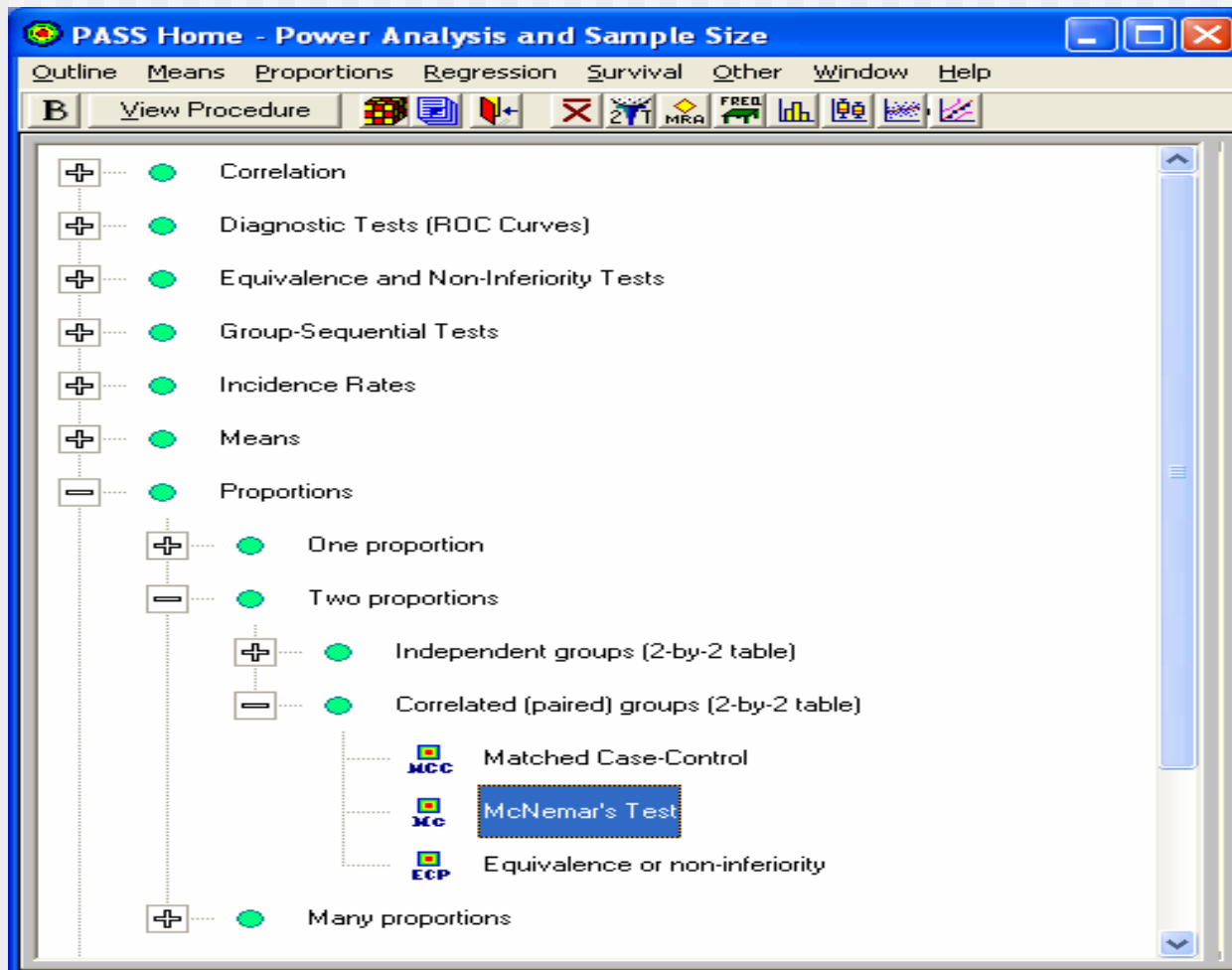
| | | Follow-up | |
|----------|-----|-----------|----------|
| | | No | Yes |
| Baseline | No | N_{11} | N_{12} |
| | Yes | N_{21} | N_{22} |

N

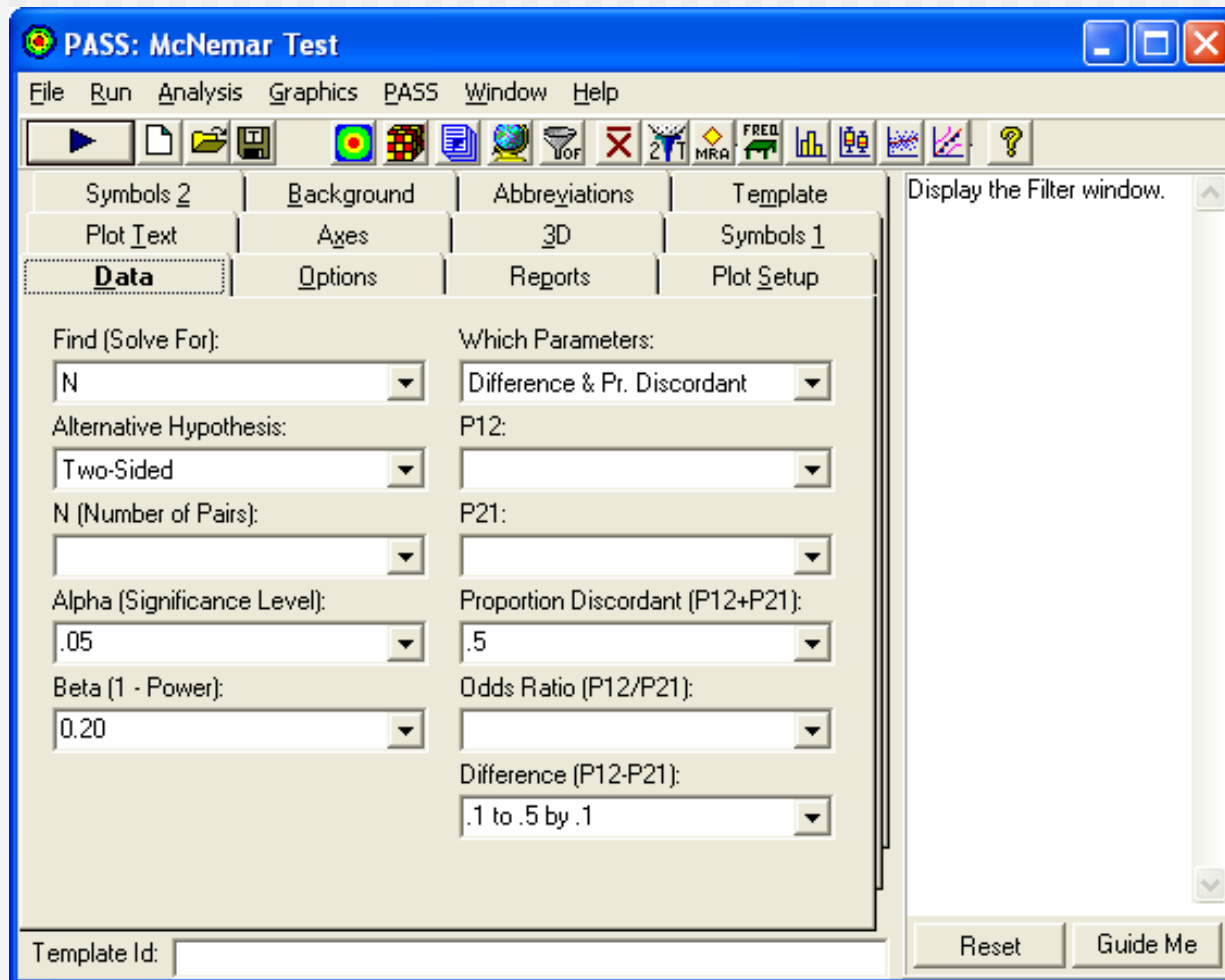
Specify values in PASS

- Use $\alpha = 0.05$, power = $1 - \beta = 0.80$
- Proportion of discordant pairs = .50
 - $= (N_{12} + N_{21}) / N$
 - Proportion of sample changing between baseline and follow-up
- Difference in proportion of discordant pairs = 0.1 to .5
- Find N

Choose program

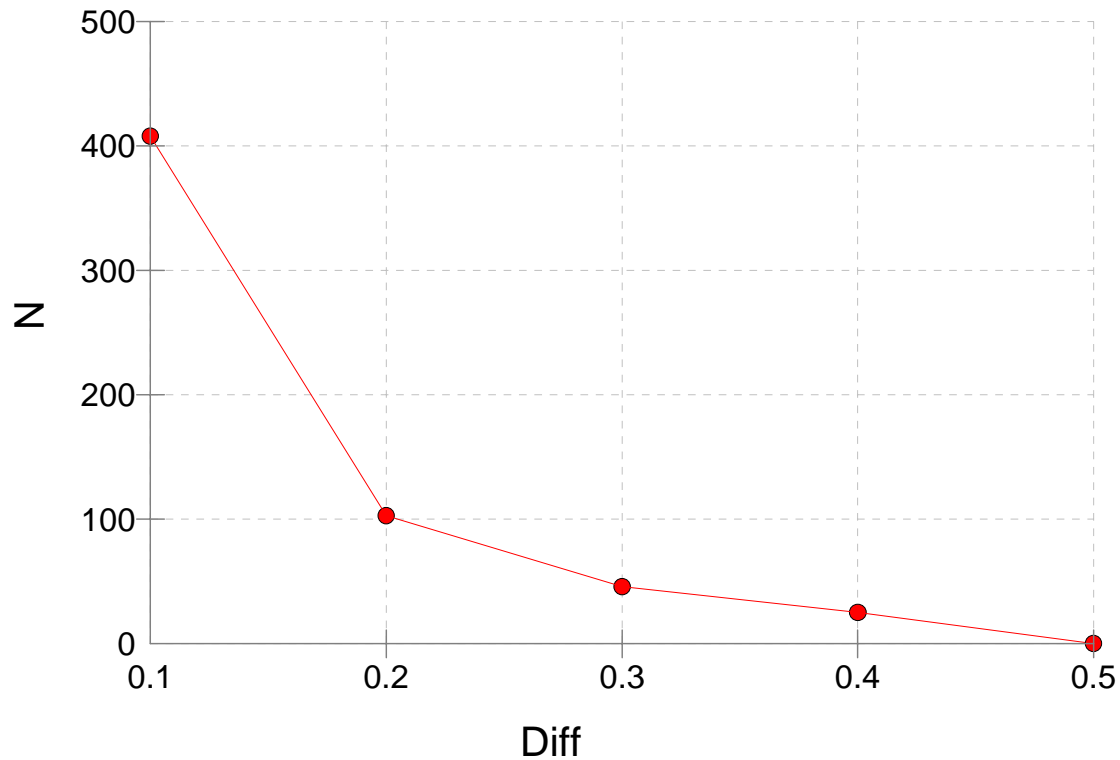


Choose Values



Output

N vs Diff with P.D.=0.500 Alpha=0.050 2-Sided McNemar Test



Longitudinal two-group design

Two situations

1. Longitudinal two-group design
 - individual randomized
2. Longitudinal two-group design
 - cluster randomized

Example for individual-randomized design

- Y = Brief Symptom Inventory (BSI)
 - 4 time points
- Intend to use mixed-effect regression to test effect of 2-arm intervention on BSI, adjusting for demographics
- Simplify to: change in slope of mean BSI over time between 2 intervention groups
- Power analysis in RMASS2 (Hedeker et al. 1999)

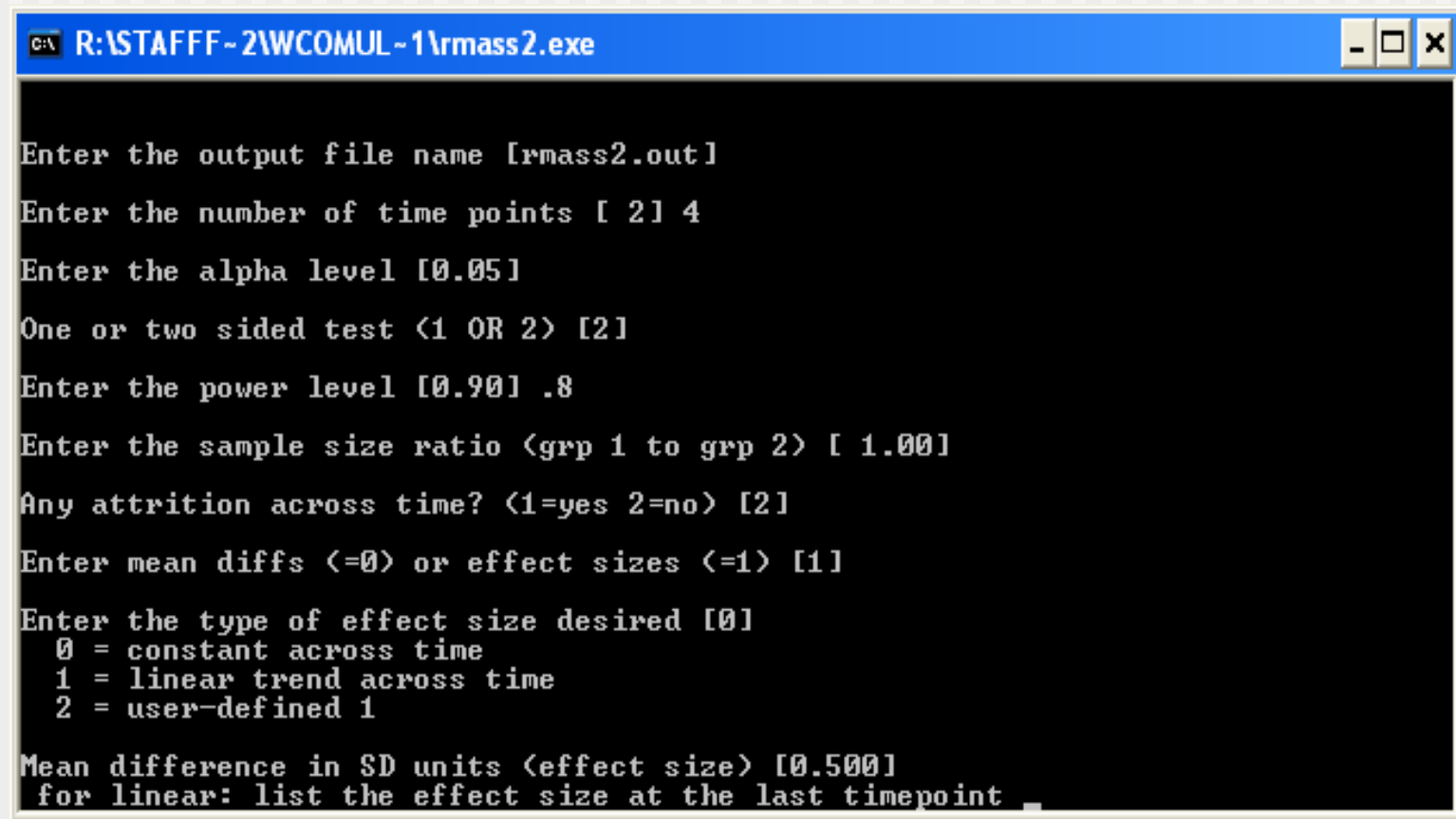
Specify values in RMASS2

- Use $\alpha = 0.05$, power = $1 - \beta = 0.80$
- Effect Size = .4
- Time points = 4
- Sample size ratio = 1
- Attrition (drop out)
 - Specified as a percent
 - Assume none for example

Specify values (continued)

- Type of correlation between repeated observations,
 - Specify random effect or autoregressive structure and specify rho (higher = more power)
- Mean difference or effect size
 - Effect size easiest, set SD to 1 and ES, e.g., .40
 - or, set SD = 10, ES = 4
- Additional specifications shown in screen shots
- Find N for each intervention arm

Choose values



```
C:\R:\STAFF-2\WCOMUL-1\rmass2.exe

Enter the output file name [rmass2.out]
Enter the number of time points [ 2] 4
Enter the alpha level [0.05]
One or two sided test <1 OR 2> [2]
Enter the power level [0.90] .8
Enter the sample size ratio <grp 1 to grp 2> [ 1.00]
Any attrition across time? <1=yes 2=no> [2]
Enter mean diffs <=0> or effect sizes <=1> [1]
Enter the type of effect size desired [0]
  0 = constant across time
  1 = linear trend across time
  2 = user-defined 1
Mean difference in SD units <effect size> [0.500]
for linear: list the effect size at the last timepoint _
```

Choose Values (continued)

```
C:\ R:\STAFF-2\WCOMUL-1\rmass2.exe
for linear: list the effect size at the last timepoint .4
Variance-covariance structure of repeated measures
no random effects (<=0) or with random effects (<=1) [0]
Enter the standard deviation at timepoint 1 [ 1.000]
Enter the standard deviation at timepoint 2 [ 1.000]
Enter the standard deviation at timepoint 3 [ 1.000]
Enter the standard deviation at timepoint 4 [ 1.000]
Enter correlation structure of repeated measures [1]
 1 = all correlations equal
 2 = stationary AR1
 3 = non-stationary AR1
 4 = toeplitz (banded) matrix 2
Enter correlation term number 1 [0.500]
Enter the type of time-related contrast desired [0]
0 = average across time; 1 = linear trend; 2 = user-defined
```

Choose Values (continued)

```
C:\ R:\STAFFF-2\WCOMUL-1\rmass2.exe

Enter the standard deviation at timepoint 2 [ 1.000]
Enter the standard deviation at timepoint 3 [ 1.000]
Enter the standard deviation at timepoint 4 [ 1.000]

Enter correlation structure of repeated measures [1]
 1 = all correlations equal
 2 = stationary AR1
 3 = non-stationary AR1
 4 = toeplitz (banded) matrix 2

Enter correlation term number 1 [0.500]

Enter the type of time-related contrast desired [0]
0 = average across time; 1 = linear trend; 2 = user-defined 1

Composite Effect size = 0.300023
N Subj for Grp1 Time 1 = 174.443945

Exit the program <1=Y OR 2=N> [2] ?
```

N = 175 in each intervention arm

Advantages and Drawbacks of RMASS2

■ Advantages:

- Allows attrition
- Several IC Correlations
- Free and downloadable:
 - <http://tiger.uic.edu/~hedeker/ml.html>

■ Draw backs

- Assumes two treatment groups only
- Assumes equally spaced time points

Example for cluster-randomized design

- Start with previous example
 - Effect of 2-arm intervention on BSI
 - 4 time points, $ES = .40$
 - $N = 175$ per intervention arm
- Now assume
 - Study conducted on patients from clinics
 - Intervention assigned at clinic level
- Find the number of clinics needed

Two options to calculate power

1. *Use software that ignores clustering (e.g. RMASS2) and adjust calculated sample size using variance inflation formula*
 2. Use software that accounts for longitudinal data and clustering
 - e.g. PASS, within and between design
- We discuss 1. further

Variance inflation formula

- When the treatments are randomized to groups (clusters), we must adjust the sample size.
- Let $M = \#$ of clusters, and
 $B = \#$ of subjects per cluster,
 $\rho =$ intraclass correlation coefficient.
- Define DEFF = design effect
 $= 1 + \rho(B-1)$, and
 $N^* =$ nominal sample size $= MB$
- For purpose of computing power, the effective (true) sample size is actually:
 Effective $N = N^*/DEFF$

Variance inflation formula (cont.)

- Example 1: $M = 30$ clinics,
 $B = 10$ subjects per clinic,
 $\rho = 0.01$

Then $DEFF = 1 + (10 - 1)(0.01) = 1.09$, and
effective $N = 300/1.09 = 275$.

- Example 2: $B = 10$ subjects per clinic,
 $\rho = 0.01$

Suppose we compute that we need $N = 175$ per group as
in example above. Then the actual sample size should be:

$$\text{actual } N = (DEFF)(N^*) = 1.09 * 175 = 191$$

or approximately 19 clinics per group.

Extensions

- *PASS Repeated Measures ANOVA*
 - Allows more factors than RMASS2
 - Several autocorrelations
 - Interactions
 - Specification of mean structure
- Drawback
 - Difficult to use

Simulations

- Good option for more complex designs
- Can use STATA for computations
- Additional research being carried out

References

- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington: University of Vermont Research Center for Children, Youth, & Families.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Erlbaum.
- Hedeker, D., Gibbons, R.D., & Waternaux, C. (1999). Sample size estimation for longitudinal designs with attrition: comparing time-related contrasts between two groups. *Journal of Educational and Behavioral Statistics*, 24:70-93.
- Hintze, J. (2004). NCSS and PASS. Number Cruncher Statistical Systems. Kaysville, Utah. www.ncss.com.
- Kraemer HC, Mintz J, Noda A, Tinklenberg J, Yesavage JA. Caution regarding the use of pilot studies to guide power calculations for study proposals. *Arch Gen Psychiatry*. 2006 May;63(5):484-9.

Thank you very much,

Abdelmonem A. Afifi
afifi@ucla.edu

Scott Comulada
scomulad@ucla.edu