Sample Size Estimation for Longitudinal Studies

Don Hedeker
University of Illinois at Chicago

www.uic.edu/∼hedeker

Hedeker, Gibbons, & Waternaux (1999). Sample size estimation for longitudinal designs with attrition. *Journal of Educational and Behavioral Statistics*, 24:70-93

Comparison of two groups at a single timepoint

Number of subjects (N) in each of two groups (Fleiss, 1986):

$$N = \frac{2(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_1 - \mu_2)^2} = \frac{2(z_\alpha + z_\beta)^2}{[(\mu_1 - \mu_2)/\sigma]^2}$$

- z_α is the value of the standardized score cutting off $\alpha/2$ proportion of each tail of a standard normal distribution (for a two-tailed hypothesis test)
- z_β is the value of the standardized score cutting off the upper β proportion
- σ^2 is the assumed common variance in the two groups
- $\mu_1 - \mu_2$ is the difference in means of the two groups

Some common choices:

- $z_\alpha = 1.645, 1.96, 2.576$ for 2-tailed .10, .05, and .01 test
- $z_\beta = .842, 1.036, 1.282$ for power = .8, .85. and .90
- effect size $= (\mu_1 - \mu_2)/\sigma = .2, .5, .8$ for "small," "medium," and "large" effects (Cohen, 1988)

Example

- $z_\alpha = 1.96$ 2-tailed .05 hypothesis test
- $z_\beta = .842$ power = .8
- effect size $(\mu_1 - \mu_2)/\sigma = .5$

$$N \;=\; \frac{2(1.96 + .842)^2}{(.5)^2} \;=\; 15.7/.25 \;=\; 62.8$$

$\Rightarrow$ need 63 subjects in each group

Rule of thumb: $N \approx (4/\delta)^2$, where δ = effect size
(for power = .8 for a 2-tailed .05 test)

effect size (δ)	N per group	$(4/\delta)^2$
.1	1571	1600
.2	394	400
.3	176	178
.4	100	100
.5	64	64
.6	45	44
.7	34	33
.8	26	25
.9	21	20
1.0	17	16

Amaze your friends with your sample size determination abilities!

Comparison of two groups across time

consistent difference across time

Number of subjects N in each of two groups (Diggle *et al.*, 2002)

$$N = \frac{2(z_\alpha + z_\beta)^2 \, (1 + (n-1)\rho)}{n[(\mu_1 - \mu_2)/\sigma]^2}$$

- σ^2 is the assumed common variance in the two groups
- $\mu_1 - \mu_2$ is the difference in means of the two groups
- n is the number of timepoints
- ρ is the assumed correlation of the repeated measures

Example

- $z_\alpha = 1.96$ 2-tailed .05 hypothesis test
- $z_\beta = .842$ power $= .8$
- effect size $(\mu_1 - \mu_2)/\sigma = .5$
- $n = 2$ timepoints
- $\rho = .6$ correlation of repeated measures

$$N = \frac{2(1.96 + .842)^2(1 + (2-1) \times .6)}{2 \times (.5)^2} = \frac{(15.7)(1.6)}{(2)(.25)} = 50.3$$

$\Rightarrow$ need approximately 50 subjects in each group

if $\rho = 0$ then $N = 31.4$ (cross-sectional)
if $\rho = 1$ then $N = 62.8$ (one-timepoint)

SAS code

```
* determines number per group;
* 5 timepoints (ICC=.4);
* effect size of .5;
* power = .8 for a 2-tailed .05 test;

DATA one;
n = 5;
za = PROBIT(.975);
zb = PROBIT(.8);
rho = .4;
effsize = .5;
num = (2*(za + zb)**2)*(1 + (n-1)*rho);
den = n*(effsize**2);
npergrp = num/den;
PROC PRINT;VAR npergrp;
RUN;
```

Comparing two groups across timepoints - balanced case

As in Overall and Doyle (1994), sample size of contrast Ψ_c of group population means across n timepoints:

$$N = \frac{2(z_\alpha + z_\beta)^2 \sigma_c^2}{\Psi_c^2}$$

with

$$\Psi_c = \sum_{i=1}^{n} c_i(\mu_{1i} - \mu_{2i})$$

$$\sigma_c^2 = \sum_{i=1}^{n} c_i^2 \sigma_i^2 + 2 \sum_{i<j}^{n} c_i c_j \sigma_{ij}$$

- σ_i^2 = common variance in the two groups at timepoint i
- σ_{ij} = common covariance in the two groups between timepoints i and j
- c_i = contrast applied at timepoint i

If the sample size is known and the degree of power is to be determined, the formula can be re-expressed as:

$$z_\beta = \sqrt{\frac{N\Psi_c^2}{2\sigma_c^2}} - z_\alpha = \sqrt{\frac{\Psi_c^2}{V(\hat{\Psi}_c)}} - z_\alpha$$

where the variance of the sample contrast $\hat{\Psi}_c$ equals

$$V(\hat{\Psi}_c) = \frac{2}{N}\sigma_c^2$$

Example

- $z_\alpha = 1.96$ 2-tailed .05 hypothesis test
- $z_\beta = .842$ power $= .8$
- $n = 2$ timepoints
- variance-covariance of repeated measures

$$V(y) = \begin{bmatrix} 1 & .6 \\ .6 & 1 \end{bmatrix}$$

I. Average group difference over time

- mean difference $\mu_1 - \mu_2 = .5$ at both $t1$ and $t2$
- time-related contrasts: $c_1 = c_2 = 1/2$ (*i.e.*, average over time)

$$\Psi_c = \frac{1}{2}(.5) + \frac{1}{2}(.5) \;=\; .5$$

$$\sigma_c^2 = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + 2\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)(.6) \;=\; .8$$

contrast effect size $\delta = \Psi_c/\sigma_c = .5/\sqrt{.8} = .56$

$$N \;=\; \frac{2(1.96 + .842)^2}{(.56)^2} \;=\; 50$$

Notice

- if $\rho = 1$, then $\sigma_c^2 = 1,\ \delta = .5,\ N = 63 \quad$ (one-timepoint)

- if $\rho = 0$, then $\sigma_c^2 = 1/2,\ \delta = 1,\ N = 16 \quad$ (cross-sectional)

where ρ is the assumed correlation of the repeated measures

II. Group difference across time

- mean difference $\mu_1 - \mu_2 = 0$ at $t1$ and .5 at $t2$
- time-related contrasts: $c_1 = -1$ and $c_2 = 1$

$$\Psi_c \; = \; -1(0) + 1(.5) \; = \; .5$$

$$\sigma_c^2 \; = \; (-1)^2(1)^2 + (1)^2(1)^2 + 2(-1)(1)(.6) \; = \; .8$$

contrast effect size $\delta = \Psi_c/\sigma_c = .5/\sqrt{.8} = .56$

$$N \; = \; \frac{2(1.96 + .842)^2}{(.56)^2} \; = \; 50$$

Notice

- if N was calculated based on $t2$ only, then $N = 63$

  $H_0 : \mu_{12} = \mu_{22} \quad \neq \quad H_0 : (\mu_{12} - \mu_{11}) = (\mu_{22} - \mu_{21})$

- if $\rho = 1$, then $\sigma_c^2 = 0$

- if $\rho = .9$, then $\sigma_c^2 = .2,\ \delta = 1.12,\ N = 14$

- if $\rho = 0$, then $\sigma_c^2 = 1/2,\ \delta = .25,\ N = 63$ cross-sectional

For average group effect over time

- as $\rho \uparrow$, then $N \uparrow$

since it's a between-subjects comparison of averages

$\Rightarrow$ less subjects needed if the averages are based on more independent data

For group difference across time

- as $\rho \uparrow$, then $N \downarrow$

since it's a between-subjects comparison of a within-subjects comparison

$\Rightarrow$ less subjects needed if the subject differences (i.e., pre to post) are based on more reliable data

More than 2 timepoints

- mean differences across time
- var-covar and/or correlation of repeated measures
- time-related contrast

3 timepoints

$t1$	$t2$	$t3$	
1/3	1/3	1/3	average across time
-1	0	1	linear trend
1	-2	1	quadratic trend

trend coefficients from tables of orthogonal polynomials

4 timepoints

$t1$	$t2$	$t3$	$t4$	
1/4	1/4	1/4	1/4	average across time
-3	-1	1	3	linear trend
1	-1	-1	1	quadratic trend
-1	3	-3	1	cubic trend

often investigators expect

- overall group difference, or
- group by (approximately) linear time interaction

SAS IML code

```
* determines number per group;
* 3 timepoints;
* linear increasing effect sizes of 0 .25 .5;
* group by linear contrast across time;
* AR1 structure with rho=.5;
* power = .8 for a 2-tailed .05 test;

PROC IML;
za = PROBIT(.975);
zb = PROBIT(.8);
meandiff = {0, .25, .5};
contrast = {-1, 0, 1};
 corrmat = { 1 .5 .25 ,
            .5  1  .5 ,
           .25 .5   1 };
contdiff = T(contrast) * meandiff;
contvar = T(contrast)*corrmat*contrast;
NperGrp = ((2*(za+zb)**2) * contvar)/(contdiff**2);
PRINT NperGrp;
```

What about Attrition?

- could use N from calculations as N for last timepoint
  - e.g., $N = 50$, retention at last timepoint $= .9$
    $\Rightarrow$ start the study with $50/.9 = 56$ subjects

- can build the retention rate information into the sample size formula
  - Hedeker, Gibbons, & Waternaux (1999), *JEBS*, 24:70-93
    program RMASS2 available at www.uic.edu/∼hedeker
    (selected publications link)

Comparing two groups across timepoints

unbalanced case

Denote sample size in first group as N_{1i} and second group as N_{2i} at timepoint i ($i = 1, \ldots, n$). The variance of the sample contrast $\hat{\Psi}_c$ equals

$$\begin{aligned} V(\hat{\Psi}_c) = & \sum_{i=1}^{n} c_i^2 \sigma_i^2 \left(\frac{1}{N_{1i}} + \frac{1}{N_{2i}} \right) \\ & + 2 \sum_{i<j}^{n} c_i c_j \sigma_{ij} \left(\frac{1}{\sqrt{N_{1i} N_{1j}}} + \frac{1}{\sqrt{N_{2i} N_{2j}}} \right) \end{aligned}$$

Notice, that if $N_{1i} = N_{2i} = N$, then

$$V(\hat{\Psi}_c) = \frac{2}{N} \sum_{i=1}^{n} c_i^2 \sigma_i^2 + 2 \sum_{i<j}^{n} c_i c_j \sigma_{ij} \qquad \text{as before}$$

The current formulation is fine to calculate power, given the varying group sample sizes across time, for the sample contrast:

$$z_\beta = \sqrt{\frac{\Psi_c^2}{V(\hat{\Psi}_c)}} - z_\alpha$$

However, to figure out the necessary group sample sizes given power, more work is needed since these (N_{1i} and N_{2i}) vary across time in the equation for $V(\hat{\Psi}_c)$:

$$\begin{aligned} V(\hat{\Psi}_c) \; = \; & \sum_{i=1}^{n} c_i^2 \sigma_i^2 \left(\frac{1}{N_{1i}} + \frac{1}{N_{2i}} \right) \\ & +2 \sum_{i<j}^{n} c_i c_j \sigma_{ij} \left(\frac{1}{\sqrt{N_{1i} N_{1j}}} + \frac{1}{\sqrt{N_{2i} N_{2j}}} \right) \end{aligned}$$

Use sample size in first group at first timepoint (N_{11}) as a reference

- define retention rates for this group as r_{1i} for timepoints $i = 1, \ldots, n$, which indicate the proportion of N_1 subjects observed at timepoint i
(note that $r_{11} = 1$ and $N_{1i} = r_{1i} N_{11}$)
- similarly, define N_{21} and r_{2i} for group two

Then,

$$V(\hat{\Psi}_c) = \frac{1}{N_{11}} \left[\sum_{i=1}^{n} c_i^2 \sigma_i^2 \left(\frac{1}{r_{1i}} + \frac{1}{r_{2i}} \frac{N_{11}}{N_{21}} \right) \right.$$
$$\left. + 2 \sum_{i<j}^{n} c_i c_j \sigma_{ij} \left(\frac{1}{\sqrt{r_{1i} r_{1j}}} + \frac{N_{11}}{N_{21}} \frac{1}{\sqrt{r_{2i} r_{2j}}} \right) \right]$$

and, denoting the ratio of sample sizes at the first timepoint (N_{11}/N_{21}) as $N_{.1}$, then

$$
\begin{aligned}
V(\hat{\Psi}_c) = {} & \frac{1}{N_{11}}\left[\sum_{i=1}^{n} c_i^2\sigma_i^2\left(\frac{1}{r_{1i}}+\frac{N_{.1}}{r_{2i}}\right)\right. \\
& \left. +2\sum_{i<j}^{n} c_i c_j \sigma_{ij}\left(\frac{1}{\sqrt{r_{1i}r_{1j}}}+N_{.1}\frac{1}{\sqrt{r_{2i}r_{2j}}}\right)\right]
\end{aligned}
$$

If the retention rates are equal for the two groups across time $r_{1i} = r_{2i} = r_i$, then

$$\begin{aligned} V(\hat{\Psi}_c) &= \frac{N_{.1}+1}{N_{11}} \left[\sum_{i=1}^{n} \frac{c_i^2 \sigma_i^2}{r_i} + 2 \sum_{i<j}^{n} \frac{c_i c_j \sigma_{ij}}{\sqrt{r_i r_j}} \right] \\ &= \frac{N_{.1}+1}{N_{11}} \, \sigma_{rc}^2 \end{aligned}$$

where σ_{rc}^2 extends σ_c^2 given earlier, namely

$$\sigma_c^2 = \sum_{i=1}^{n} c_i^2 \sigma_i^2 + 2 \sum_{i<j}^{n} c_i c_j \sigma_{ij}$$

for the case where sample sizes vary across timepoints (although group retention rates are assumed equal)

To calculate power for any of the above variance formulations of the sample contrast,

$$z_\beta = \sqrt{\frac{\Psi_c^2}{V(\hat{\Psi}_c)}} - z_\alpha$$

In particular, for the case of common retention rates across time

$$z_\beta = \sqrt{\left(\frac{N_{11}}{N_{.1} + 1}\right) \frac{\Psi_c^2}{\sigma_{rc}^2}} - z_\alpha$$

where $N_{.1}$ is the sample size ratio between groups

Re-expressing, the number of subjects needed in the first group at the first timepoint equals:

$$N_{11} = \frac{(N_{.1} + 1)(z_\alpha + z_\beta)^2 \sigma_{rc}^2}{\Psi_c^2}$$

Based on

- sample size ratio between groups $N_{.1}$ at the first timepoint
- equal retention rates r_i across time

$\Rightarrow$ required sample size at each timepoint for both groups can be calculated in a relatively simple way

RMASS2: Repeated Measures with Attrition: Sample Sizes for 2 Groups

Donald Hedeker and Suna Barlas

- Calculates sample size for a 2-group repeated measures design
- Allows for attrition and a variety of variance-covariance structures for the repeated measures
- Details on the methods can be found in Hedeker, Gibbons, and Waternaux (1999, *Journal of Educational and Behavioral Statistics*, 24:70-93)
- Program runs at the "Command Prompt" and the user is queried for program parameters
- For each query, the default parameter value is given in [ ]; hitting a carriage return sets the parameter to the default

Program Parameters

fout - output file name

n - number of timepoints (maximum is 20)

alpha - alpha level for statistical test (possible values = .01, .05, .10)

nside - sided test (1 or 2)

beta - level of power (from .5 to .95 in multiples of .05)

ratio - ratio of sample sizes (group 1 to group 2)

attrit - attrition across time (1=yes, 2=no)

- *if attrit=1* - attrition rates between adjacent timepoints (assumed equal for both groups)

mtype - type of expected group differences (0=means, 1=effect sizes)

- *if mtype=0* - expected difference in group means at each timepoint
- *if mtype=1* - estype - effect size type (0=constant, 1=linear trend, 2=user-defined)
  - *if estype=0* - expected effect size (equal across time)
  - *if estype=1* - expected effect size at last timepoint
  - *if estype=2* - expected effect size at each timepoint

vtype - variance-covariance structure of repeated measures

- *if vtype=0* (no random effects) $\boldsymbol{\Sigma}_y = \sigma_j^2 \boldsymbol{R}, \quad j = 1, ...n$ timepoints
  - standard deviation at each timepoint σ_j
  - correlation structure of repeated measures ($\boldsymbol{R}$: 1=all correlations equal, 2=stationary AR1; 3=non-stationary AR1; 4=toeplitz or banded matrix)
- *if vtype=1* (random-effects structure) $\boldsymbol{\Sigma}_y = \boldsymbol{X}\boldsymbol{\Sigma}_v\boldsymbol{X}' + \sigma^2\boldsymbol{\Omega}$
  - nr = number of random effects (maximum is 4)
  - random-effects variance-covariance matrix $\boldsymbol{\Sigma}_v$
  - random-effects design matrix $\boldsymbol{X}$ ($n \times nr$ elements)
  - error variance σ^2 and autocorrelated error structure $\boldsymbol{\Omega}$

contrast - type of time-related contrast for statistical test
(0=average across time, 1=linear trend, 2=user-defined)

- *if contrast=2* - contrast coefficient at each timepoint
- this selection should generally match the effect size type selected

Example

- $z_\alpha = 1.96$ 2-tailed .05 hypothesis test
- $z_\beta = .842$ power = .8
- $n = 2$ timepoints, retention rates $r_1 = 1$ and $r_2 = .8$
- sample size ratio $N_{.1} = 1$
- variance-covariance of repeated measures

$$V(y) = \begin{bmatrix} 1 & .6 \\ .6 & 1 \end{bmatrix}$$

I. Average group difference over time

- mean difference $\mu_1 - \mu_2 = .5$ at both $t1$ and $t2$
- time-related contrasts: $c_1 = c_2 = 1/2$

$$\Psi_c \;=\; \frac{1}{2}(.5) + \frac{1}{2}(.5) \;=\; .5$$

$$\sigma^2_{rc} \;=\; \left(\frac{1}{2}\right)^2 + \frac{(1/2)^2}{.8} + 2\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)(.6)/\sqrt{.8} \;=\; .9$$

contrast effect size $\delta = \Psi_c/\sigma_{rc} = .5/\sqrt{.9} = .53$

$$N_{11} \;=\; \frac{2(1.96 + .842)^2}{(.53)^2} \;=\; 56.4$$

Note:if $r_2 = 1$ then $N_{11} = 50$

```
Command Prompt - rmass2

Enter the output file name [rmass2.out]

Enter the number of time points [ 2]

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] 1

 Enter the attrition rate between timepoints  1 and  2 [0.050] .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

Mean difference in SD units (effect size) [0.500]
 for constant: list the eff. size that is assumed equal across time

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

 Enter correlation term number  1 [0.500] .6
```

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

Composite Effect size (without attrition)      =     0.559017
N Subj for Grp1 Time 1 (without attrition)     =    50.247707

Composite Effect size (adjusted for attrition) =     0.527659
N Subj for Grp1 Time 1 (adjusted for attrition)=    56.397411


Exit the program (1=Y OR 2=N) [2] ?
```

RMASS2.OUT - Notepad

File Edit Format View Help

```
RMASS2 - Repeated Measures with Attrition: Sample Sizes for 2 Group Designs
-----------------------------------------------------------------------------


 Correlation Matrix of Y across time

                1      2
    1          1.000
    2          0.600  1.000

 Number of Timepoints           =      2
 Alpha level                    =  0.050 (2-sided)
 Power level                    =  0.800
 Grp1 to Grp2 Sample Size Ratio =  1.000

 Retention rate =   1.000   0.800
 Mean Diffs     =   0.500   0.500
 Stand. Devs.   =   1.000   1.000
 Effect Sizes   =   0.500   0.500
 Contrasts      =   0.707   0.707

 Composite Mean Difference                          =     0.707107
 Composite Variance (without attrition)             =     1.600000
 Composite Variance (adjusted for attrition)        =     1.795820

 Composite Effect size (without attrition)          =     0.559017
 N Subj for Grp1 at Time 1 (without attrition)       =    50.247707

 Composite Effect size (adjusted for attrition)      =     0.527659
 N Subj for Grp1 at Time 1 (adjusted for attrition) =    56.397411

 Sample Sizes by Group across Time - without Attrition

 Group 1      50.2      50.2
 Group 2      50.2      50.2

 Sample Sizes by Group across Time - with Attrition

 Group 1      56.4      45.1
 Group 2      56.4      45.1
 -----------------------------------------------------------------------------
```

II. Group difference across time

- mean difference $\mu_1 - \mu_2 = 0$ at $t1$ and .5 at $t2$
- time-related contrasts: $c_1 = -1$ and $c_2 = 1$

$$\Psi_c \;=\; -1(0) + 1(.5) \;=\; .5$$

$$\sigma^2_{rc} \;=\; (-1)^2(1)^2 + \frac{(1)^2(1)^2}{.8} + 2(-1)(1)(.6)/\sqrt{.8} \;=\; .91$$

contrast effect size $\delta = \Psi_c/\sigma_{rc} = .5/\sqrt{.91} = .525$

$$N_{11} \;=\; \frac{2(1.96 + .842)^2}{(.525)^2} \;=\; 57.1$$

Note:if $r_2 = 1$ then $N_{11} = 50$

Command Prompt - rmass2

```
Enter the output file name [rmass2.out]

Enter the number of time points [ 2]

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] 1

 Enter the attrition rate between timepoints  1 and  2 [0.050] .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 .5

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

 Enter correlation term number  1 [0.500] .6
 Enter the type of time-related contrast desired [0]
 0 = average across time; 1 = linear trend; 2 = user-defined 1

 Composite Effect size (without attrition)      =     0.559017
 N Subj for Grp1 Time 1 (without attrition)     =    50.247707

 Composite Effect size (adjusted for attrition) =     0.524616
 N Subj for Grp1 Time 1 (adjusted for attrition)=    57.053710


 Exit the program (1=Y OR 2=N) [2] ?
```

RMASS2.OUT - Notepad

File Edit Format View Help

```
RMASS2 - Repeated Measures with Attrition: Sample Sizes for 2 Group Designs
---------------------------------------------------------------------------


 Correlation Matrix of Y across time


                  1      2
    1           1.000
    2           0.600  1.000

 Number of Timepoints           =      2
 Alpha level                    =  0.050 (2-sided)
 Power level                    =  0.800
 Grp1 to Grp2 Sample Size Ratio =  1.000

 Retention rate =   1.000   0.800
 Mean Diffs     =   0.000   0.500
 Stand. Devs.   =   1.000   1.000
 Effect Sizes   =   0.000   0.500
 Contrasts      =  -0.707   0.707

 Composite Mean Difference                         =     0.353553
 Composite Variance (without attrition)            =     0.400000
 Composite Variance (adjusted for attrition)       =     0.454180

 Composite Effect size (without attrition)         =     0.559017
 N Subj for Grp1 at Time 1 (without attrition)      =    50.247707

 Composite Effect size (adjusted for attrition)     =     0.524616
 N Subj for Grp1 at Time 1 (adjusted for attrition) =    57.053710

 Sample Sizes by Group across Time - without Attrition

 Group 1     50.2     50.2
 Group 2     50.2     50.2

 Sample Sizes by Group across Time - with Attrition

 Group 1     57.1     45.6
 Group 2     57.1     45.6
 ---------------------------------------------------------------------------
```

Dichotomous outcomes

Comparison of two groups at a single timepoint

Number of subjects (N) in each of two groups (Fleiss, 1981):

$$N = \frac{\left[z_\alpha(2\bar{p}\bar{q})^{1/2} + z_\beta(p_1q_1 + p_2q_2)^{1/2}\right]^2}{(p_1 - p_2)^2}$$

- p_1 = response proportion in group 1 $\quad (q_1 = 1 - p_1)$
- p_2 = response proportion in group 2 $\quad (q_2 = 1 - p_2)$
- $\bar{p} = (p_1 + p_2)/2$
- $\bar{q} = 1 - \bar{p}$

Example

- $z_\alpha = 1.96$ 2-tailed .05 hypothesis test
- $z_\beta = .842$ power $= .8$
- $p_1 = .5$ and $p_2 = .7$

$$
\begin{aligned}
N &= \frac{\left[1.96(2 \times .6 \times .4)^{1/2} + .842(.5 \times .5 + .7 \times .3)^{1/2}\right]^2}{(.5 - .7)^2} \\
&= 93.03
\end{aligned}
$$

Dichotomous outcomes - longitudinal case

The number of subjects (N) in each of two groups for a consistent difference in proportions $p_1 - p_2$ between two groups across n timepoints (Diggle *et al.*, (2002):

$$N = \frac{\left[z_\alpha (2\bar{p}\bar{q})^{\frac{1}{2}} + z_\beta (p_1 q_1 + p_2 q_2)^{\frac{1}{2}}\right]^2 (1 + (n-1)\rho)}{n(p_1 - p_2)^2}$$

- p_1 = response proportion in group 1 $\quad (q_1 = 1 - p_1)$
- p_2 = response proportion in group 2 $\quad (q_2 = 1 - p_2)$
- $\bar{p} = (p_1 + p_2)/2$
- $\bar{q} = 1 - \bar{p}$
- ρ is the common correlation across the n observations

Example

- $z_\alpha = 1.96$ 2-tailed .05 hypothesis test
- $z_\beta = .842$ power $= .8$
- $n = 2$ timepoints
- correlation of repeated outcomes $= .6$
- $p_1 = .5$ and $p_2 = .7$

$$N = \frac{\left[1.96(2 \times .6 \times .4)^{\frac{1}{2}} + .842(.5 \times .5 + .7 \times .3)^{\frac{1}{2}}\right]^2 (1 + (2-1).6)}{2(.5 - .7)^2}$$

$$= 74.42$$

if $\rho = 0$ then $N = 46.51$ (cross-sectional)
if $\rho = 1$ then $N = 93.03$ (one-timepoint)

SAS code

```
* determines number per group;
* 5 timepoints (ICC=.4);
* difference in proportions of .5 and .67 (OR=2);
* power = .8 for a 2-tailed .05 test;

DATA one;
za = PROBIT(.975);
zb = PROBIT(.8);
n = 5;
p1 = .5; p2 = 2/3;
q1 = 1-p1; q2 = 1-p2;
pbar = (p1+p2)/2;
qbar = (q1+q2)/2;
rho = .4;
num = ((za*SQRT(2*pbar*qbar) + zb*SQRT(p1*q1 + p2*q2))**2)
      *(1 + (n-1)*rho);
den = n*((p1-p2)**2);
npergrp = num/den;
PROC PRINT;VAR npergrp;
RUN;
```

Ordinal outcomes

- methods not as developed
- use methods for continuous outcomes, but adjust the detectable effect sizes by an efficiency loss (*e.g.*, 80%)
  - Armstrong & Sloan (1989, Amer Jrn of Epid) report efficiency losses between 89% to 99% comparing an ordinal to continuous outcome, depending on the number of categories and distribution within the ordinal categories
  - Strömberg (1996, Amer Jrn of Epid) report efficiency losses between 87% to 97% comparing an ordinal outcome with 3 or 4 categories to one with 5 categories

Calculate Power via Simulation

- Randomly generate large number of datasets (NumDat) with assumed parameter values
  - NumDat = 5,000 and N per group = 63
  - Observations are normally distributed with $\mu_1 = 0$, $\mu_2 = .5$, $\sigma = 1$ (*e.g.*, effect size of .5)
- Analyze each dataset (5,000 t-tests) and count the number of times $H0 : \mu_1 = \mu_2$ is rejected (NumRej)
- Power = NumRej / NumDat

$\Rightarrow$ with above specifications, power = .8018 via simulation

Why simulate to get power?

- For simple situations where formulas exist, no real advantage to simulation approach
- However, for not-so-simple situations, simulation comes to the rescue
  - more complicated models (can easily include covariates and interactions)
  - different kinds of outcomes (binary, ordinal, counts)
  - can deal with longitudinal and/or clustered data

Comparison of Pre Post models

X_i = pre, Y_i = post, G_i = group (0=control, 1=test)

Post t-test

$$Y_i = \beta_0 + \beta_1 G_i + \epsilon_i$$

Change score t-test

$$(Y_i - X_i) = \beta_0 + \beta_1 G_i + \epsilon_i$$

ANCOVA

$$Y_i = \beta_0 + \beta_1 G_i + \beta_2 X_i + \epsilon_i$$

$H_0 : \beta_1 = 0$ is test of interest in all cases

Simulation results: tests of $H_0 : \beta_1 = 0$

- 10000 datasets with 100 subjects in each of 2 groups
- mean difference of 0 at pre, .4 at post
- variance = 1 at both timepoints for both groups
- correlation = .4, .45, .5, .55, .6 between pre and post measurements

correlation	model	rejection rate
0.400	ttest	0.81
0.400	change	0.73
0.400	ancova	0.87
0.450	ttest	0.81
0.450	change	0.77
0.450	ancova	0.89
0.500	ttest	0.81
0.500	change	0.81
0.500	ancova	0.91
0.550	ttest	0.81
0.550	change	0.85
0.550	ancova	0.92
0.600	ttest	0.81
0.600	change	0.88
0.600	ancova	0.94

SAS simulation program: 2 sample t-test

```
* t-test power for .5 effect size and 63 per group;
OPTIONS NOCENTER NOSOURCE NONOTES NOSPOOL NODATE NONUMBER;
* writing log and output info to external files;
PROC PRINTTO LOG='c:\POWER\Ttest_power_N=63.log';
PROC PRINTTO PRINT='c:\POWER\Ttest_power_N=63.lst';
RUN;

* Specify the parameters;
DATA parms;
* number of datasets;
ndats = 5000;
* number of subjects per group;
npergrp = 63;
* total number of subjects;
nsubj = 2*npergrp;
* effect size for the group difference;
effsize = .5;
* seed for random number generation;
SEED = 974657747;
```

```
* Generate the data;
DATA simdat; SET parms;
ndat=1;
DO WHILE (ndat LE ndats);
   n=1;
   DO WHILE (n LE nsubj);
      grp = 0; IF n > npergrp THEN grp = 1;
      err = RANNOR(SEED);
      y = effsize*grp + err;
      OUTPUT simdat;
   n+1; END;
ndat+1; END;

* t-test via regression model;
PROC REG;
MODEL y = grp;
BY ndat;
ODS OUTPUT 'Parameter Estimates'=parmest;
RUN;
```

```
* select only the grp estimate;
DATA grpest; SET parmest; BY ndat;
IF VARIABLE = 'grp';

* summarize the results of the regression model;
DATA sumres;
IF _N_ = 1 THEN SET parms; SET grpest;

reject = 0;
IF (ABS(ESTIMATE/STDERR) > 1.95996) THEN reject = 1;

PROC MEANS NOPRINT;
VAR ndats npergrp estimate effsize reject;
OUTPUT OUT = outres
MEAN(ndats npergrp estimate effsize reject) =
     ndatasets numpergrp estm truev rejrate;
RUN;
```

```
* write out the results;
DATA outfile; SET outres
(KEEP = ndatasets numpergrp truev estm rejrate);
FILE 'c:\POWER\Ttest_power_N=63.txt';
PUT (ndatasets numpergrp) (6.0) (truev estm rejrate) (12.6);
RUN;

PROC EXPORT DATA=outfile
OUTFILE= 'c:\POWER\Ttest_power_N=63.xls'
DBMS=EXCEL2000 REPLACE;
RUN;
```

SAS simulation program: Logistic Regression with 2 groups

```
* logistic regression power for OR=2 and 137 per group;
OPTIONS NOCENTER NOSOURCE NONOTES NOSPOOL NODATE NONUMBER;
* writing log and output info to external files;
PROC PRINTTO LOG='c:\POWER\LReg_power_N=137.log';
PROC PRINTTO PRINT='c:\POWER\LReg_power_N=137.lst';
RUN;

* Specify the parameters;
DATA parms;
* number of datasets;
ndats = 5000;
* number of subjects per group;
npergrp = 137;
* total number of subjects;
nsubj = 2*npergrp;
* odds ratio for the group difference;
oddsratio = 2;
* seed for random number generation;
SEED = 974657747;

* Generate the data;
DATA simdat; SET parms;
ndat=1;
DO WHILE (ndat LE ndats);
    n=1;
    DO WHILE (n LE nsubj);
          grp = 0; IF n > npergrp THEN grp = 1;
          exp1 = RANEXP(SEED); exp2 = RANEXP(SEED);
          err = LOG(exp1/exp2);
          ystar = LOG(oddsratio)*grp + err;
          y = 0;if ystar > 0 then y = 1;
          OUTPUT simdat;
    n+1; END;
ndat+1; END;
```

```
* logistic regression model;
PROC LOGISTIC DESCENDING;
MODEL y = grp;
BY ndat;
ODS OUTPUT 'Parameter Estimates'=parmest; RUN;

* select only the grp estimate;
DATA grpest; SET parmest; BY ndat;
IF VARIABLE = 'grp';
EstOddsRatio = EXP(ESTIMATE);

* summarize the results of the logistic regression model;
DATA sumres;
IF _N_ = 1 THEN SET parms; SET grpest;

reject = 0;
IF (ABS(ESTIMATE/STDERR) > 1.95996) THEN reject = 1;

PROC MEANS NOPRINT;
VAR ndats npergrp EstOddsRatio oddsratio reject;
OUTPUT OUT = outres
MEAN(ndats npergrp EstOddsRatio oddsratio reject) =
        ndatasets numpergrp estm truev rejrate;
RUN;

* write out the results;
DATA outfile; SET outres
(KEEP = ndatasets numpergrp truev estm rejrate);
FILE 'c:\POWER\LReg_power_N=137.txt';
PUT (ndatasets numpergrp) (6.0) (truev estm rejrate) (12.6);
RUN;

PROC EXPORT DATA=outfile
OUTFILE= 'c:\POWER\LReg_power_N=137.xls'
DBMS=EXCEL2000 REPLACE;
RUN;
```

SAS simulation program: Random intercept model

```
* Random Intercept model power for .5 effect size and 33 per group;
OPTIONS NOCENTER NOSOURCE NONOTES NOSPOOL NODATE NONUMBER;
* writing log and output info to external files;
PROC PRINTTO LOG='c:\POWER\RandInt_power_N=33.log';
PROC PRINTTO PRINT='c:\POWER\RandInt_power_N=33.lst';
RUN;

* Specify the parameters;
DATA parms;
ndats = 5000;       * number of datasets;
npergrp = 33;       * number of subjects per group;
nsubj = 2*npergrp;  * total number of subjects;
ntime = 5;          * number of timepoints;
effsize = .5;       * effect size for the group difference;
icc = .4;           * ICC for repeated outcomes;
SEED = 974657747;   * seed for random number generation;

* Generate the data;
* calculate the subject & error variances;
* assuming the subject plus error variance equals 1;
varsub = icc; varerr = (1-icc);
DATA simdat; SET parms;
ndat=1;
DO WHILE (ndat LE ndats);
    n=1;
    DO WHILE (n LE nsubj);
        grp = 0; IF n > npergrp THEN grp = 1;
        subj = SQRT(varsub)*RANNOR(SEED);
        j=1;
        DO WHILE (j LE ntime);
            err = SQRT(varerr)*RANNOR(SEED);
            y = effsize*grp + subj + err;
            OUTPUT simdat;
        j+1; END;
    n+1; END;
ndat+1; END;
```

```
* random intercept model;
PROC MIXED NOCLPRINT;
CLASS n;
MODEL y = grp /SOLUTION;
RANDOM INTERCEPT/SUBJECT=n;
BY ndat;
ODS OUTPUT SOLUTIONF=parmest;
RUN;

* select only the grp estimate;
DATA grpest; SET parmest; BY ndat;
IF EFFECT = 'grp';

* summarize the results of the random intercept model;
DATA sumres;
IF _N_ = 1 THEN SET parms; SET grpest;
reject = 0;
IF (ABS(ESTIMATE/STDERR) > 1.95996) THEN reject = 1;

PROC MEANS NOPRINT;
VAR ndats npergrp ntime icc estimate effsize reject;
OUTPUT OUT = outres
MEAN(ndats npergrp ntime icc estimate effsize reject) =
        ndatasets numpergrp numtime intclass estm truev rejrate;
RUN;

* write out the results;
DATA outfile; SET outres
(KEEP = ndatasets numpergrp numtime intclass truev estm rejrate);
FILE 'c:\POWER\RandInt_power_N=33.txt';
PUT (ndatasets numpergrp numtime) (6.0) (intclass) (8.4) (truev estm rejrate) (12.6);
RUN;

PROC EXPORT DATA=outfile
OUTFILE= 'c:\POWER\RandInt_power_N=33.xls'
DBMS=EXCEL2000 REPLACE; RUN;
```

SAS simulation program: Random intercept Logistic model

```
* Random Int Logistic power for marginal OR=2 and 70 per group;
OPTIONS NOCENTER NOSOURCE NONOTES NOSPOOL NODATE NONUMBER;
* writing log and output info to external files;
PROC PRINTTO LOG='c:\POWER\LR_RandInt_power_N=70.log';
PROC PRINTTO PRINT='c:\POWER\LR_RandInt_power_N=70.lst';
RUN;

* Specify the parameters;
DATA parms;
ndats = 5000;        * number of datasets;
npergrp = 70;        * number of subjects per group;
nsubj = 2*npergrp;   * total number of subjects;
ntime = 5;           * number of timepoints;
mod = 2;             * marginal odds ratio for the group difference;
icc = .4;            * ICC for repeated outcomes;
SEED = 974657747;    * seed for random number generation;

* Generate the data;
* calculate the error & subject variances, and the subject-specific beta;
varerr = ((ATAN(1)*4)**2)/3;
varsub = varerr*(icc/(1-icc));
ssbeta = LOG(mod)*SQRT((varsub + varerr)/varerr);
DATA simdat; SET parms;
ndat=1; DO WHILE (ndat LE ndats);
    n=1; DO WHILE (n LE nsubj);
         grp = 0; IF n > npergrp THEN grp = 1;
         subj = SQRT(varsub)*RANNOR(SEED);
         j=1; DO WHILE (j LE ntime);
              exp1 = RANEXP(SEED); exp2 = RANEXP(SEED);
              err = LOG(exp1/exp2);
              ystar = ssbeta*grp + subj + err;
              y = 0;IF ystar > 0 THEN y = 1;
              OUTPUT simdat;
         j+1; END;
    n+1; END;
ndat+1; END;
```

```
* random intercept logistic model;
PROC GLIMMIX NOCLPRINT METHOD=QUAD(QPOINTS=5);
CLASS n;
MODEL y (DESCENDING) = grp /DIST=BINARY SOLUTION;
RANDOM INTERCEPT/SUBJECT=n;
BY ndat;
ODS OUTPUT PARAMETERESTIMATES=parmest;
RUN;

* select only the grp estimate;
DATA grpest; SET parmest; BY ndat;
IF EFFECT = 'grp';

* summarize the results of the random intercept logistic model;
DATA sumres;
IF _N_ = 1 THEN SET parms; SET grpest;
reject = 0;
IF (ABS(ESTIMATE/STDERR) > 1.95996) THEN reject = 1;

PROC MEANS NOPRINT;
VAR ndats npergrp ntime icc estimate ssbeta reject;
OUTPUT OUT = outres
MEAN(ndats npergrp ntime icc estimate ssbeta reject) =
        ndatasets numpergrp numtime intclass estm truev rejrate;
RUN;

* write out the results;
DATA outfile; SET outres
(KEEP = ndatasets numpergrp numtime intclass truev estm rejrate);
FILE 'c:\POWER\LR_RandInt_power_N=70.txt';
PUT (ndatasets numpergrp numtime) (6.0) (intclass) (8.4) (truev estm rejrate) (12.6);
RUN;

PROC EXPORT DATA=outfile
OUTFILE= 'c:\POWER\LR_RandInt_power_N=70.xls'
DBMS=EXCEL2000 REPLACE; RUN;
```

Additional papers, some with programs, on power for longitudinal studies

- Basagaña X. & Spiegelman D. (2010). Power and sample size calculations for longitudinal studies comparing rates of change with a time-varying exposure. *Statistics in Medicine*, 29(2):181-92.
- Comulada W.S. & Weiss R.E. (2010). Sample size and power calculations for correlations between bivariate longitudinal data. *Statistics in Medicine*, 29(27):2811-24.
- Dang, Q., Mazumdar, S. & Houck, P.R. (2008). Sample size and power calculations based on generalized linear mixed models with correlated binary outcomes. *Computer Methods and Programs in Biomedicine*, 91(2), 122-127.
- Donohue, M.C., Gamst, A.C., & Edland, S.D. (2010). longpower: Sample size calculations for longitudinal data. http://cran.r-project.org/web/packages/longpower/index.html
- Jung, S.H. & Ahn, C. (2003). Sample size estimation for GEE method for comparing slopes in repeated measurements data. *Statistics in Medicine*, 22(8):130515.
- Raudenbush, S.W., Xiao-Feng L. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*, 6(4):387401.
- Rochon, J. (1998). Application of GEE procedures for sample size calculations in repeated measures experiments. *Statistics in Medicine*, 17(14):1643-1658.
- Roy, A., Bhaumik, D.K., Subhash, S. & Gibbons R.D. (2007). Sample size determination for hierarchical longitudinal designs with differential attrition rates. *Biometrics*, 63(3):699-707. http://healthstats.org/rmass/
- Tu, X.M., Kowalski, J., Zhang, J., Lynch, K.G., & Crits-Christoph P. (2004). Power analyses for longitudinal trials and other clustered designs. *Statistics in Medicine*, 23(18):2799-815.
- Tu, X.M., Zhang, J., Kowalski, J., Shults, J., Feng, C., Sun, W., & Tang, W. (2007). Power analyses for longitudinal study designs with missing data. *Statistics in Medicine*, 26(15):2958-81.
- Zhang, Z., & Wang, L. (2009). Power analysis for growth curve models using SAS. *Behavior Research Methods*, 41(4), 1083-1094. http://www.psychstat.org/us