*Simulations to Assess Power for Treatment Effect on a Vector of Binary Events

MULTBINPOW 1.0 software assists researches in choosing the most powerful statistical test in the design of randomized and non-randomized studies in which the outcome is a composite binary-event endpoint MULTBINPOW calculates empirical power via simulations for 8 distinct statistical tests, each comparing 2 vectors of binary events (treatment versus control) under user-specified ranges of within-subject correlations, covariance structures, sample sizes and incidences. Details of the statistical methods are given in Mascha and Imrey (2010). Simulations are based on the method of Oman (2009).

Power is computed for the following statistical tests:

1. Collapsed composite of any-versus-none of the events
2. Count of events within individual (Mann-Whitney and T-test)
3. Minimum P-value test
4. GEE common effect test (also called "global odds ratio" test).
5. GEE average relative effect (distinct effects test, removes influence of high incident components)
6. GEE K-df distinct effects test (analogous to Hotelling's T2, not sensitive to direction)
7. GEE covariance-weighted distinct effects test (usually very similar power to GEE common effect)
8. GEE treatment-component interaction distinct effects test (whether treatment effect varies across vector)

The user specifies a range of underlying within-subject correlations, covariance matrix, sample sizes, and response vectors for treatment and control groups;

*All parameter option for MULTBINPOW are described in detail below;

* AUTHOR: Edward J. Mascha, PhD, Cleveland Clinic. Address questions/comments to maschae@ccf.org;
   *Citing this program if using it to design a study is much appreciated;
   *Program is not guaranteed, although has been carefully tested;
   *Updates will be made periodically to enhance available options and outputs;

*A SAMPLE design and CALL are as follows:

```
     Design: Calculate power (at alpha=0.05) to compare 2 vectors of 3 proportions each,
       Treatment: (pr_t=.18 .16 .09), Control: (pr_c= .20  .20  .15), Weight outcomes 1-3 as .2 .4 .4, respectively
       Assess power at within-subject correlations of .10, .30, .50 and N/group of 200, 400, 600.
       Number of simulations: B=1000 for each combination of within-subject correlation and N/group.
       Create simulations  using exchangeable correlation (rmat=2) and use unstructured (use_rmat=3) to analyze data;

*    %LET myPATH=your directory to store results;
*    %inc "/YOUR DIRECTORY/multbinpow.sas";
*    %multbinpow(path=&mypath, reset=0, sims=1, summary=1, makesum=1, outdata=results,imlwt=0,
       startsim=1, numsims=1000, pr_t= .18 .16 .09, pr_c= .20  .20  .15, rmin=10, rmax=50, rby=20, nmin=200,
       nmax=600,nby=200, rmat=2, use_rmat=3, wt=.2 .4 .4, spar_ck=0, where=,SIMSEED=1234, alpha=.05, listres=0);
```

****

*PARAMETERS   DEFINITIONS/OPTIONS

path=	User-defined path to which results will be sent
outdata=multpower	User-specified name of output SAS dataset containing power results
sims=1	1= create simulations (RUNs loops.sas which calls simulation macro (must=1 to create simulations)), 0= No
runanal=1	1= analyze simulated data (usually=1 when SIMS=1. Set to 0 with SIMS=1 to create datasets w/o analysis)
summary=1	Print results (assumes makesum=1 was run at least once after simulations finished)
makesum=1	Create summary SAS dataset &libr..&outdata._sum based on specified simulations (requires summary=1)
startsim=1	Index number to begin current simulation runs (usually 1, but can be any number > 0)
numsims=100	Number of simulated samples to create (eg, if startsim=1 and numsims=100, 100 runs will be made. To add 400 to initial 100 simulations, for example, use startsim=101 and numsims=500)
pr_t=	Vector of outcome proportions for components 1 - K in Treatment group (e.g., pr_t = .10 .15 .20)
pr_c=	Vector of outcome proportions for components 1 - K in Control group (e.g., pr_c = .15 .18 .25)

*Below: set parameters for within-subject correlation and sample size combinations. Total simulation strata are number of distinct correlations times the number of distinct sample sizes.

The specific R values below (from Rmin to Rmax by Rby) indicates the starting correlation parameter within a particular set of simulations. For example, if Rmin=0.10, Rmax = 0.30 and Rby = 0.20, separate simulations will be run using an R value of 0.10 and 0.30. If compound symmetry correlation structure is specified by RMAT=2, then the first set of simulations will use R of 0.10 for the correlation between all pairs of the components of the composite outcome. Another set of simulations will use R = 0.30 for each pair of components.

If AR(1) is specified by RMAT=1, then Rmin = .10 will use correlation of 0.10 for the first pair of components in the use list in PR_T and PR_C, and then smaller correlations for components further apart in the input listing. The ar(1) correlation will be estimated as rho to the power | j – k |, where "j – k" is the distance between the proportions, which would be 1 between components 1 and 2 and 2 between components 1 and 3, and so forth. The ar(1) is merely used as a tool to create binary outcome data that are not all equally correlated. If a vector of 3 outcome component proportions are specified (3 treatment, 3 control), then the 3 x 3 correlation matrix with rho = 0.10 would give a correlation of approximately 0.10 for the first and second components listed, and 0.1 to the power 2, or 0.01, for component 1 versus 3. Then all would be rerun for rho = .30 (since Rmax = 0.30), with correlation .04 for components 1 and 3.

Rmin=10	Starting (i.e., lowest) value of within-subject correlation (x 100) to use in simulations. The specific R value indicates the correlation parameter used within a particular set of simulations.
Rmax=40	Maximum (i.e., highest) value of within-subject correlation (x 100) to use in a simulations. The specific R value indicates the correlation parameter used within a particular set of simulations.
Rby=10	By value for within-subject correlations (x 100). Example: rmin=10, rmax=40, rby=10 creates 4 sets of simulations, corresponding to within-subject correlations equal to 0.1, 0.2, 0.3 and 0.4

```
Nmin=100            Lowest PER GROUP sample size to be used in simulations
Nmax=500            Highest PER GROUP sample size to be used in simulations
Nby=100             By value for PER GROUP sample sizes. Example, nmin=50, nmax=150, nby=50 creates simulations for
                     PER GROUP N=50, 100 and 150 and thus total N of 100, 200 and 300

** additional parameters

listres=0           0= Do not show every SAS proc result for each simulation, 1= yes
rmat=1              Simulation working corr  1= AR(1) first order autoregressive, 2= CS (compound symmetry)
                       Example: AR(1) useful to specify high correlation for some components, lower for others
use_rmat= rmat      Analysis working correlation.  Default=rmat. Otherwise, 1= AR(1), 2= CS (compound symmetry),
                       3=UN (unspecified)
wt=1                Input either single weight of 1 for all components, or vector (eg, wt= .1 .2 .4 .3)summing to 1

reset=0             0=do not delete data in (outdata=), 1= delete bootstrap results for given dataset (outdata=)

spar_ck=1           Check sparcity of simulated data in treatment-component crossings
imlwt=0             RUN PROC IML to obtain test for covariance-matrix weighted average of estimated treatment effects
                    -- if this is chosen, must have PROC IML module live.  otherwise, PROC IML is not needed

SIMSEED=1234        Simulation seed. Actual seed used for sims is function of bseed, ith boot, nper and rho
alpha=.05           Alpha at which power calculated

figpath=&path       Physical path where figure should be put
figname=powerfig    Filename for JPEG figure
figdata=figdata     SAS filename for power results as presented in figure
figtype=jpg         Valid SAS figure types. Also used as extension to figure.
FigHeight =800px    Figure height
FigWidth = 600px    Figure width
figDpi=150          Figure DPI

where=              Specify subset of data to be used in proc tabulate summary of results (eg, where= where rho= .20)
progname=mySAS      Insert the name of your SAS program to be used as a label in output title
clearlog=1          Clear log after each simulation run (0= do not clear logs)

*************************************************************************************************

Output dataset (&outdata._sum.sas7bdat) key variables and labels:

      nper="N/group"
      r=”Rho”
      wcorr2="GEE working corr(1,2)"
      k="Number of components"
      b_coll=   "Beta collapsed composite"
      se_b_coll="SE(B) collapsed composite"
      b_gc=     "Beta common effect"
      se_b_gc=  "SE(B) common effect"
```

```
      b_avg_wald=    "Beta average relative effect"
      se_avg_wald=   "SE(B) avg rel effect"

                *Pow= Power;
      rp_b_coll="Pow collapsed comp (Wald)"
      rpz_b_gc="Pow common effect (Wald)"
      rp_avg_wald="Pow avg rel effect (Wald)"
      rp_distinct_kdf="Pow K-df distinct(Wald)"
      rp_avg_sc="Pow avg rel effect(Score)"
      rp_chi_INTER="Pow TX-component interaction (Score)"
      rp_w_count="Pow Count (MW)"
      rp_diff_ct="Pow Count (pooled t)"
      rmin_pval="Pow minimum P-value ???"
      rmin_bootp="Pow min bootstrap P-val"
      rpwald_wtavg="Pow var-cov wt avg"

Figure dataset (&figdata..sas7bdat) key variables and labels:
       nper="N per group"
       meanvar="Rejections" (proportion of null hypotheses rejected = power)
       r="Rho"
       test="Test"
```