Online Supplementary Appendix for

Dziak, Nahum-Shani and Collins (2012),

"Multilevel Factorial Experiments for Developing Behavioral Interventions: Power, Sample Size, and Resource Considerations"

Appendix D

Resources for Implementing Power Calculations

Within-Clusters Macro

The following is a copy of a SAS macro which computes power for between-clusters factorial experiments. The current version of this macro is also found at http://methodology.psu.edu/multilevelfactorial/ . The arguments "PretestICC," "ChangeICC," "PretestPosttestCorr," "NumFactors," "NumClusters," "MeanNumMembers," "StandardizedEffectSize," "AlphaLevel" and "ModelOrder" provide the input and settings for how the macro will work.

```
 %MACRO WithinClustersPower(PretestICC=,
                           ChangeICC=,
                           PretestPosttestCorr=,
                           NumFactors=,
                           NumClusters=,
                           MeanNumMembers=,
                           StandardizedEffectSize=,
                           AlphaLevel=.05,
                           ModelOrder=2);
/*
 By John Dziak, The Methodology Center at Penn State.
 Copyright 2012 The Pennsylvania State University.
 This program is free software: you can redistribute
 it and/or modify it under the terms of the GNU General
 Public License as published by the Free Software Foundation,
 either version 3 of the License, or (at your option) any
 later version. This program is distributed in the hope that
 it will be useful, but WITHOUT ANY WARRANTY; without even
 the implied warranty of MERCHANTABILITY or FITNESS FOR A
 PARTICULAR PURPOSE. See the GNU General Public License
 for more details (http://www.gnu.org/licenses/).
```

```
  First of two SAS Macros to accompany Dziak, Nahum-Shani and
  Collins, "Multilevel Factorial Experiments for Developing
  Behavioral Interventions: Power, Sample Size and Resource
  Considerations". Finds power per component given effect
  size for multisite individually-randomized factorial designs.
*/
PROC IML;
  /* Initial Calculations of Design Options and Parameters; */
  PretestICC = &PretestICC;
  ChangeIcc = &ChangeICC;
  PretestPosttestCorr = &PretestPosttestCorr;
  NumFactors = &NumFactors;
  NumClusters = &NumClusters;
  MeanNumMembers = &MeanNumMembers;
  PretestICC = &PretestICC;
  ChangeICC = &ChangeICC;
  PretestPosttestCorr = &PretestPosttestCorr;
  AlphaLevel = &AlphaLevel;
  StandardizedEffectSize = &StandardizedEffectSize;
  N = NumClusters*MeanNumMembers;
  /* Here the variance components are calculated using the results
  in Appendix B, assuming sigma2 total=1.  This assumption is
  without loss of generality because the input effect size is
  assumed to be relative to sigma total. The effect-coded
  regression coefficient is half the standardized effect
  size (see Appendix A). */
  SigmaSqd = (1-PretestPosttestCorr)*(1-PretestICC);
  TauSqdPi0 = PretestPosttestCorr*(1-PretestICC);
  TauSqdBeta1 = 2*SigmaSqd*(ChangeICC)/(1-ChangeICC);
  TauSqdBeta0 = 1 - .25*TauSqdBeta1 - TauSqdPi0 - SigmaSqd;
  Gamma = StandardizedEffectSize/2;
  N = MeanNumMembers*NumClusters;
  /* The following input checking code is for detecting errors
```

```
and input which the macro would not be able to process.
If the input checking code determines that the power
calculations would not be meaningful or trustworthy with this
input, the variable "try" is set to zero, in which case the
power calculations will not be done. */
try = 1;
Message1 = "";
IF ((&PretestICC < 0) | (&PretestICC >= 1)) THEN DO;
  Message1 = "The posttest ICC estimate must be greater than
              or equal to zero and less than one.";
  try = 0;
END;
IF ((&ChangeIcc < 0) | (&ChangeIcc >= 1)) THEN DO;
  Message1 = "The change-score ICC estimate must be greater than
              or equal to zero and less than one.";
  try = 0;
END;
IF ((&PretestPosttestCorr < 0) | (&PretestPosttestCorr >= 1)) THEN DO;
  Message1 = "The pretest-posttest correlation must be greater than
              or equal to zero and less than one.";
  try = 0;
END;
IF ((&NumFactors < 1) | (&NumFactors > 10)) THEN DO;
  Message1 = "The number of factors should be between 1 and 10.";
  try = 0;
END;
IF ((&MeanNumMembers < 1) | (&MeanNumMembers > 10000)) THEN DO;
  Message1 = "The mean cluster size should be between 1 and 10000.";
  try = 0;
END;
IF ((&StandardizedEffectSize < 0) | (&StandardizedEffectSize > 2)) THEN DO;
  Message1 = "The standardized effect size should be between 0 and 2.";
  try = 0;
```

```
END;
IF ((&alphaLevel <= 0) | (&alphaLevel > .25)) THEN DO;
  Message1= "The alpha level should be above 0 and below .25";
  try = 0;
END;
/* Check model order (1=fit main effects only, */
/* 2=fit main effects and two-way interactions)*/
ModelOrder = %EVAL(&ModelOrder)+0;
IF ((ModelOrder=1) | (ModelOrder=2)) THEN; ELSE DO;
Message1 = "The model order should be 1 or 2.";
try = 0;
END;
/* Determine the error df.
The number of parameters in the model are determined here.  It
is assumed that the factors are all dichotomous and that either
a first order (main effects only) or second order model will be fit.
*/
IF ModelOrder=1 THEN DO;
    NumParams = 1 + INT(&NumFactors);
    ErrorDF = N - NumParams;
    END;
IF ModelOrder=2 THEN DO;
    NumParams = 1 + INT(&NumFactors) +
                  INT(&NumFactors)*(INT(&NumFactors) - 1)/2;
    ErrorDF = N - NumParams;
END;
/* Calculate the power requested.  The power for detecting a main
effect of size d is computed here under the two-level
covariate-adjusted model.  The work is done by the SAS functions
PROBF and FINV.  */
IF Try = 1 THEN DO;
Noncentrality = N * Gamma * Gamma / (2*SigmaSqd);
CritVal = FINV(1 - AlphaLevel,1,ErrorDF);
```

```
PowerForMain = 1 - PROBF(CritVal, 1, ErrorDF, Noncentrality);
/* The corresponding calculation for an interaction of size d
is performed, if the model is second order. */
PowerForInteraction = 1 - PROBF(CritVal,
     1,
     ErrorDF,
     (Noncentrality)/4);
IF ModelOrder = 1 THEN PowerForInteraction = 0;
/*Here the results of the computation are output to the screen.*/
DesignParameters = PretestICC //
      ChangeIcc //
      PretestPosttestCorr //
      StandardizedEffectSize //
      NumFactors //
      NumParams //
      MeanNumMembers //
      NumClusters;
r = "Posttest ICC" //
     "Change ICC" //
     "Pretest-Posttest Corr." //
     "Effect Size" //
     "# Factors in Experiment" //
     "# Parameters in Regression Model" //
     "Mean # Members / Cluster" //
     "# Clusters" ;
c = "";
Message2="";
IF (ModelOrder=1) THEN
   Message2="Assuming a model without interactions," //
       "and a two-sided test for each main effect.";
IF (ModelOrder=2) THEN
   Message2="Assuming a model with 2-way interactions," //
      "and a two-sided test for each main effect and interaction.";
```

```
  PRINT DesignParameters[rowname=r colname=c format=12.4
        label="Design Parameters specified by user:"], Message2[label=""];
  VarianceParameters = TauSqdPi0 //
                       TauSqdBeta0 //
                       TauSqdBeta1 //
                       SigmaSqd;
  r = "Tau^2[pi,0] (Individual)" //  "Tau^2[beta,0] (Cluster)" //
    "Tau^2[beta,1] (Cluster * Time)" // "Sigma^2 (Error)" ;
  PRINT VarianceParameters[rowname=r colname=c format=12.4
   label="Variance Components:"] " ";
  PRINT("Variance components above have been scaled " //
        "so that total posttest variance is " //
        "Tau^2[pi,0]+Tau^2[beta,0]+Sigma^2+.25*Tau^2[beta,1]=1") ;
  PRINT("Estimated Power for a Main Effect of" //
     "Standardized Size &StandardizedEffectSize at alpha=&AlphaLevel: ");
  PRINT PowerForMain[label="" format=6.4];
  PRINT("Estimated Power for an Interaction of" //
      "Standardized Size &StandardizedEffectSize at alpha=&AlphaLevel:");
  PRINT PowerForInteraction[label="" format=6.4 ];
/* Here a power table is displayed, showing the power that would be obtained
      with different numbers of clusters of the same size. */
  PRINT "Please see next page for a power table ..." / ;
  NumClustersForTable = ( 4+2*(0:11)  );
  NoncentralityForTable = MeanNumMembers * NumClustersForTable *
                               Gamma * Gamma / (2*SigmaSqd);
  ErrorDFForTable = NumClustersForTable*MeanNumMembers - NumParams;
  PowerForMainForTable = J(12,1,0);
  PowerForInteractionForTable = J(12,1,0);
  DO i = 1 TO 12;
    /* This is basically the same power calculation as before, but
         repeated for several possible values for the sample size
          in terms of number of clusters. */
    PowerForMainForTable[i] = 1 - PROBF(
```

```
            FINV(1 - AlphaLevel,1,ErrorDFForTable[i]),
            1,
            ErrorDFForTable[i],
            NoncentralityForTable[i]);
  PowerForInteractionForTable[i] = 1 - PROBF(
       FINV(1-AlphaLevel,1,ErrorDFForTable[i]),
            1,
            ErrorDFForTable[i],
            NoncentralityForTable[i]/4);
 END;
 r = CHAR(NumClustersForTable);
 PRINT "Power for Main Effects by # Clusters:"
       PowerForMainForTable[r=r c="Power" label=" " format=10.4];
 IF ModelOrder=2 THEN DO;
   PRINT "Power for Interactions by # Clusters:"
         PowerForInteractionForTable[r=r c="Power" label=" " format=10.4];
 END;
/* Here the results are written to a SAS dataset.*/
 CREATE WithinClustersPowerFormulaOut
       VAR {     PretestICC
                 ChangeIcc
                 PretestPosttestCorr
                 NumFactors
                 MeanNumMembers
                 NumClusters
                 AlphaLevel
                 StandardizedEffectSize
                 ModelOrder
                 NumParams
                 SigmaSqd
                 TauSqdPi0
                 TauSqdBeta1
                 TauSqdBeta0
```

```
                Noncentrality
                ErrorDF
                PowerForMain
                PowerForInteraction
                CritVal
                Gamma };
  APPEND;
  CLOSE WithinClustersPowerFormulaOut;
     END;
  ELSE DO;
  PRINT "The macro encountered a problem:";
  PRINT Message1;
  END;
QUIT;
%MEND;
```

The syntax and information required by the macro is as follows.

- PretestICC is the researcher's guess at the value of the intraclass correlation (ICC) of the pretest response (Y_{pre}) within clusters. It is assumed to be the same as the ICC of the posttest Y_{post} within clusters, holding treatment condition constant, i.e., the same as $ICC(Y_{\text{post}}|\mathbf{X})$. The values used in the simulations in the article for low, medium and high are .05, .15, and .30.

- ChangeICC is the researcher's guess, based on the literature, at the ICC of the change scores defined by $Y_{\text{post}} - Y_{\text{pre}}$. In the simulations in the article it was set to be .50 times PretestICC, i.e., .025, .075, or .15.

- PretestPosttestCorr is the researcher's guess at the correlation between Y_{pre} and Y_{post} after adjusting for cluster effects and treatment effects. This is something like pretest-posttest reliability and could also be thought of as a kind of within-subject ICC. For the simulations in the article it was set to be .65.

- NumFactors is the number (K) of factors or components included in the study. It is assumed that all factors are dichotomous. Thus, the complete factorial has 2^K cells, and a $1/2^f$ factorial has 2^{K-f+1} nonempty cells. We assume that it does not matter for power whether a complete or fractional factorial is being used (i.e., we ignore aliasing in this macro).

- NumClusters is the number J of clusters (e.g., clinics or schools) assumed to be available in the sample.

- MeanNumMembers is the mean number $\bar{n}$ of members assumed to be available within any given cluster (e.g., number of patients per clinic or students per school).

- StandardizedEffectSize is the standardized effect size for one factor, expressed as a scaled difference d relative to the overall error standard deviation σ_{tot}. This should be the minimum effect size the researcher hopes to be able to detect with adequate testwise (per-effect) power for any given factor. σ_{tot} is the standard deviation of Y_{post} in a naïve model adjusting for treatment effects but not adjusting for clustering or for posttest, i.e., $\sqrt{\text{Var}\left(Y_{\text{post}}|\mathbf{X}\right)}$. Also, σ_{tot} is equal to $\sqrt{\text{Var}\left(Y_{\text{pre}}\right)}$ if time is centered in Model (8) or (9). For the simulations in the article the standardized effect size was approximately .20. Estimated power will be calculated both for a main effect of this size and for an interaction of this size.

- AlphaLevel is the Type I error rate of the test, set to .05 by default.

- ModelOrder is 1 if only main effects are to be estimated in the model, or 2 if main effects and two-way interactions are to be estimated. In the article we essentially used option 2.

Here is an example of calling the within-clusters power formula macro. Suppose a factorial experiment is planned, which will have 5 factors, 5 clusters, 50 members per cluster, a pretest ICC within clusters of .05, a change-score ICC within clusters of .025, a pretest-posttest correlation (within-subject ICC) of .65, and a standardized effect size of .2306 relative to σ_{tot}. (This scenario corresponds to the condition in the first row of Table 5 in the paper). Suppose the macro file is saved in a folder named `c:\Documents`. Then use the code

```
%INCLUDE "C:\Documents\WithinClustersPower.sas";
%WithinClustersPower( PretestICC=.05,
                     ChangeICC=.025,
                     PretestPosttestCorr=.65,
                     NumFactors=5,
                     NumClusters=5,
                     MeanNumMembers=50,
                     StandardizedEffectSize=.2306,
                     AlphaLevel=.05,
                     ModelOrder=2);
```

The following results are obtained:

```
        Posttest ICC                              0.0500
        Change ICC                                0.0250
        Pretest-Posttest Corr.                    0.6500
        Effect Size                               0.2306
        # Factors in Experiment                   5.0000
        # Parameters in Regression Model         16.0000
        Mean # Members / Cluster                 50.0000
        # Clusters                                5.0000

  Assuming a model with 2-way interactions,
  and a two-sided test for each main effect and interaction.

        Variance Components:

        Tau^2[pi,0] (Individual)                0.6175
        Tau^2[beta,0] (Cluster)                 0.0457
        Tau^2[beta,1] (Cluster * Time)          0.0171
```

Sigma^2 (Error) 0.3325

Variance components above have been scaled
so that total posttest variance is
Tau^2[pi,0]+Tau^2[beta,0]+Sigma^2+.25*Tau^2[beta,1]=1

Estimated Power for a Main Effect of
Standardized Size .2306 at alpha=.05:

0.6051

Estimated Power for an Interaction of
Standardized Size .2306 at alpha=.05:

0.1996

		Power
Power for Main Effects by # Clusters:	4	0.5117
	6	0.6846
	8	0.8053
	10	0.8840
	12	0.9329
	14	0.9621
	16	0.9790
	18	0.9886
	20	0.9939
	22	0.9968
	24	0.9983
	26	0.9991

Power

```
Power for Interactions by # Clusters:      4     0.1687
                                           6     0.2305
                                           8     0.2917
                                          10     0.3513
                                          12     0.4087
                                          14     0.4633
                                          16     0.5149
                                          18     0.5630
                                          20     0.6078
                                          22     0.6490
                                          24     0.6869
                                          26     0.7214
```

The predicted value in Table 5 differs very slightly due to rounding. A dataset called WithinClustersPowerFormulaOut is also created, with the following contents.

PRETESTICC	CHANGEICC	PRETEST POSTTESTCORR	NUMFACTORS	MEANNUM MEMBERS
0.05	0.025	0.65	5	50
NUMCLUSTERS	ALPHALEVEL	STANDARDIZED EFFECTSIZE	MODELORDER	NUMPARAMS
5	0.05	0.2306	2	16
SIGMASQD	TAUSQDPI0	TAUSQDBETA1	TAUSQDBETA0	NONCENTRALITY
0.3325	0.6175	0.017051	0.045737	4.99778
ERRORDF	POWERFOR MAIN	POWERFOR INTERACTION	CRITVAL	GAMMA
234	0.60506	0.19962	3.88151	0.1153

The most important entries above are POWERFORMAIN and POWERFORINTERACTION, containing the calculated power estimates for a main effect of the given size and for a two-way interaction of the given size. Intermediary results calculated during the computation of power are also included.

The macro's calculations are done using the appropriate formula from Table 2. In the given example, assuming Model (8), the noncentrality parameter can be calculated as $\frac{N\gamma^2}{2\sigma^2} = \frac{5\times 50\times\left(\frac{0.2306}{2}\right)^2}{2\times 0.3325} = 4.9978$ (left column of Table 2) or equivalently $\frac{Nd^2}{8\frac{\sigma^2}{\sigma^2_{\mathrm{tot}}}} = \frac{5\times 50\times 0.2306^2}{8\times 0.3325} = 4.9978$

(right column of Table 2). The numerator degrees of freedom for testing a single effect here is 1. The number of regression parameters in the model is 16, for 1 intercept, 5 main effects, and $\binom{5}{2} = 10$ interactions. The overall sample size is expected to be about $N = 5 \times 50 = 250$. The denominator (error) degrees of freedom are then $250 - 16 = 234$. The critical value for a $\alpha = .05$ test under the null hypothesis is the constant f_0 such that $P(F_{1,234} > f_0) = .05$ for a central F distribution, and SAS's FINV function finds that this value is 3.8815. The power of the test under the alternative hypothesis is then the probability that a noncentral F distribution with degrees of freedom 1 and 234 and noncentrality parameter 4.9978 will exceed 3.8815. SAS's PROBF function determines that this is $1 - \mathtt{PROBF}(3.8815, 1, 234, 4.9978) = 0.605$.

Between-Clusters Macro

The following is a copy of a similar macro which computes power for *between*-clusters factorial experiments. The current version of this macro is also found at `http://methodology.psu.edu/multilevelfactorial/` . The arguments (inputs) to this macro are the same as those for WithinClustersPower with one addition: StdDevMembers, the anticipated standard deviation of the number of members per cluster. To specify a mean cluster size of approximately $\bar{n}$ and a coefficient of variation of cluster sizes CV_n, specify $\bar{n}$ for MeanNumMembers and $\bar{n}CV_n$ for StdDevMembers. The variability of cluster sizes was not as important in the WithinClustersPower macro because it was assumed that treatments were assigned at the individual rather than cluster level. However, for between-clusters power the distribution of the cluster sizes is more important.

```
%MACRO BetweenClustersPower(PretestICC=,
                            ChangeICC=,
                            PretestPosttestCorr=,
                            NumFactors=,
                            NumClusters=,
                            MeanNumMembers=,
```

```
                            StdDevMembers=,
                            StandardizedEffectSize=,
                            AlphaLevel=.05,
                            ModelOrder=2);
/*
  By John Dziak, The Methodology Center at Penn State.
  Copyright 2012 The Pennsylvania State University.
  This program is free software: you can redistribute it and/or modify
  it under the terms of the GNU General Public License as published by
  the Free Software Foundation, either version 3 of the License, or
  (at your option) any later version.
  This program is distributed in the hope that it will be useful,
  but WITHOUT ANY WARRANTY; without even the implied warranty of
  MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See
  the GNU General Public License for more details
  (http://www.gnu.org/licenses/).

  Second of two SAS Macros to accompany Dziak, Nahum-Shani
and Collins, "Multilevel Factorial Experiments for Developing
Behavioral Interventions".  Finds power per component given
effect size for cluster-randomized factorial designs.
*/
PROC IML;
  /* Initial Calculations of Design Options and Parameters; */
  PretestICC = &PretestICC;
  ChangeIcc = &ChangeICC;
  PretestPosttestCorr = &PretestPosttestCorr;
  NumFactors = &NumFactors;
  NumClusters = &NumClusters;
  MeanNumMembers = &MeanNumMembers;
  StdDevMembers = &StdDevMembers;
  PretestICC = &PretestICC;
  ChangeICC = &ChangeICC;
```

```
  PretestPosttestCorr = &PretestPosttestCorr;
  AlphaLevel = &AlphaLevel;
  StandardizedEffectSize = &StandardizedEffectSize;
  SigmaSqd = (1-PretestPosttestCorr)*(1-PretestICC);
  TauSqdPi0 = PretestPosttestCorr*(1-PretestICC);
  TauSqdBeta1 = 2*SigmaSqd*(ChangeICC)/(1-ChangeICC);
  TauSqdBeta0 = 1 - .25*TauSqdBeta1 - TauSqdPi0 - SigmaSqd;
  Gamma = StandardizedEffectSize/2;
  N = MeanNumMembers*NumClusters;
  CVMembers = StdDevMembers/MeanNumMembers;
/* Input Checking. The conditions are similar to those in
  the other macro.  It is assumed that the standard deviation
  of cluster sizes is not greater than the mean of cluster sizes,
  although such a situation is technically possible in unusual
  or very skewed distributions. */
  try = 1;
  Message1 = "";
  IF ((&PretestICC < 0) | (&PretestICC >= 1)) THEN DO;
  Message1 = "The posttest ICC estimate must be greater than" //
             "or equal to zero and less than one.";
  try = 0;
  END;
  IF ((&ChangeIcc < 0) | (&ChangeIcc >= 1)) THEN DO;
  Message1 = "The change-score ICC estimate must be greater than" //
       "or equal to zero and less than one.";
  try = 0;
  END;
  IF ((&PretestPosttestCorr < 0) | (&PretestPosttestCorr >= 1)) THEN DO;
  Message1 = "The pretest-posttest correlation must be greater than" //
             "or equal to zero and less than one.";
  try = 0;
  END;
  IF ((&NumFactors < 1) | (&NumFactors > 10)) THEN DO;
```

```
    Message1 = "The number of factors should be between 1 and 10.";
    try = 0;
END;
IF ((&NumClusters < 20) | (&NumClusters > 1000)) THEN DO;
    Message1 = "The number of clusters should be between 20 and 1000.";
    try = 0;
END;
IF ((&MeanNumMembers < 1) | (&MeanNumMembers > 10000)) THEN DO;
    Message1 = "The mean cluster size should be between 1 and 10000.";
    try = 0;
END;
IF ((&StdDevMembers < 0) | (&StdDevMembers > &MeanNumMembers)) THEN DO;
    Message1 = "The standard deviation of cluster size should be " //
           "between 0 and &MeanNumMembers.";
    try = 0;
END;
IF ((&StandardizedEffectSize < 0) | (&StandardizedEffectSize > 2)) THEN DO;
Message1 = "The standardized effect size should be between 0 and 2.";
try = 0;
END;
IF ((&alphaLevel <= 0) | (&alphaLevel > .25)) THEN DO;
Message1= "The alpha level should be above 0 and below .25";
try = 0;
END;
/* Check model order (1=fit main effects only,
    2=fit main effects and two-way interactions); */
ModelOrder = %EVAL(&ModelOrder)+0;
IF ((ModelOrder=1) | (ModelOrder=2)) THEN; ELSE DO;
Message1 = "The model order should be 1 or 2.";
try = 0;
END;
/* Determine the error df; */
IF ModelOrder=1 THEN DO;
```

```
    NumParams = 1 + INT(&NumFactors);
     ErrorDF = NumClusters - NumParams;
END;
IF ModelOrder=2 THEN DO;
NumParams = 1 + INT(&NumFactors) +
             INT(&NumFactors)*(INT(&NumFactors) - 1)/2;
ErrorDF = NumClusters - NumParams;
END;
/* Here the power is calculated, using FINV and PROBF.*/
IF Try = 1 THEN DO;
Noncentrality = N * Gamma * Gamma / (2*SigmaSqd +
                MeanNumMembers*(1+CVMembers**2)*TauSqdBeta1) ;
CritVal = FINV(1 - AlphaLevel,1,ErrorDF);
PowerForMain = 1 - PROBF(CritVal,
     1,
     ErrorDF,
     Noncentrality);
PowerForInteraction = 1 - PROBF(CritVal,
     1,
     ErrorDF,
     (Noncentrality)/4);
IF ModelOrder = 1 THEN PowerForInteraction = 0;
/* Now the output is prepared and presented.*/
DesignParameters = PretestICC //
       ChangeIcc //
       PretestPosttestCorr //
       StandardizedEffectSize //
       NumFactors //
       NumParams //
       MeanNumMembers //
       StdDevMembers //
       NumClusters;
r = "Posttest ICC" //
```

```
  "Change ICC" //
  "Pretest-Posttest Corr." //
  "Effect Size" //
  "# Factors in Experiment" //
  "# Parameters in Regression Model" //
  "Mean # Members / Cluster" //
  "Std. Dev. of Members / Cluster" //
  "# Clusters" ;
c = "";
Message2="";
IF (ModelOrder=1) THEN Message2=
      "Assuming a model without interactions," //
      "and a two-sided test for each main effect.";
IF (ModelOrder=2) THEN Message2=
      "Assuming a model with 2-way interactions," //
      "and a two-sided test for each main effect and interaction.";
PRINT DesignParameters[rowname=r colname=c format=12.4
       label="Design Parameters specified by user:"], Message2[label=""];
VarianceParameters = TauSqdPi0 //
       TauSqdBeta0 //
       TauSqdBeta1 //
       SigmaSqd;
r = "Tau^2[pi,0] (Individual)" //  "Tau^2[beta,0] (Cluster)" //
  "Tau^2[beta,1] (Cluster * Time)" // "Sigma^2 (Error)" ;
PRINT VarianceParameters[rowname=r colname=c format=12.4
 label="Variance Components:"] " ";
PRINT("Variance components above have been scaled " //
      "so that total posttest variance is " //
      "Tau^2[pi,0]+Tau^2[beta,0]+Sigma^2+.25*Tau^2[beta,1]=1") ;
PRINT("Estimated Power for a Main Effect of" //
   "Standardized Size &StandardizedEffectSize at alpha=&AlphaLevel: ");
PRINT PowerForMain[label="" format=6.4];
PRINT("Estimated Power for an Interaction of" //
```

```
    "Standardized Size &StandardizedEffectSize at alpha=&AlphaLevel:");
PRINT PowerForInteraction[label="" format=6.4 ];
/* Here a power table is displayed, showing the power that would be obtained
    with different numbers of clusters of the same size. */
PRINT "Please see next page for a power table ..." / ;
NumClustersForTable =  INT(NumParams/5)*5+5*(1:12);
NoncentralityForTable = MeanNumMembers * NumClustersForTable * Gamma * Gamma /
      (2*SigmaSqd + MeanNumMembers*(1+CVMembers**2)*TauSqdBeta1);
ErrorDFForTable = NumClustersForTable - NumParams;
PowerForMainForTable = J(12,1,0);
PowerForInteractionForTable = J(12,1,0);
DO i = 1 TO 12;
/* This is basically the same power calculation as before, but
   repeated for several possible values for the sample size in
   terms of number of clusters. */
  PowerForMainForTable[i] = 1 - PROBF(
             FINV(1 - AlphaLevel,1,ErrorDFForTable[i]),
             1,
             ErrorDF,
             NoncentralityForTable[i]);
  PowerForInteractionForTable[i] = 1 - PROBF(
             FINV(1 - AlphaLevel,1,ErrorDFForTable[i]),
             1,
             ErrorDF,
             NoncentralityForTable[i]/4);
END;
r = CHAR(NumClustersForTable);
PRINT "Power for Main Effects by # Clusters:"
       PowerForMainForTable[r=r c="Power" label=" " format=10.4];
IF ModelOrder=2 THEN DO;
  PRINT "Power for Interactions by # Clusters:"
       PowerForInteractionForTable[r=r c="Power" label=" " format=10.4];
END;
```

```
  CREATE BetweenClustersPowerFormulaOut
    VAR {   PretestICC
      ChangeIcc
      PretestPosttestCorr
      NumFactors
      MeanNumMembers
      StdDevMembers
      NumClusters
      AlphaLevel
      StandardizedEffectSize
      ModelOrder
      NumParams
      SigmaSqd
      TauSqdPi0
      TauSqdBeta1
      TauSqdBeta0
      Noncentrality
      ErrorDF
      PowerForMain
      PowerForInteraction
      CritVal
      Gamma };
  APPEND;
  CLOSE BetweenClustersPowerFormulaOut;
     END;
  ELSE DO;
  PRINT "The macro encountered a problem:";
  PRINT Message1;
  END;
QUIT;
%MEND;
```

The syntax is similar to that for the previous macro, with one additional input argument: StdDevMembers is the assumed standard deviation of number of members per cluster.

Here is an example of calling the between-clusters power formula macro. Suppose a factorial experiment is planned under the assumptions of 5 factors, 25 clusters, 20 members per cluster, a coefficient of variability of cluster size of about .29, a pretest ICC of .05, a change-score ICC of .025, a pretest-posttest correlation of 0.65, and a standardized effect size of about .2306 relative to σ_{tot}. (This scenario corresponds to the condition in the first row of Table 7 in the paper. It is practical only for a fractional factorial since there are fewer than 2^5 clusters available so not all of the cells in the complete factorial could be filled) Suppose the macro file is saved in a folder named `C:\Documents\` . Then use the code

```
%INCLUDE "C:\Documents\BetweenClustersPower.sas";
%BetweenClustersPower( PretestICC=.05,
               ChangeICC=.025,
               PretestPosttestCorr=.65,
               NumFactors=5,
               NumClusters=25,
               MeanNumMembers=20,
           StdDevMembers=5.8,
               StandardizedEffectSize=.2306,
               AlphaLevel=.05,
               ModelOrder=2);
```

to obtain the following output:

```
Posttest ICC                          0.0500
Change ICC                            0.0250
Pretest-Posttest Corr.                0.6500
Effect Size                           0.2306
# Factors in Experiment               5.0000
# Parameters in Regression Model     16.0000
Mean # Members / Cluster             20.0000
```

```
     Std. Dev. of Members / Cluster         5.8000
     # Clusters                            25.0000
Assuming a model with 2-way interactions,
and a two-sided test for each main effect and interaction.


     Variance Components:

     Tau^2[pi,0] (Individual)             0.6175
     Tau^2[beta,0] (Cluster)              0.0457
     Tau^2[beta,1] (Cluster * Time)       0.0171
     Sigma^2 (Error)                      0.3325


  Variance components above have been scaled
  so that total posttest variance is
  Tau^2[pi,0]+Tau^2[beta,0]+Sigma^2+.25*Tau^2[beta,1]=1


         Estimated Power for a Main Effect of
         Standardized Size .2306 at alpha=.05:


                         0.6178


         Estimated Power for an Interaction of
         Standardized Size .2306 at alpha=.05:


                         0.2057
```

```
Power for Main Effects by # Clusters:        20      0.3600
                                             25      0.6178
                                             30      0.7324
                                             35      0.8071
                                             40      0.8600
                                             45      0.8984
                                             50      0.9265
                                             55      0.9470
                                             60      0.9620
                                             65      0.9728
                                             70      0.9806
                                             75      0.9862


                                                      Power

Power for Interactions by # Clusters:        20      0.0936
                                             25      0.2057
                                             30      0.2677
                                             35      0.3164
                                             40      0.3593
                                             45      0.3986
                                             50      0.4353
                                             55      0.4699
                                             60      0.5025
                                             65      0.5335
                                             70      0.5627
                                             75      0.5905
```

The dataset BetweenClustersPowerFormulaOut is also created, with the following contents.

PRETEST ICC	CHANGE ICC	PRETEST POSTTESTCORR	NUMFACTORS	MEANNUM MEMBERS
0.05	0.025	0.65	5	20
STDDEV MEMBERS	NUMCLUSTERS	ALPHALEVEL	STANDARDIZED EFFECTSIZE	MODEL ORDER
5.8	25	0.05	0.2306	2
NUMPARAMS	SIGMASQD	TAUSQDPI0	TAUSQDBETA1	TAUSQDBETA0
16	0.3325	0.6175	0.017051	0.045737
NON CENTRALITY	ERRORDF	POWERFOR MAIN	POWERFOR INTERACTION	CRITVAL
6.42409	9	0.61784	0.20574	5.11736
GAMMA				
0.1153				

The variance components above, like "SIGMASQD" and "TAUSQDPI0," are expressed as proportions of σ^2_{tot}. Note that $\sigma^2_{\text{tot}} = \tau^2_{\beta 0} + \frac{1}{4}\tau^2_{\beta 1} + \tau^2_{\pi 0} + \sigma^2$ so $\frac{\tau^2_{\beta 0}}{\sigma^2_{\text{tot}}} + \frac{1}{4}\frac{\tau^2_{\beta 1}}{\sigma^2_{\text{tot}}} + \frac{\tau^2_{\pi 0}}{\sigma^2_{\text{tot}}} + \frac{\sigma^2}{\sigma^2_{\text{tot}}} = 1$ (see Appendix B). Accordingly, in the above output table, TAUSQDBETA0 + $\frac{1}{4}$ TAUSQDBETA1 + TAUSQDPI0 + SIGMASQD = 1.

The macro's calculations were done using the appropriate formula from Table 2. Consider the first row of Table 7, with 5 factors, 25 clusters, 20 members per cluster, a pretest-posttest correlation of 0.65, a pretest ICC of 0.05, a change score ICC of 0.025, a coefficient of variability of cluster size of about 0.29, and standardized effect size $d \approx 0.23$. First, assuming Model (9), the noncentrality parameter is calculated as $\lambda = \frac{Nd^2}{8\frac{\sigma^2}{\sigma^2_{\text{tot}}} + 4\bar{n}\left(\text{CV}^2_n + 1\right)\frac{\tau^2_{\beta 1}}{\sigma^2_{\text{tot}}}}$. From Appendix B, $\frac{\sigma^2}{\sigma^2_{\text{tot}}} = (1 - \rho_{\text{pre,post}})(1 - \rho_{\text{pre}})$ where we assumed $\rho_{\text{pre,post}} = .65$, $\rho_{\text{pre}} = .05$, so $\frac{\sigma^2}{\sigma^2_{\text{tot}}} = .3325$. Next, from Appendix B, $\frac{\tau^2_{\beta 1}}{\sigma^2_{\text{tot}}} = 2\left(\frac{\sigma^2}{\sigma^2_{\text{tot}}}\right)\left(\frac{\rho_{\text{change}}}{1-\rho_{\text{change}}}\right)$ where we assumed $\rho_{\text{change}} = .025$, so $\frac{\tau^2_{\beta 1}}{\sigma^2_{\text{tot}}} = 2\,(0.3325)\left(\frac{0.025}{1-0.025}\right) = 0.01705$. The overall sample size is $N = 25 \times 20 = 500$. Then $\lambda = \frac{500 \times 0.2306^2}{8(0.3325)+4(20)(0.29^2+1)(0.01705)} \approx 6.42$. The numerator degrees of freedom for testing a single effect here is 1. The number of regression parameters in the model is 16, for 1 intercept, 5 main effects, and $\binom{5}{2} = 10$ interactions. The number of clusters is 25. The denominator (error) degrees of freedom are then $25 - 16 = 9$. The critical value for an $\alpha = .05$ test under the null hypothesis is the constant f_0 such that $P(F_{1,9} > f_0) = .05$ for a central F distribution, and SAS's FINV function finds that this value is 5.117. The power of the test under the alternative hypothesis is then the probability that a noncentral

F distribution with degrees of freedom 1 and 9 and noncentrality parameter 6.42 will exceed 5.117. SAS's PROBF function determines that this is 0.618.

SAS Code for Analyzing Simulated Data

We used the following SAS code for the within-clusters adjusted two-level models:

```
PROC MIXED DATA=randomDataWide NOCLPRINT NOITPRINT ;
    CLASS clusterID personID;
    MODEL yPost = yPre F1|F2|F3|F4|F5@2 ;
    RANDOM INT / SUBJECT=clusterID;
    ODS OUTPUT COVPARMS=theseParamsAncova TESTS3=theseTestsAncova;
RUN;
```

For between-clusters scenarios, we replaced the RANDOM command above with `REPEATED / SUBJECT=clusterID(condition) TYPE=CS`; where condition was a `CLASS` variable labeled 1 through 16 or 1 through 32 for each cell. Instead of `(condition)` we could have used `(F1*F2*F3*F4*F5)`, except that PROC MIXED would not allow nesting within a numerical variable. We did not want to specify the factors as `CLASS` variables because they might then be internally dummy-coded rather than effect-coded; this would not matter for the ANOVA tests of interest, at least if Type 3 sum squares are being used, but it still seemed better to remain with effect coding.

The corresponding code used for the within-cluster three-level models was:

```
PROC MIXED DATA=randomDataLong NOCLPRINT NOITPRINT ;
    CLASS clusterID personID F1 F2 F3 F4 F5;
    MODEL y = time time*F1 time*F2 time*F3 time*F4 time*F5
              time*F1*F2 time*F1*F3 time*F1*F4 time*F1*F5
              time*F2*F3 time*F2*F4 time*F2*F5
              time*F3*F4 time*F3*F5
              time*F4*F5;
    RANDOM INT time / SUBJECT=clusterID;
```

```
            RANDOM INT / SUBJECT=personID(clusterID);
            ODS OUTPUT COVPARMS=theseParams TESTS3=theseTests;
        RUN;
```

For reasons described above, although it should not matter for the test output shown in the paper, it might be better to exclude `F1 F2 F3 F4 F5` from the `CLASS` statement above if the values of coefficients are to be interpreted.

For between-clusters models, the code used was

```
PROC MIXED DATA=randomDataLong NOCLPRINT NOITPRINT ;
            CLASS clusterID personID condition;
            MODEL y = time  time*F1 time*F2 time*F3 time*F4 time*F5
                      time*F1*F2 time*F1*F3 time*F1*F4 time*F1*F5
                      time*F2*F3 time*F2*F4 time*F2*F5
                      time*F3*F4 time*F3*F5
                      time*F4*F5  /   DDF=&df,&df,&df,&df,&df,
          &df,&df,&df,&df,&df,
            &df,&df,&df,&df,&df,&df ;
            RANDOM INT time / SUBJECT=clusterID(condition) ;
            RANDOM INT / SUBJECT=personID(clusterID);
            ODS OUTPUT COVPARMS=theseParams TESTS3=theseTests;
        RUN;
```

where the macro variable `&df` had previously been defined as the number of clusters minus 16 (the number of effects in the second-order model). Code like

```
PROC MIXED DATA=here.randomDataLong NOCLPRINT NOITPRINT ;
            CLASS clusterID personID F1 F2 F3 F4 F5;
            MODEL y = time time*F1 time*F2 time*F3 time*F4 time*F5
                      time*F1*F2 time*F1*F3 time*F1*F4 time*F1*F5
                      time*F2*F3 time*F2*F4 time*F2*F5
                      time*F3*F4 time*F3*F5
                      time*F4*F5 ;
```

```
          RANDOM INT time / SUBJECT=clusterID(F1*F2*F3*F4*F5) ;
          RANDOM INT / SUBJECT=personID(clusterID);
          ODS OUTPUT COVPARMS=theseParams TESTS3=theseTests;
  RUN;
```

would give the same Type III significance test results but different estimates and interpretations for the model coefficients because F1 through F5 would be automatically recoded.

These coefficients can be obtained using the `/SOLUTION` option in the `MODEL` statement. The R code (R Development Core Team, 2010; Pinheiro et al., 2010) for the adjusted two-level model, with X variables coded as -1 and +1, was

```
lme( fixed = yPost ~ yPre + f1 + f2 + f3 + f4 + f5 +
   f1*f2 + f1*f3 + f1*f4 + f1*f5 + f2*f3 + f2*f4 + f2*f5 +
   f3*f4 + f3*f5 + f4*f5 , random = ~ 1 | clusterId, data=sim.data)
```

in either the between or within clusters case. In either SAS or R, additional covariates such as gender, and possibly their interactions with the factors, could also be included in the model statements along with the factors. In this case the investigator should check to make sure that the degrees of freedom are being counted correctly (see Murray 1998), i.e., that the software is correctly distinguishing between individual-level and cluster-level covariates.

References

Murray, D. M. (1998). *Design and analysis of cluster-randomized trials.* New York: Oxford.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & The R Development Core Team (2010). *nlme: Linear and Nonlinear Mixed Effects Models.* R package version 3.1-97. Retrieved from `http://cran.r-project.org/web/packages/nlme/index.html`

R Development Core Team (2010). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from `http://cran.r-project.org/`

SAS Institute. (2004). *SAS/QC 9.1 users guide.* Cary, NC: Author.