

Estimating Power for Interaction Tests in Logistic Regression: A Case-Study of Tobacco Cessation among Cancer Survivors

Zoran Bursac¹, D. Keith Williams², C. Heath Gauss², Robert Klesges¹

¹Center for Population Sciences, Department of Preventive Medicine, College of Medicine, University of Tennessee Health Science Center, 66 N. Pauline St., Memphis, TN

²Department of Biostatistics, College(s) of Medicine and Public Health, University of Arkansas for Medical Sciences, 4301 W. Markham, Little Rock, AR

Abstract

At the present time, researchers are limited in available methods to conduct power analysis for an interaction term between two main variables of interest in a study that utilizes logistic regression. We propose a method and a SAS macro tool for estimating the power associated with an interaction term in a logistic regression model. This method empirically calculates the power for an interaction term, based on cell counts from a 2x2x2 table, and several other intuitive input parameters. We illustrate the method with an example from a randomized controlled trial (RCT) of tobacco cessation among cancer survivors, which investigates interaction between two-level treatment assignment and cancer staging.

Key Words: Logistic regression, interaction, power

1. Introduction

Power analysis is most effective when carried out at the study planning stage. When researchers are planning a study, they commonly must determine an adequate sample size required for their selected statistical test to have the power to detect an effect in the sample population if one exists. Presently there are available methods researchers can employ to conduct power analyses when the statistical method of choice is logistic regression. However, researchers are currently limited if they want to power a study using logistic regression if they suspect a significant interaction, or effect modification, exists between two explanatory variables to be collected during the course of the study.

Due to the lack of available methods, the objective of this project was to develop a practical and straightforward method of estimating power for a proposed study utilizing logistic regression when researchers anticipate a significant interaction in their study results.

The motivating example behind this project originated from a grant proposal in which researchers were interested in exploring the effect of cancer stage at diagnosis on the relationship between smoking cessation intervention and continuous smoking abstinence. They suspected that there may be an interaction between cancer stage and intervention on the smoking abstinence outcome. In other words, researchers suspected that those patients with early stage diagnosis might have higher odds of cessation as a result of active treatment compared to those patients with the late stage diagnosis. However, no known research had investigated this relationship in this exact setting.

In order to conduct this type of study, investigators must determine an appropriate level of power to detect the effects of treatment and cancer stage in the presence of the interaction between the two variables. That is to say, researchers need to know the power of the statistical method to detect the differential effect of stage on smoking abstinence among individuals receiving active treatment.

A SAS macro has been developed to conduct power analysis for logistic regression models (Northern, Williams, & Bursac, 2009; Williams, Bursac, & Wooten, 2010). However, the existing version of the macro does not carry out power analysis for an interaction term. We have modified this existing macro to incorporate an interaction term within the model and to calculate sufficient power for investigators who suspect a significant interactive effect between two study variables.

2. Background

2.1 SAS PROC POWER

The Power Procedure in SAS (PROC POWER) provides power and sample size analyses for a variety of statistical methods. PROC POWER since SASv.9.2 offers power analysis for logistic regression; one can execute power and sample size analyses for the chi-square likelihood ratio test. Users also have the option of adding covariates or potential confounders to the model. However, this approach allows for only a single predictor variable in binary logistic regression. In addition, SAS assumes independence among the explanatory variables. Because of this, PROC POWER is not appropriate for studies with some correlation between explanatory variables, which characterize most observational studies, i.e., studies that are common in the public health setting (SAS Institute, 2015).

The LOGISTIC statement is used to carry out power analysis for a logistic regression. Users specify a distribution for the predictor variable, either regression coefficients or odds ratios associated with the predictor, an alpha level, the sample size(s), and the power. One selects the desired result, either power or sample size, by designating the parameter as a missing value. However, PROC POWER does not have the capability to compute power with the addition of an interaction term to a logistic regression model (SAS Institute, 2015).

2.2 SAS POWERLOG Macro

The SAS POWERLOG macro was developed by Michael Friendly (1998) and calculates the sample size needed to reach specified power values for a logistic regression model that contains a single predictor variable of interest in the presence of one or more covariates. POWERLOG also allows users to specify correlation between the predictor and the covariate(s). Results are presented both as a table and a graphical display of sample sizes for a range of power values specified in the macro. Friendly's POWERLOG macro works well for calculating power for logistic regression and is relatively simple to understand. However, similar to PROC POWER, the POWERLOG macro does not allow for the addition of an interaction term to the model (Friendly, 1998).

2.3 Power Analysis and Sample Size Software (PASS)

PASS (2013) is interactive software designed for power analysis. It allows the user to solve for power, sample size, effect size and alpha level, as well as validate these parameters once data have been collected. Output is displayed as text summaries and numeric tables as well as charts and graphs. Summary statistics presented are clear, concise and easy to interpret. Output is produced in a portable format that is cut and paste compatible with many other programs. PASS can perform power and sample size analysis for a logistic regression with a single predictor of interest. Yet, as with the other methods mentioned above, PASS cannot compute power for logistic regression with an interaction term in the model (Hintze, 2013).

2.4 Demidenko's Applet (2008)

Research by Eugene Demidenko (2008) discusses sample size for a logistic regression with interaction present. He presents a method of estimating power for proposed studies using a logistic regression model with a binary explanatory variable and a covariate, and their interaction. He has developed an online application (an applet) that will carry out power or sample size calculations, based on input parameters,

required for a logistic regression with an interaction term in the model. The applet can be accessed at <http://www.dartmouth.edu/~eugened/power-samplesize.php> (Demidenko, 2008).

The applet allows a user to calculate power (or a sample size) for two binary explanatory variables and their interaction. Thus, if researchers anticipate two variables may interact, they can determine the appropriate sample size required or the desired power for a specified sample size by accessing this applet (Demidenko, 2008).

Figure 1 presents an image of Demidenko's applet. Users specify the alpha level, the desired power or sample size, and the odds ratio they want to detect if an effect is present. If power is specified, the applet returns the sample size required to reach the specified power level. If sample size is specified, the applet returns the power level that would be attained if a proposed study is conducted and the specified sample size and other input parameters are observed (Demidenko, 2008).

Significance level = 0.05

What to compute:
 Power = 0.896
 Sample size = 1500
 Detectable/alternative OR = 4.7

One variable with exposure, x:
 $Prx=Pr(x=1)=$ $Pry=Pr(y=1|x=0)=$

Two variables with exposure, x and confounder, z:
 $Prx=Pr(x=1)=$ $Prz=Pr(z=1)=$ $ORyz=$ $ORxz=$
 Disease prevalence rate $Pry=Pr(y=1|x=0,z=0)=$

Two binary variables, x and z, with their interaction, x^*z :
 $Prx=Pr(x=1)=$ 0.66 $ORyx=$ 1 $Prz=Pr(z=1)=$ 0.6 $ORyz=$ 1.07 $ORxz=$ 1
 Disease prevalence rate $Pry=Pr(y=1|x=0,z=0)=$ 0.05

Figure 1. Example of Demidenko's Applet Calculations of Power for the Interaction Term in 2x2x2 Case

Other input parameters are required to return power for a study with an interaction term. These are the estimated probabilities for each of the explanatory variables of interest, the effects of each of the explanatory variables on the outcome in the presence of the interaction in the form of odds ratios, the association between the two explanatory variables in the form of an odds ratio, and the probability of the outcome being a "1" (the outcome of interest) when both explanatory variables are zero. The drawback to Demidenko's method is that these required input parameters, in the form of odds ratios, may be difficult to conceptualize and entail some calculations that may be challenging. Therefore, we propose a method to estimate power for a logistic regression with interaction that is perhaps more user-friendly than Demidenko's method. Our method will estimate power derived from a sample size and several other input parameters within a SAS macro for a specified interaction between two explanatory variables.

3. Methods

3.1 Assessing Model Fit

In general, there are four groups of statistical tests used to conduct power and sample size analysis for logistic regression with interaction (Demidenko, 2008). Two of these methods are discussed here as they are the theoretical concepts behind Demidenko's and our approaches to estimating power for a logistic regression with interaction. These are the Wald test and the likelihood ratio test (LRT).

Both power calculation tests use the likelihood to assess the fit of the models being compared. The likelihood is a function of the estimates of the beta coefficients and the data. The goal when fitting a model is to determine values of the parameter estimates, or beta coefficients, that maximize the value of the likelihood function. In other words, it optimizes a set of parameter estimates that yield the most likely data. Hosmer, Lemeshow, and Sturdivant (2013) describes this as the "most parsimonious, yet biologically reasonable model" that describes a relationship between an outcome and a set of explanatory variables.

According to Demidenko (2008), the statistical test selected to compute power should be the equivalent of the type used to test the null hypothesis. He argues that since the Z-test is a common method to test significance of beta coefficients or the fit of the model, the same test should be used to conduct power analysis or determine sample size. Consequently, Demidenko asserts that the Wald test is the appropriate test statistic when estimating power for logistic regression. Demidenko (2008) further asserts that, although many maintain that both tests are equivalent in large samples, the Wald and LRT may produce comparable results when the null hypothesis is true but yield considerable differences when the alternative hypothesis is true.

An advantage of the Wald test is that it entails estimating one model, whereas the likelihood ratio test estimates two models. The null hypothesis under the Wald test is that a set of parameters, in this case, beta coefficients, is equal to some value, commonly zero. The Wald test assesses how far from zero the estimated beta coefficients are in standard errors. The implication is that, if the test fails to reject the null hypothesis, the fit of the model is improved without the additional variables in the model.

The null hypothesis being tested using the LRT is that the reduced model produces a better fit of the data than the full model. The LRT estimates two models and then compares the fit of one model to the other. This is performed by comparing the log-likelihood of both models, that is, by taking the difference between the log-likelihood of the full model and the log-likelihood of the reduced model. This difference is the LRT test statistic and has an approximate chi-squared distribution, with degrees of freedom equal to the number of coefficients being tested to be equal to zero, or the number of parameters removed from the model. If the difference is statistically significant, then the full model is believed to fit the data better.

The LRT is commonly recommended when conducting model fit testing for regression. This is indeed the approach we utilized in our SAS macro to conduct our power analysis for a logistic regression with interaction.

3.2 Theory and Calculations

In order to determine the correct way to accomplish the power calculations we had to ascertain the theory behind the calculations. We set up contingency tables to help us attain what would ultimately be used as input parameters for the SAS macro. Tables 1 and 2 illustrate these theoretical contingency tables.

Table 1. Relationship between Predictor and Outcome for Effect Modifier at Level 1

Outcome	Predictor		
	No	Yes	Total
No	a	b	a+b
Yes	c	d	c+d
Total	a+c	b+d	n

Table 2. Relationship between Predictor and Outcome for Effect Modifier at Level 2

Outcome	Predictor		
	No	Yes	Total
No	A	B	A+B
Yes	C	D	C+D
Total	A+C	B+D	N

The basic logistic regression model is shown as

$$\text{Logit} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 \quad (1)$$

We used the above tables to determine how to calculate the beta coefficients. Results of more lengthy derivations are shown below. Calculation of the intercept term is shown at (2). The intercept (β_0) is the log-odds associated with the outcome when the predictor variables are set to zero. The main adjusted effect of the predictor on the outcome (β_1) is demonstrated at (3); the main adjusted effect of the effect modifier on the outcome (β_2) is demonstrated at (4); and the effect of the interaction between the predictor and the effect modifier on the outcome (β_3) is demonstrated at (5).

$$\beta_0 = \ln(c/a) \quad (2)$$

$$\beta_1 = \ln(ad/bc) \quad (3)$$

$$\beta_2 = \ln(aC/Ac) \quad (4)$$

$$\beta_3 = \ln(AD/BC) - \ln(ad/bc) \quad (5)$$

As discussed earlier, researchers were interested in planning a study to explore effects of cancer stage at diagnosis on the relationship between active treatment and smoking cessation. They suspected that there may be an interaction between cancer stage and treatment with respect to smoking cessation. For this particular study, the β_3 coefficient in the logistic regression model is the effect of the interaction between cancer stage and treatment on smoking cessation and is the effect for which power is being computed.

Any time researchers wish to estimate power for a proposed study, they will need to know something about the proposed population they are studying before a sufficient sample size or power result can be attained. One usually accomplishes this by doing literature reviews in relation to their research topic of interest or using some pilot data results. For this situation, the theory behind determining what information will ultimately be used to compute power is derived from the proportion of individuals in the predictor of interest and the effect modifier, as well as assumed odds ratios for each strata of the effect modifier. Following that, one can fill in hypothetical cell counts or frequencies to create 2x2x2 contingency tables.

The following example is a scenario that came from the assumed population proportions from a prior research and grant proposal. Thus, the data used are based on known proportions of individuals who are going to be randomized to active treatment in a 2:1 fashion, and patients with early cancer stage diagnosis. For our scenario, treatment is the explanatory variable of interest, and cancer stage at diagnosis is the effect modifier. To obtain the numbers for the tables, we used the assumed proportion of individuals in the population who were going to be randomized to active treatment (0.66 or 66%) and the assumed proportion who had an early stage diagnosis (0.6 or 60%). If we were to use a hypothetical population of 809, we could assume some odds ratios and determine our cell sizes. Tables 3 and 4 illustrate the tables with data derived from our conjectured scenario.

Table 3. Relationship between Treatment and Smoking Abstinence for Cancer Stage = 3-4
OR ~ 1

Outcome	Treatment		
	Control	Active	Total
No	102	203	305
Yes	5	10	15
Total	107	213	320

Table 4. Relationship between Treatment and Smoking Abstinence for Cancer Stage = 1-2
OR ~ 4.77

Outcome	Treatment		
	Control	Active	Total
No	152	263	415
Yes	8	66	74
Total	160	329	489

Again, we assumed a population of 809; however, that can be scaled up or down, or even cell proportions can be used. We assumed no relationship between treatment and smoking abstinence for the late stage cancer diagnosis (OR=1), and we assumed an odds ratio of 4.77 for the association between treatment and smoking abstinence among early stage cancer diagnosis. Using these odds ratios, we were able to determine the cell counts shown in Tables 3 and 4. These are the numbers that are used to determine assumed beta coefficients for the logistic regression model in (1) above and are ultimately used for power computations.

These cell sizes are used in formulas (2) through (5) above to obtain assumed beta coefficients in the model.

$$\beta_0 = \ln(c/a) = \ln(5/102) = \ln(0.049) \quad (6)$$

$$\beta_1 = \ln(ad/bc) = \ln(1020/1015) = \ln(1.005) \quad (7)$$

$$\beta_2 = \ln(aC/Ac) = \ln(816/760) = \ln(1.074) \quad (8)$$

$$\beta_3 = \ln(AD/BC) - \ln(ad/bc) = \ln(4.768) - \ln(1.005) = \ln(4.745) \quad (9)$$

The method we developed to compute power for this study allows users to input the cell sizes from the contingency tables along with several other input parameters, and our SAS macro then performs the calculations to produce parameter estimates, such as the ones shown above in (6) through (9). These calculations of the parameter estimates and study proportions are similar calculations one would have to carry out to obtain the parameters required for Demidenko's (2008) applet (Figure 1).

Tables 3 and 4 also illustrate interaction between cancer stage and treatment assignment. One can see that the odds of experiencing a more favorable outcome depend on the cancer stage at diagnosis. A differential effect of treatment on smoking cessation exists between individuals that have different cancer staging. The objective here is to obtain a sample large enough for the logistic regression model to have sufficient power to detect an interaction effect of approximately $\ln(4.745) = 1.557$.

3.3 SAS Macro %LRPowerCorr10Int

As noted before a SAS macro has been developed that estimates power for logistic regression models with two or more predictors of interest in the presence of additional confounders (Northern et al., 2009; Williams et al., 2010). However, the existing version of the macro does not carry out a power analysis for an interaction term. Therefore, we extended this existing macro into %LRPowerCorr10Int to estimate power for a logistic regression model that includes two main predictors of interest and an interaction term. This study involved modifying the existing SAS macro to incorporate code that computes beta coefficients using inputs for the macro. These inputs (among others) are the cell counts from the tables of population proportions and assumed associations discussed above. The existing macro was modified to use the calculated beta coefficients for logistic regression models used in the power analysis. The new macro empirically calculates the power for an interaction term based on sample size and other parameters that are determined by the user.

The algorithm takes a specified number of random samples with replacement from the true underlying logit (the logit is based on the beta coefficients calculated from user-specified input cell counts). Two logistic regression models are fitted, a full model which includes the coefficients for the main effects and the interaction term, and a reduced model with the interaction term removed. An LRT statistic is calculated to test the differences between the full model and the reduced model. Each time it is found to be significant (correctly rejecting the null), it is recorded as 1, otherwise as 0. As mentioned previously, this process is repeated as many times as specified by the user; approximately 1000 simulations are usually sufficient. After the simulation process is completed, the proportion of correct rejections is tabulated. This proportion is the estimate of the power for the conjectured scenario.

The SAS macro is invoked by using the %LRPowerCorr10Int call. The input parameters are described in the paragraph below. The entire macro is available upon request from the authors (see contact information).

```
%LRPowerCorr10Int (1500,1000,0,
                    102,203,5,10,152,263,8,66,
                    1,1,1,
                    1,1,1,1,
                    cx1 cx2 cx3,
                    cx1 cx2,
                    .05,
                    1,
                    .66,.6)
```

Line one specifies the sample size that is to be evaluated ($n=1,500$), the number of simulations to be run (1000), and the correlation between the explanatory variables (0 or no correlation in this case, which is recommended). Line two specifies the cell counts from the contingency tables discussed above. The

macro calculates the beta coefficients based on these numbers and uses them to compute the power. Our macro allows the addition of more explanatory variables to be included in the model if desired. These can be other covariates of interest or confounders. They are specified in lines three and four and are entered as odds ratios. We did not include such covariates in the model, so the odds ratios are designated as one, or no association. Line five specifies the full logistic regression model to be tested. Line six specifies the reduced model. Line seven specifies the alpha or significance level. Line eight specifies the degrees of freedom associated with the likelihood ratio test statistic, designated as one because it is being tested whether or not one explanatory variable can be removed from the full model. Line nine specifies the probabilities associated with the predictor variable of interest and the effect modifier. In our example, these are the proportions of subjects randomized to treatment group (66%), and the proportion of subjects with the early stage diagnosis (60%).

4. Results

Below is the main portion of the output generated by the SAS macro.

Sample size = 1500; Simulations = 1000;		
Power	LCL	UCL
80%	77%	83%

This macro has the capability to list certain macro parameters in the output as well as functions of the parameter estimates that were computed for each term in the model. The computed beta coefficients match the ones detailed in results (6) through (9) above. Additionally, the odds ratios that are specified for any additional covariates can be listed in the output. The output can also include the description of the full and reduced models. The estimated level of power is displayed along with a computed 95% confidence interval. A power of 80% was achieved with a specified sample size of 1,500.

5. Discussion

This project expands available methods that are currently limited in estimating the power for logistic regression with interaction present. Moreover, it does it in a perhaps slightly more intuitive fashion than Demidenko's (2008) method.

Demidenko's (2008) applet will calculate power for two binary explanatory variables and their interaction. Our macro does not limit the power calculation to binary variables. In addition, our macro is not limited to two variables and the interaction term. Other predictor variables and/or covariates can be added to the model. Correlation between covariates in the model can also be specified, which is likely to be present in many observational studies; however, it should be used with caution if this correlation becomes large.

A limitation of this project is that we used a simple example to test our macro, basically a 2x2x2 scenario. We assumed 0 correlation. Overall, however, this macro has extended capabilities as discussed above and offers an additional option to researchers dealing with this or a similar scenario.

References

- Agresti A. (2007). An Introduction to Categorical Analysis. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Demidenko E. (2008). Sample size and optimal design for logistic regression with binary interaction. *Statistics in Medicine*, 27:36-46. <http://www.dartmouth.edu/~eugened/power-samplesize.php>
- Friendly M. (1998). Visualizing Categorical Data: Powerlog - Power for Logistic Regression, Quantitative Predictor. <http://www.math.yorku.ca/SCS/vcd/powerlog.html>
- Hintze J. (2013). PASS 12. NCSS, LLC. Kaysville, Utah, USA. www.ncss.com
- Hosmer DW, Lemeshow S, and Sturdivant RX. (2013). Applied Logistic Regression. 3rd ed. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Northern J, Williams DK, Bursac Z. (2009). An Algorithm and Macro for Estimating Power and Sample Size for Logistic Models with One or More Independent Variables of Interest in the Presence of Covariates. ASA Proceedings of the Joint Statistical Meetings, Biometrics Section, Alexandria, VA: 3476-3482.
- SASv9.4 Documentation. (2015). SAS Institute Inc., Cary, NC. www.sas.com
- Szklo M and Nieto FJ. (2007). Epidemiology Beyond the Basics. Jones and Bartell, Inc.
- UCLA ATS, n.d. FAQ: How are the likelihood ratio, Wald, and Lagrange multiplier (score) tests different and/or similar? UCLA: Academic Technology Services, Statistical Consulting Group. http://www.ats.ucla.edu/stat/mult_pkg/faq/general/nested_tests.htm
- Williams DK, Bursac Z, Wooten T. (2010). Three Algorithms and SAS Macros for Estimating Power and Sample Size for Logistic Models with One or More Independent Variables of Interest in the Presence of Covariates. ASA Proceedings of the Joint Statistical Meetings, Biometrics Section, Alexandria, VA: 3286-3293.

Contact Information

Zoran Bursac, PhD
 Professor
 Division of Biostatistics and Center for Population Sciences
 Department of Preventive Medicine
 University of Tennessee Health Science Center
 Tel. (901) 448-1195
 Email. zbursac@uthsc.edu