American Journal of Public Health

Simulation-based Power Analysis for Detecting Interactions in Longitudinal Data Analysis

--Manuscript Draft--

Manuscript Number:	
Article Type:	Statistically Speaking
Keywords:	Power Analysis; Sample Size Calculations; Longitudinal Data Analysis; Simulation; Statistical Computing; Interaction Effects; Multilevel Models
Corresponding Author:	Brady T. West, Ph.D. University of Michigan-Ann Arbor Ann Arbor, MI UNITED STATES
First Author:	Brady T. West, Ph.D.
Order of Authors:	Brady T. West, Ph.D.
	Sean Esteban McCabe, Ph.D.
Abstract:	Abstracts are not required for Statistically Speaking articles.
Full Title:	Simulation-based Power Analysis for Detecting Interactions in Longitudinal Data Analysis
Manuscript Classifications:	68: Statistics/Evaluation/Research
Author Comments:	This article is being re-submitted after discussions with Dr. Vaughan and Dr. Northridge about the reviews of the initial submitted article. Our cover letter discusses this in detail, including our responses to the comments from the first two reviewers. Also, please note that we have once again provided the SAS macros in a supplemental file, and that we would like the reviewers to have access to these SAS macros.
Suggested Reviewers:	
Additional Information:	
Question	**Response**
Theme Issue Question	Technological Innovations in Public Health
Please list the Abstract word count for this document. This count is forwarded to the Editors and Reviewers as part of the evaluation process.	N/A
Please list the main text word count for this document. This count excludes the abstract, figures and tables, and references and is forwarded to the Editors and Reviewers as part of the evaluation process.	982
Please list the number of references in this document. This count is forwarded to the Editors and Reviewers as part of the evaluation process.	13
Please list the number of Tables and Figures and Supplemental Tables and Figures in this document. This count is forwarded to the Editors and Reviewers as part of the evaluation process. Please report counts as follows: Figures: #; Tables: #; Supplemental Tables: #; Supplemental Figures: # (where #	0

represents the actual count of submitted Tables and Figures).	
Affiliations Statement	Brady T. West is with the Survey Research Center and the Center for Statistical Consultation and Research at the University of Michigan-Ann Arbor. Sean E. McCabe is with the Institute for Research on Women and Gender at the University of Michigan-Ann Arbor.
Acknowledgments and Disclosure	Brady T. West and Sean E. McCabe were supported by research grant R01DA031160 from the National Institute on Drug Abuse, National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Drug Abuse or the National Institutes of Health.
Human Participant Protection statement	

July 15, 2014

To the Editors of AJPH,

On behalf of my co-author Dr. Sean Esteban McCabe, I am re-submitting this article to be considered once again for publication in the "Statistically Speaking" section of AJPH. This manuscript is in compliance with the *Principles of the Ethical Practice of Public Health*, and has less than 1,000 words of text, with no tables or figures. We do include more than five references, but we feel that all of these references are directly relevant to the issue being discussed.

In the initial round of reviews, a first reviewer was very positive about the article, stating “This short note presents an entirely reasonable simulation-based approach to conducting power analyses in settings where simpler tools are unavailable. This subject and the advice given are likely to be broadly interesting.” The first reviewer also provided suggestions for improvement that have been addressed in this revised article. A second reviewer questioned the unique contribution of the article, citing a 2013 *Statistica Sinica* paper (which was cited in the original submission) and the supplementary SAS code provided for that paper. We have more thoroughly acknowledged the importance of that work and clearly indicated the contribution of our proposed approach for non-statisticians who might find it difficult to modify the code in the 2013 paper for the purpose of testing interactions. We clearly state that we present a simulation-based approach as an alternative for more specific tests of interest.

This article stems from several years of collaborations, consultations, and grant proposals, where we have both frequently encountered stress and difficulty at the study design stage due to the need to perform power calculations for complex analytic techniques. A review of the literature does not find much support for non-statisticians, and we have found simulations to be very useful in these settings over the years. We wrote this article in an effort to assist public health researchers in thinking about how to perform these kinds of simulations in more complex settings, and we provide free software code (in SAS) that can be used by readers to perform the simulations.

Please do not hesitate to contact us with any questions about this article. We look forward to receiving additional feedback on the revised article.

Best Regards,

Brady West

Brady T. West

bwest@umich.edu

Simulation-based Power Analysis for Detecting Interactions in Longitudinal Data Analysis

Public health researchers frequently need to perform power and sample size calculations when planning studies and determining study budgets. When study designs are fairly simple (e.g., simple random sampling), the corresponding statistical tests designed to answer research questions may be relatively straightforward (e.g., two-sample t-tests for comparing means in two groups). In these settings, researchers can take advantage of well-known results, including "canned" software designed to facilitate these calculations (e.g., nQuery Advisor), to perform power calculations.[1] However, many study designs, especially those resulting in clustered or longitudinal data sets, tend to be more complex, and the corresponding analytic techniques (e.g., multilevel models including interaction terms) may not lend themselves to "easy" power analyses at the design stage.

We consider a longitudinal study where researchers are interested in testing the interaction between a categorical "group" variable and time. For example, researchers may be interested in testing whether two groups follow different trends over time for a measure of interest. Three key aspects of this type of study complicate power calculations: 1) the repeated measurements introduce a non-zero correlation in measures collected on the same individual, eliminating power analysis methods that assume independent observations; 2) the repeated measures themselves may not necessarily be normally distributed; and 3) the specific parameter of interest is an *interaction* coefficient capturing differences in effects between groups, rather than a "main effect." Similar issues apply to clustered study designs (e.g., people in neighborhoods) in cross-sectional settings. The question of how to calculate power in these more complex settings has been addressed by several previous authors,[2-9] but none of these studies provide user-friendly tools that simultaneously handle all three aspects discussed above, and some of the studies only provide analytic results without accompanying software.

Li and McKeague[10] have provided the most general and widely applicable approach to this problem to date, presenting important theoretical developments and illustrations of a fast and accurate methodology for power and sample size calculations based on Generalized Estimation Equations

(GEE). These authors also provide supplementary SAS code implementing their methodology, but the code only allows for a single binary covariate in a logistic regression model fitted using GEE. Analysts interested in testing specific interaction coefficients would need to carefully modify this code, consistent with the theory presented in this paper, to meet their needs. We outline a flexible, simulation-based alternative to this method and the other important methods developed to date that can be easily implemented by non-statisticians designing longitudinal studies with complex time-by-group interactions in mind.

We begin with a general recipe for developing simulations in these more complex settings:[11]

1. Write down the model of interest, including expected values for the coefficients in the model based on analyses of prior data or subject matter knowledge.
2. Estimate the correlation of repeated measures collected on the same individual or from the same cluster (again from prior data), and translate this correlation to a between-individual / between-cluster variance component.
3. Repeatedly simulate samples from the specified model using alternative sample size values (given the specified coefficients and variance component), fit the model of interest to the data simulated from each sample, and test the specific interaction coefficient(s) for significance at a pre-specified significance level (alpha).
4. Record the proportion of samples where the interaction was found to be significant at the specified alpha level; this is the empirical power of a given sample size choice.

We now consider an example. Suppose that a researcher is interested in being able to detect different changes over time in the prevalence of past-year ecstasy (MDMA) use among high school seniors. The researcher has baseline information available, where based on samples of 2,300 male and 3,500 female high school seniors, the respective prevalence estimates are 11.5% and 6.5%.[12-13] The researcher believes that the prevalence for males will drop markedly over the next year (to 6.3%), while the prevalence for females will stay about the same (6.2%), and wants to know whether having 1,000 males and 1,000 females with data at both time points will be sufficient for detecting this time-by-gender interaction effect as being significant at the 0.05 level. Assuming that all baseline high school seniors respond at both time points and fitting a simple logistic regression model to these hypothetical data, we obtain estimated coefficients of -0.050 (for time), 0.626 (for male), -0.608 (for the interaction term), and -2.666 (for the intercept). We therefore want to detect an interaction of this magnitude as being significant at the 0.05 level, but also adjusting for an expected within-person correlation in the binary outcomes (ecstasy use) of 0.2. This correlation (ICC) translates to be a between-person variance component (VC) of 0.822 [where ICC = VC / (VC + (3.1415^2 / 3))]. We plug these numbers into the SAS macro available at [web site], as follows, requesting 100 simulated samples:

```
%geesim(nreps=100, nint=1000, bsubvar=0.822, int=-2.666,
betat=-0.050, betag=0.626, betatg=-0.608, alpha=0.05);
```

The resulting simulated power is 0.74 (or 74%), suggesting that a larger number of male and female respondents with data at both time points may be needed to detect this interaction at the 0.05 level. The 100 simulated samples took less than 20 seconds, and more samples could be used if one desires calculations with a higher level of precision.

Public health researchers may also be interested in fitting models to repeatedly-measured continuous outcomes with normally-distributed residuals. In addition to the case of testing a time-by-group interaction in a logistic regression model for repeated measures of binary data, we also provide a simulation macro (at the web site indicated above) for a normally distributed outcome, with time and a normally-distributed time-varying predictor as covariates. These SAS macros are freely available and can be modified depending on the research context, or adapted to other software packages. We hope that these examples will motivate additional exchange of software and ideas for performing these simulations, and simplify this important aspect of study design for public health researchers.

References

1. Weller SC. Sample size estimation: The easy way. *Field Methods*. 2014; doi: 10.1177/1525822X14530086.

2. Donohue MC, Edland SD, Gamst AC. Power for linear models of longitudinal data with applications to Alzheimer's Disease Phase II study design (Technical Report, Division of Biostatistics and Bioinformatics, University of California, San Diego). http://cran.r-project.org/web/packages/longpower/vignettes/longpower.pdf. Published September 14, 2013. Accessed June 13, 2014.

3. Spybrook J, Bloom H, Congdon R, Hill C, Martinez A, Raudenbush S. Optimal Design Plus Empirical Evidence: Documentation for the “Optimal Design” Software. http://hlmsoft.net/od/od-manual-20111016-v300.pdf. Published October 16, 2011. Accessed June 13, 2014.

4. Demidenko E. Sample size and optimal design for logistic regression with binary interaction. *Statistics in Medicine*. 2008;27:36-46.

5. Jung S-H, Ahn C. Sample size estimation for GEE method for comparing slopes in repeated measurement data. *Statistics in Medicine*. 2003;22:1305-1315.

6. Hsieh FY, Lavori PW, Cohen HJ, Feussner JR. An overview of variance inflation factors for sample-

size calculation. *Evaluation and the Health Professions*. 2003;26:239-257.

7. Rochon J. Application of GEE procedures for sample size calculations in repeated measures experiments. *Statistics in Medicine*. 1998;17:1643-1658.

8. Liu G, Liang K-Y. Sample size calculations for studies with correlated observations. *Biometrics*. 1997;53(3):937-947.

9. Lipsitz SR, Fitzmaurice, GM. Sample size for repeated measures studies with binary responses. *Statistics in Medicine*. 1994;13:1233-1239.

10. Li Z, McKeague IW. Power and sample size calculations for generalized estimating equations via local asymptotics. *Statistica Sinica*. 2013;23:231-250.

11. Gelman A, Hill J. *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge, NY: Cambridge University Press; 2006.

12. Johnston LD, O'Malley PM, Bachman JG. *Monitoring the Future national survey results on drug use, 1975-2000 (Volume I: Secondary school students (NIH Publication No. 01-4924))*. Bethesda, MD: National Institute on Drug Abuse; 2001.

13. Johnston LD, O'Malley PM, Bachman JG. *Monitoring the Future national survey results on drug use, 1975-2001 (Volume I: Secondary school students (NIH Publication No. 02-5106))*. Bethesda, MD: National Institute on Drug Abuse; 2002.

```
/* Simulation program that computes empirical power for GEE logistic regression models fitted to
longitudinal data sets, with two binary covariates (time and group) and their interaction */

/* Author: XXXXX */

/* Date: June 5, 2014 */

/* Simulation macro */

/* INPUT PARAMETERS:

nreps = number of simulations
nint = number of subjects in each group (e.g., number of intervention cases, number of control
cases)
bsubvar = between-subject variance (defined as (rho x pi^2/3) / (1 - rho), rho is exchangeable
correlation)
int = desired intercept coefficient
betat = desired coefficient for time (post = 1, pre = 0) when group = 0 (time effect for control
group)
betag = desired coefficient for group (int = 1, cont = 0) when time = 0 (group effect at time 1)
betatg = desired interaction coefficient for time x group
alpha = desired significance level

*/

%macro geesim(nreps, nint, bsubvar, int, betat, betag, betatg, alpha);

%let jplusone = %eval(&nint + 1);
%let total = %eval(2 * &nint);

/* Repeatedly simulate samples based on the given model */

%do reps = 1 %to &nreps;

data simsample (drop = j k u2);
   do j = 1 to &nint;
      intid = j;
          u2 = sqrt(&bsubvar)*rannor(0);
          do k = 1 to 2; /* NOTE: two time points per subject */
                 rand = ranuni(0);
                 group = 0;
                 if k = 1 then post = 0;
                 else if k = 2 then post = 1;
                 p2 = exp(&int + &betat * post + &betag * group + &betatg * group*post + u2) / (1
+ exp(&int + &betat *post + &betag * group + &betatg *group*post + u2));
              if rand <= p2 then trbz_bin = 1; /* binary outcome */
                 else trbz_bin = 0;
         output;
          end;
   end;
   do j = &jplusone to &total;
      intid = j;
          u2 = sqrt(&bsubvar)*rannor(0);
          do k = 1 to 2;
                 rand = ranuni(0);
                 group = 1;
                 if k = 1 then post = 0;
                 else if k = 2 then post = 1;
                 p2 = exp(&int + &betat * post + &betag * group + &betatg * group*post + u2) / (1
+ exp(&int + &betat *post + &betag * group + &betatg *group*post + u2));
              if rand <= p2 then trbz_bin = 1;
                 else trbz_bin = 0;
                 output;
          end;
   end;
run;

/* Fit the model of interest to a simulated sample */

proc genmod data = simsample desc;
   class intid;
```

```
   model trbz_bin = group post group*post / dist = bin link = logit;
   repeated subject = intid / corr = exch corrw;
   ods output GEEEmpPEst = parmest;
run;

/* Determine whether the interaction of interest was significant at the alpha level */

%if &reps = 1 %then %do;
data parmests (keep = probz power_bin);
   set parmest;
   if probz < &alpha then power_bin = 1;
   else power_bin = 0;
   if Parm = "group*post";
run;
%end;
%else %do;
data parmest (keep = probz power_bin);
   set parmest;
   if probz < &alpha then power_bin = 1;
   else power_bin = 0;
   if Parm = "group*post";
run;
data parmests;
   set parmests parmest;
run;
%end;

%end;

/* Compute empirical power based on the simulated samples */

proc means data = parmests mean;
   var power_bin;
run;

%mend geesim;

/* Example of macro call */

%geesim(nreps=100, nint=1000, bsubvar=0.822, int=-2.666, betat=-0.050, betag=0.626, betatg=-
0.608, alpha=0.05);



/* Simulation program that computes empirical power for GEE linear regression models fitted to
longitudinal data sets, with general sets of normally distributed covariates */

/* Author: XXXXX */

/* Date: June 5, 2014 */

/* Simulation macro */

/* INPUT PARAMETERS:

nreps = number of simulations
nint = number of subjects on which the dependent variable will be measured
bsubvar = variance of random intercepts
residvar = variance of residuals
int = desired intercept coefficient
betat = desired coefficient for time
betaPA = desired coefficient for a time-varying covariate (PA)
alpha = desired significance level

*/

%macro geenormsim(nreps, nint, bsubvar, residvar, int, betat, betaPA, alpha);

/* Repeatedly simulate samples based on the given model */

%do reps = 1 %to &nreps;
```

```
data simsample (drop = j u2);
   do j = 1 to &nint;
      intid = j;
         u2 = sqrt(&bsubvar)*rannor(0); * Random subject intercept;
         do k = 1 to 12; * 12 time points;
            PA = 3.00 + 1.1475*rannor(0); * Note PA independent of time, with mean 3 and SD 
1.1475;
               engage = &int + &betat * k + &betaPA * PA + u2 + sqrt(&residvar)*rannor(0);
         output;
         end;
   end;
run;

/* Fit the model of interest to a simulated sample */

proc genmod data = simsample;
   class intid; *intid = person ID;
   model engage = k PA;
   repeated subject = intid / corr = exch corrw;
   ods output GEEEmpPEst = parmest;
run;

/* Determine whether the coefficient of interest was significant at the alpha level */

%if &reps = 1 %then %do;
data parmests (keep = probz power_bin);
   set parmest;
   if probz < &alpha then power_bin = 1;
   else power_bin = 0;
   if Parm = "PA"; * predictor that you want to assess power for;
run;
%end;
%else %do;
data parmest (keep = probz power_bin);
   set parmest;
   if probz < &alpha then power_bin = 1;
   else power_bin = 0;
   if Parm = "PA"; * predictor that you want to assess power for;
run;
data parmests;
   set parmests parmest;
run;
%end;

%end;

/* Compute empirical power based on the simulated samples */

proc means data = parmests mean;
   var power_bin;
run;

%mend geenormsim;

/* Example of macro call */

%geenormsim(nreps=100, nint=50, bsubvar=0.015, residvar=0.112, int=1.38, betat=2, betaPA=0.10, 
alpha=0.05);
```