

A New Effect Modification P Value Test Demonstrated

Manoj B. Agravat, MPH, University of South Florida, SESUG 2009

Abstract:

Effect modification P value is a method to determine if there is a condition called homogeneous odds ratio which if present, interaction is possible and must be analyzed. Currently the PROC FREQ CMH SAS® command is used to test for whether the odds ratio is homogeneous for large sample sizes with fixed effects. One of the tests currently used is the Breslow -Day test in the PROC FREQ CMH test (SAS). If the P value < alpha, for the new Method, you reject the null and say the condition of homogeneous odds ratio is rejected, and that there is effect modification. The Breslow-Day test used is meant for large sample sizes, linear relationship, and fixed effects. The new Method can be used for large or small sample sizes, meant for random variables with a non- normal distribution, and includes a chi-square test of independence, as well as a test for large sampling approximation from P values of LRT, Score, and Wald tests, and the Durbin Watson statistic test for autocorrelation. The PROC LOGISTIC (SAS) Lackfit command is used for this new Method. A new procedure to transform and fit count data is demonstrated that allows you to use predicted versus expected counts and regression to obtain a P value from regression and residuals to evaluate effect modification P value. The area under the curve, ROC curves, and power is discussed in this paper to support the new method of the author.

Introduction

In this type of test of interaction, the problem of effect modification P value is meant for determining if there is a condition that is called homogeneity of odds ratio. The odds ratios have to be somewhat equal because that is the question. No one category can have an odds ratio that is much greater than the others. In fact this becomes the null hypothesis, while the alternative becomes that you fail to prove that the odds ratios are not homogeneous which is when the Breslow- Day P value and the new Method have $P > \alpha$. If the null is rejected, then a condition of effect modification exists. SAS tests for the standard condition of effect modification , are tested for the Breslow-Day test which is coded for by using the PROC FREQ (SAS) command with CMH as the option. Breslow- Day test is meant for linear data that has fixed effects that affect the inference made and is meant for large sample sizes only. The new Method you will see in this paper is using the PROC LOGISTIC (SAS) and the Lackfit command in SAS is indicated for random variables with a non-normal distribution that have independent covariates for data that can be for small or

larger sample sizes. If the data is non-normal and covariates are independent of the outcome, then, you can consider that the assumptions of random effects have been met for this new Method using PROC UNIVARIATE command and PROC FREQ Chisq (SAS), and large sample approximation tests from LRT, Score, Wald test, and PROC AUTOREG. The Power for the new Method is higher than Breslow- Day's test. The ROC curve shows higher area under the curve for the author's method than Breslow- Day test. The standard error for small sample sizes is smaller than Breslow –Day test. The P value obtained produces a lower value however the algorithm converges so the results for MLE are valid. There are other points such as maxRsqr being good, C statistic being higher, and convergence of algorithm when testing for power (equaling 1) as well as P value that support s the benefits and use of this new Method that is shown in SAS outputs.

Fixed vs. Random Effects:

As stated Breslow –Day's test is a fixed effect test, while the new Method is meant for random effects. To begin with, the importance of understanding the meaning and significance of fixed versus random effects is discussed. Fixed effects are considered non-experimenting controlling for other variables using linear and logistic regression. One has to include the variables to estimate the effects of the variables. Next, the variable has to be measured and chosen for model selection. Ordinarily, if dependent variables are quantitative, then fixed effects can be implemented through ordinary least squares. As a result, stable characteristics are able to be controlled eliminating bias. Fixed effects ignore the between-person variation and focuses on the within-person variation. In a linear regression model:

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \alpha_i + \epsilon_{ij}$$

β_0 is fixed and all x 's are measured while β_1 is fixed. ϵ or the error term is defined as a random variable with a probability distribution that is normal and mean 0 and variance σ^2 . Thus the models can be both fixed and random. Fixed effects are also considered as pertaining to treatment parameters where there is only one variable of interest. They are used to generalize results for within effects. In a random effects model, the ϵ can exist and do not have to be zero and be taken into account. Random effects exist when the variable is drawn from a probability distribution. Blocking, controls, and repeated measures belong to random effects. Random effects are involved in determining possible effects and confidence intervals. Unbalanced data may cause problems in inference about treatments. Random effects are involved in clinical trials and in making causal inferences. Random effects do not measure variables that are stable and unmeasured characteristics. Thus the α 's are uncorrelated with the measured

parameters. Random effects model can be used to generalize effects for all of the variables that belongs in the same population.

New Effect Modification P value test

The author calculates a new effect modification P value from adjusted data. The zx , yz , z beta estimates are calculated from survival analysis. Zx represents confounder interaction with explanatory variable. Yz represents interaction with outcome and confounder. Z represents the confounder. The model of the SAS (TM) code for proc logistic class level variables is chosen sometimes dependent on whether the outcome converges with all the variables in. In other words, one variable may be the same as another and can be left out. The format is same for each level, the outcome comes first with $y=1$ positive for lung cancer then follows 1 for a 'fit' variable next. This row with 1 means not fitted values. Next there is a 1 for the confounder and 1 for the explanatory variables. Then, the count or n is from the data set directly. For the next row, the outcome will be same $y=1$ and 1's for confounder and explanatory variable followed by the raw count. The next two rows will have the fit values, so both fit variables will be 0 in each row, and the fit values will come from the sequence shown: for the 'z', or the confounding variable, the adjusted for count is the: $\text{observed count}/zx*z$. For the explanatory variable, it is $\text{observed count}/yz$. The value of count comes from the observed count, and this method is used to calculate a new count. You must alternate it until the even data is finished. One uses the Lackfit command in SAS (TM) software and obtains the P value. If the P value is greater than the alpha, then we fail to reject the null and say there is no effect modification.

Table 1. Lung Cancer Study Data (Agresti, 1996)

Country	Spouse Smoked	Cases	Controls
Japan	No	21	82
	Yes	73	188
Great Britain	No	5	16
	Yes	19	38
United States	No	71	249
	Yes	137	363

Diagram 1. Sample Size versus Breslow-Day and New Method for P value

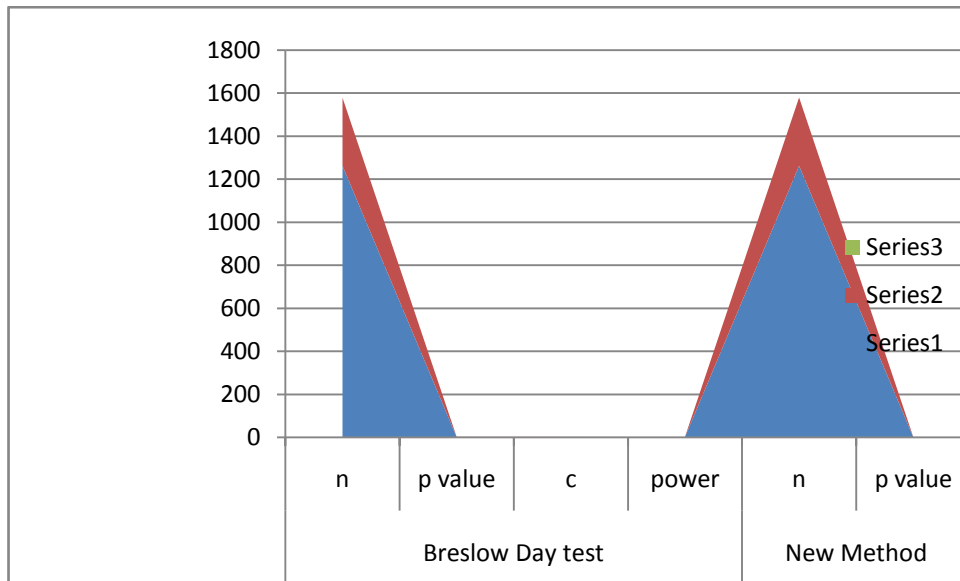


Diagram 2. Sample Size Versus Average Power

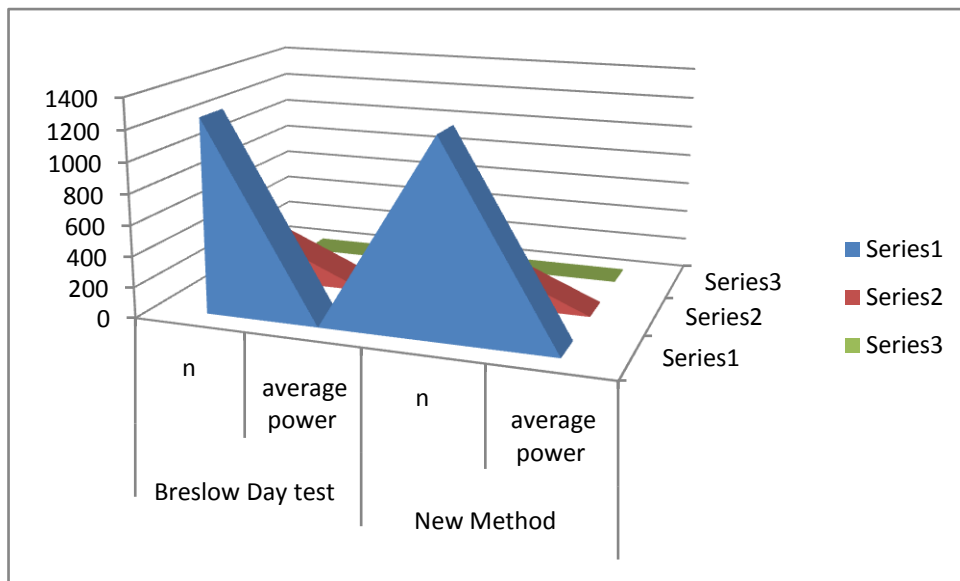
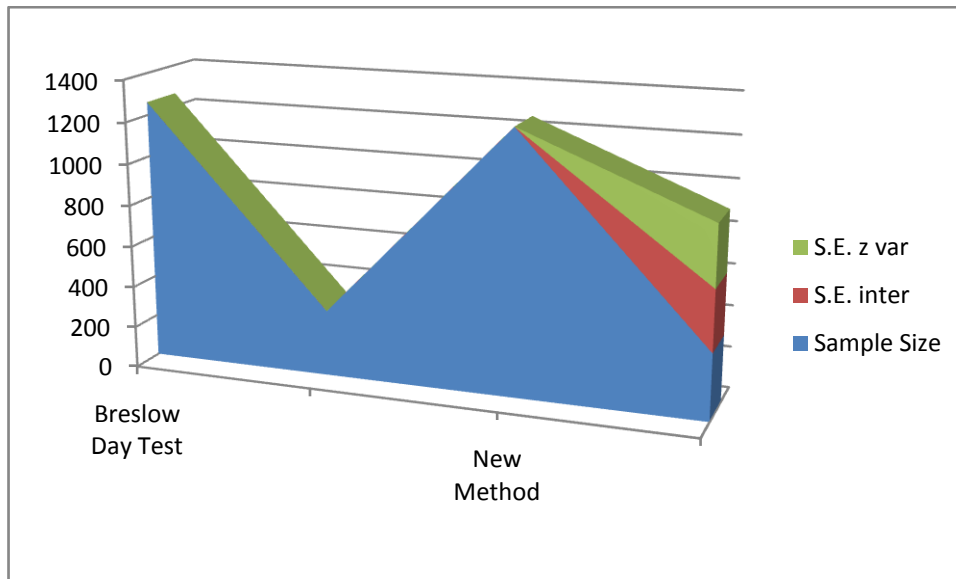


Diagram 3. Sample Size versus Standard Error



Here diagram 3, shows that the distribution is symmetric for sample size but not for standard error (S.E.) though S.E. is accounted for more in the new Method versus Breslow-Day.

The standard error distribution of both the new and old test has more variation for the new Method possibly implying that there is more variation in standard error intercept while controlling for unmeasured covariates and adjusting for lack of independence. The variability in standard error of the β_2 intercept is accounted for far less in the standard test than the new Method. The Breslow-Day test shows less standard error for large sample size for the confounder than the new Method (.07 vs. .38) but has less power (.5 vs. 1). The new Method has less standard error for the small sample size (.04 vs. .52) and greater power (1 vs. .77 for confounder).

Code 1. New Method for Effect Modification P Value:

```

Data passlungexp5d;
input cases fit zxy xzy n;
datalines;
1 1 1 1 73
0 1 1 1 188
1 0 1332 101 27
0 0 5201 307 82
1 1 1 1 19
0 1 1 1 38
1 0 317 19 5
0 0 1015 60 16

```

```

1 1 1 1 137
0 1 1 1 363
1 0 4503 266 71
0 0 15793 933 249
;
run;
proc logistic data=passlungexp5d descending;
weight count;
class xzy ;
model cases= fit xzy / rsq lackfit;
run;

```

Output 1. Large Sample Size Test of Effect Modification for New Method

Testing Global Null Hypothesis: BETA:				
	Test	Chi-Square	DF	Pr > ChiSq
	Likelihood Ratio	229.5875	2	<.0001
	Score	173.7039	2	<.0001
	Wald	66.2636	2	<.0001
R-Square	1.0000	Max-rescaled R-Square	1.0000	
The LOGISTIC Procedure				
	Pairs	36	c	0.653
Hosmer and Lemeshow Goodness-of-Fit Test				
	Chi-Square	DF	Pr > ChiSq	
	8.6778	5	0.1226	

The global null hypothesis of $\beta=0$ shows that LRT, Score, and Wald $P < .0001$ which means that large sampling approximation are working well and the results are trustworthy for $\beta=0$. The Max-r-Square=1 which is 100%. This shows 100% correlation of predicted outcome.

Code 2. Testing for Non-Normal Distribution

```

proc univariate data=passlungexp5d normal;
var cases;
qqplot cases / normal(mu=est sigma=est);
run;

```

Output 2. Non-Normal Distribution test

The UNIVARIATE Procedure		Variable: cases	
Tests for Normality			
Test	--Statistic---	-----p Value-----	
Shapiro-Wilk	W 0.649783	Pr < W	0.0003
Kolmogorov-Smirnov	D 0.330824	Pr > D	<0.0100
Cramer-von Mises	W-Sq 0.32839	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq 1.996967	Pr > A-Sq	<0.0050

As you can see, the Shapiro Wilk test shows $P < \alpha$, with $P < .0003$ hence you can conclude that this is a non-normal distribution you are dealing with by rejecting the H_0 : of existence of normality.

Independence of large Sample Size:

Code 3. Independence

```
proc freq data=passlungexp5d;
weight n;
tables cases*(zxy xzy) / expected chisq;
run;
```

The chi square independence shows that for both variables you fail to reject independence from the outcome cases of lung cancer with p values $< .0001$.

Output 3. Power Analysis of New Method

power of passlung5 count data							
Obs	Source	DF	ChiSq	Prob ChiSq	test	power	
1	fit	1	152.60	<.0001	3.84146	1	
2	zxy	1	85.479	<.0001	3.84146	1	
3	xzy	1	83.235	<.0001	3.84146	1	

Figure 1 ROC Curve for Author's Effect Modification P value Large Sample Size

Code 4. Autocorrelation and Randomness

```
proc autoreg data = passlungexp5d;
model cases = fit zxy/ dw = 2 dwprob;
run;
quit;
```

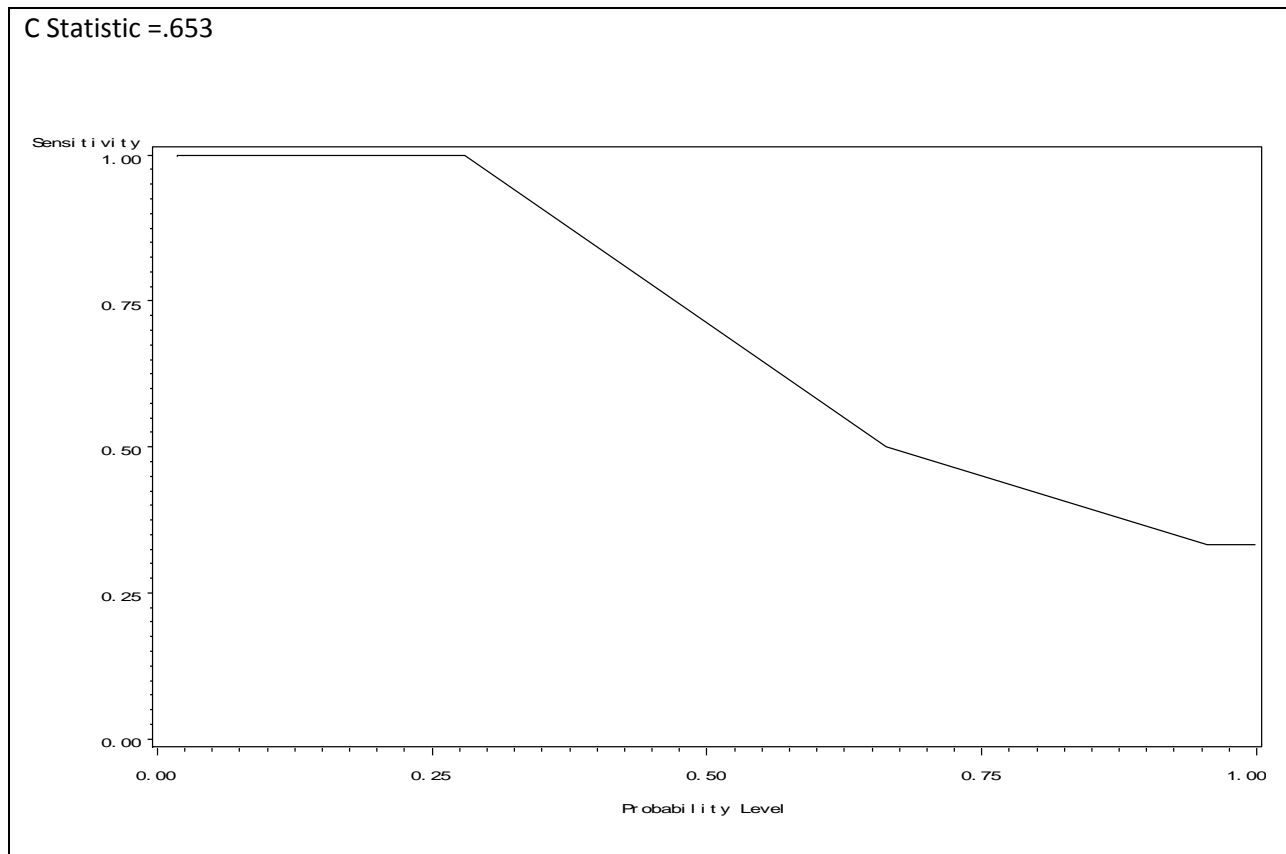
Output 4. Autocorrelation and Randomness

Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	3.6072	0.9994	0.0006
2	0.1808	<.0001	1.0000

NOTE: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

Interpretation of order 1 test of Durbin Watson statistic shows that there is no autocorrelation hence one may have randomness. Plus DW statistic is greater than 2 indicating no autocorrelation.

Figure 1. ROC Curve for Author's Effect Modification P value Large Sample Size



In this study, the probability of lung cancer as outcome is cases=1 (positive for lung cancer), and 0 for no lung cancer. 'Country' represents the confounding variable or the countries: 1=Japan; 2=Great Britain; 3=United States. 'Smokers' represents passive smoke exposure or the spouse who smoked with x=0 or no passive smoke exposure; x=1 as positive smoke exposure.

Code 5. Traditional Test for Effect Modification P Value: Breslow-Day Test for Large Dataset

```

data passlung5;
input country smokers cases count;
datalines;
1 0 1 21
1 0 0 82
1 1 1 73
1 1 0 188
2 0 1 5
2 0 0 16
2 1 1 19
2 1 0 38
3 0 1 71
3 0 0 249
3 1 1 137
3 1 0 363
;
run;
proc freq data=passlung5 order=data;
weight count;
tables country*smokers*cases/cmh chisq;
run;

```

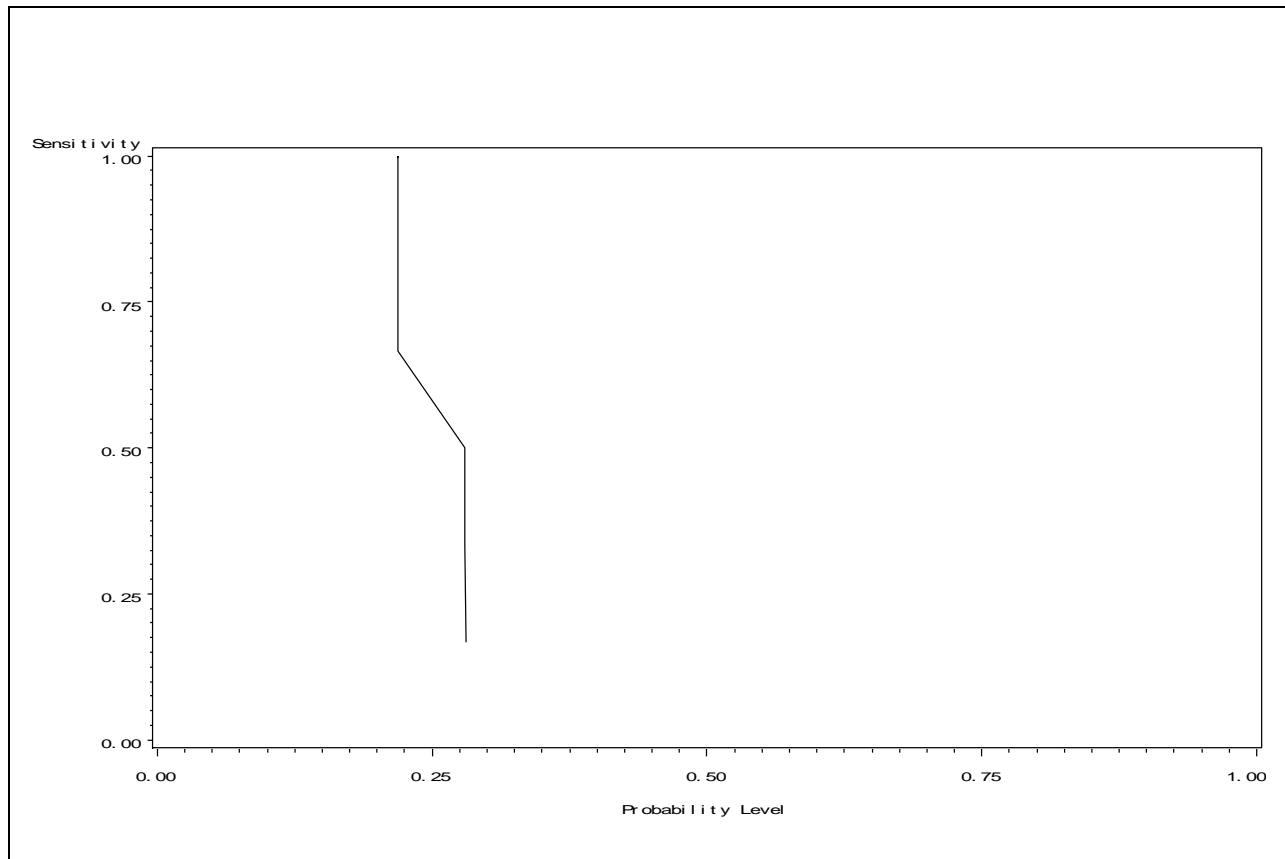
Output 5. Breslow-Day Test for Large Dataset:

Controlling for country					
Cochran-Mantel-Haenszel Statistics (Based on Table Scores)					
	Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	5.4497	0.0196	
Breslow-Day test					
Homogeneity of the Odds Ratios					
	Chi-Square	0.2381			
	DF	2			
	Pr < Chisq	<.8878			

Total Sample Size =1262

You can see that the sample size is 1262. The Nonzero correlation P value $< .0196$ hence there is independence. Next, the Breslow -Day P value is $.88 > .05$ (alpha), hence you conclude that you fail to reject that homogeneity of odds ratio exists.

Figure 2. ROC for Breslow–Day Version for Large Dataset



Output 6. Power Analysis on Breslow-Day Test for Large Dataset:

Power of Passlung5 count Data							
Obs	Prob Source	DF	ChiSq	ChiSq	test	power	
1	country	1	0.0005	0.9819	3.84146	0.05006	
2	smokers	1	5.7062	0.0169	3.84146	0.66598	

The new Method procedure applied to a small sample size (318 or 317) giving the results summarized in the table 1 below.

```

data hrpdeath;
input defendantr victimr penalty count;
datalines;
0 0 1 6
0 0 0 97
0 1 1 11
0 1 0 52
1 0 1 0
1 0 0 9
1 1 1 19
1 1 0 132
;

```

Table 2. Comparing Two Sets of Data and Two Tests

	Sample Size	P value	C statistic	Average power	S.E. Intercept	S.E. (Z var.)
Breslow- Day test	1262	.88	.5	.5	.21	.07
	317	.52	.62	.19, .77	.42	.52
New Method	1262	.12	.653	1	.38	.38
	318	.86	.813	1	.65	.04

Table 2 given above shows the sample size , P value, standard error, and C statistic for the two tests and a small data set and a larger dataset. The Breslow- Day test for larger sample size has a lower C statistic yet higher P value. The new Method has lower P value but higher C statistic and power, hence you want to believe the new Method. The next one shows that Breslow Day test is symmetric for P value, standard error of 'z', and C statistic, while the new Method shows a sharp slope in daigram 3. Ideally one may expect the symmetric relationship between power and sample size as seen with the new Method for effect modification as seen. One can clearly see that the power is very good for the new Method versus not so good for Breslow –Day.

**Conclusion: New Effect Modification P Value Test Has Advantages
The New Method:**

The effect modification P value using my new Method is <.12. This P value > α =.05 hence one concludes that there is no effect modification and that the null of homogeneous odds ratios is not rejected. Since this method is indicated for random variables with non-normal data that have independent covariates, the inference will deal with differences in levels into design by chance not design that is found in variables of the same type of variable in the

same population. The relative difference among levels of interest chosen while the author's method tells of the variable not typically chosen for their unique personal attributes, while the new Method collectively represents random variables with non-normal distributions that are independent. From the data, the PROC UNIVARIATE (SAS TM) test show that $P < .0003$ hence there is evidence to reject normality. The chi square independence test of PROC FREQ (SAS) shows that $P < .0001$ which support independence between outcome of cases and the variable zxy. The PROC LOGISTIC (SAS TM) test shows that the likelihood ratio test, Score test, and the Wald test shows that $P < .0001$ which support s the validity of results from large sampling approximations. The Durbin Watson statistic is 3.6 indicating there is no autocorrelation hence randomness is expected. One can generalize from the new Method as result that for outcome lung cancer there is no effect modification for countries across all three levels. The effect of lung cancer may not vary for all countries in the same population.

Comparison of Two Tests:

For the standard test, the nonzero correlation P value is $< .0196$ hence the variables are independent, thus you continue to Breslow- Day test (PROC FREQ CMH, SAS TM). This conclusion is that the Breslow –Day's P value of .88 which is meant for large sample size and fixed effects is much larger than $P < .12$ of the new Method, and the power is higher for the new Method, hence Breslow- Day may have an overestimate with large sample size. The Breslow- Day test deals with independent variables because they are directly related to the topic of interest. One may conclude that the conditional odds ratio of smokers for lung cancer has independent variables that do not change with different levels of country. The C statistic for the Breslow-Day test is .5 using PROC LOGISTIC for the large dataset. For the author's technique, the new tool has C statistic of .65 for this data for large sample size dataset used, and the new Method produces a C statistic of .875. The author's technique has greater area under the curve (or C statistic) for the effect modification P value therefore there may be greater confidence in the author's method versus the Breslow-Day test for these datasets. The Breslow–Day Test is meant for large sample size and fixed effects but in this case their C statistic is only .5 for this data which is minimal, and their power is minimal for the coefficients. The Breslow- Day test is not meant for either too large or too small sample sizes because the power needs to be high enough for there to be validity. Quite often count data studies' datasets have smaller sample size like this study on lung cancer. Perhaps as a consequence there is less confidence as well as less accuracy with the Breslow-Day test for this dataset. The new Method is showing less standard error for the intercept of confounder than Breslow-Day test which indicates that the new test is better, as well as for having higher power and higher area under the curve. The usefulness of the new Method is good because smaller sample size studies can be done when checking for interaction which makes the new method more useful cost wise, plus the power of the methods are good for both

large to small sample sizes. Inferences regarding between subject variations can be made and so there is no statistically significant difference in odds ratio across strata for countries and lung cancer due to passive smoke exposure for the large group of countries mentioned (Japan, Great Britain, and United States). Randomization in this type of effect means that bias is adjusted for and taken into account which makes this method useful for randomized clinical trials.

Comparison of Small Data set for Two Methods

For a small dataset, the Breslow -Day test shows a P value of .52 for the small data set indicating that you fail to reject the null of lack of homogeneity of odds ratio and power of .76 for victim's race and .19 for defendant's race. You may not expect interaction between death penalty and defendant's race or that they may not differ significantly. The power for the Breslow- Day Test is very much less than the author's method. This disparity in power may indicate that perhaps the author's method should be used since there is balanced and greater overall power. Perhaps the new Method has greater power (1) so the P value estimate (.86) is a more liberal and a closer estimate. The new Method has higher c statistic (.813) and lower standard error (.04 for β_2). Therefore if one correctly identifies the status of death penalty, and optimize the chance of not making error saying there is no effect modification or there is no change in odds ratio across the status of defendants' race based on non-significant P value. The new Method also has a maxRsqr of 1 with a converging algorithm for P value for this dataset, and a power of 1. Higher power is necessary for validity of the study.

SAS (TM) as a tool and a means for analysis is very helpful for analysis of this problem of Effect Modification P value when the previous method is meant for large datasets and only fixed effects. The variability of this new Method will allow SAS (TM) to produce more meaningful results with higher power and more area under the curve for large and small sample sizes with non-normal distribution, independence, and random effects.

SAS (TM) and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

References

1. Agresti A, 'Introduction to *Categorical Data Analysis*'. NY: John Wiley and Sons; 1996.
2. Blot W, Fraumeni J, 'Passive smoking and lung cancer', *Journal of National Cancer Institute*. 1986;77 No.5:993-1000.
3. SAS (TM) institute, Cary, North Carolina.

4. Wackerly D, Mendenhall, Schaeffer , 'Mathematical Statistics with Applications'; US, Duxbury Thomson Learning; 2002.
5. 'Logistic Regression Model chapter 1'. (www.cda.morris.umn.edu/~anderson/math4601/notes/logistic.pdf). (Retrieved 10/07/07).
6. 'Logistic Regression Model chapter 8'.
(<http://cda.morris.umn.edu/~anderson/math3611/notes/logistic.pdf>). (Retrieved 10/07/07).
7. 'New Method for Calculating Odds Ratios and Relative Risks and Confidence Intervals after Controlling for Confounder with Three Levels' by Manoj Agravat, MPH (Unpublished, 2009).
8. Mandy Webb, Jeffrey Wilson, Jenny Chong; 'An Analysis of Quasi-complete Binary Data with Logistic Models: Applications to Alcohol Abuse Data'; *Journal of Data Science*; 2:273-285, (2004).
9. Theodore Holford, Peter Van Hess, Joel Dubin, Program on Aging, Yale University School of Medicine, New Haven ,CT., 'Power Simulation for Categorical Data Using the RANTBL Function', SUGI30 Statistics and Data Analysis,pg, 207-230.
10. 'Applied Logistic Regression, Second Edition by Hosmer and Lemeshow Chapter 5: Assessing the fit of the model', (<http://www.ats.ucla.edu/stat/SAS/examples/alr2/hlch5sas.htm>), (Retrieved,10/07).
11. 'Introduction to Fixed Effects';
<http://support.sas.com/publishing/pubcat/chaps/58343.pdf> (Retrieved, 6/10/09).
12. 'Random Effects'
http://faculty.ucr.edu/~hanneman/linear_models/c4.html (Retrieved, 6/11/09).
13. 'Lesson 6: Logistic Regression-Binary Logistic Regression for Two Way Tables'
http://www.stat.psu.edu/online/courses/stat504/06_logreg/11_logreg_fitmodel.htm (Retrieved, 6/20/09).
14. Nathan A. Curtis, SAS Institute Inc., Cary, NC ', 'Are Histograms Giving you Fits? New SAS Software for Analyzing Distributions.'
<http://support.sas.com/md/app/papers/distributionanalysis.pdf> (Retrieved, 7/11/09).
15. 'Regression Analysis by Example by Chatterjee, Hadi and Price Chapter 8: The Problem of Correlated Errors'
<http://www.ats.ucla.edu/stat/sas/examples/chp/default.htm> (Retrieved, 7/11/09).

Acknowledgements:

I want to thank my family, wife, son, parents, and brothers who always gave me support for my endeavors. I was inspired by my poem 'Epilogue.'

Manoj B Agravat MPH
University of South Florida
20108 Bluff Oak Blvd
Tampa Florida 33647
813-991-5537
M_agravat@yahoo.com