A New Method of Odds Ratio and Hazard Analysis of Head and Neck Cancer

Manoj B. Agravat*

University of South Florida – COPH, Biostatistics/Epidemiology, 20108 Bluff Oak Blvd Tampa, Florida 33647, USA

Abstract: The main topic of this paper is to focus on a new method for calculating odds ratios and hazard ratios through probabilities and effect modification. This probability is derived through an odds ratio proof for the common conditional odds ratio of Cochran Mantel Hansel showing theta equals one. Subsequently, the probability formula is obtained and the hazard ratio expression derived. However, the new relation of this proof is to show that logits equals itself through probability. From this derivation, an expression of risk is obtained which is an odds ratio. Parameters are obtained through a novel method of Survreg and its proportional hazard assumption. The odds ratio obtained is given as per strata as well as hazard ratio method demonstrated which is curvilinear to probability in comparison for the interaction model to represent percent change. The odds ratios from PROC GLIMMIX for interaction model has odds ratio of 1.76 vs 1.73 and 1.83 vs 1.76 for white and black males of a logit expression another expression of a logit. A parametric analysis shows correlation to the odds ratios for strata and probability Pr(z) that can work from a new derivation for an odds ratio with for the exposure shown to have power with the RANTBL function of about 83 % with effect modification included at 100% power. The comparison of effect modification P values to hazard ratio is then made for differences across strata.

Keywords: Logit, Odds ratio, Hazard ratio, Non-normal probability, Effect modification, Head neck cancer.

1. BACKGROUND

The data set chosen is from an INHANCE study from Mia Hasibe *et al*. 2007 [4] from the Journal of National Cancer Institute. There are many sites where this survey is taken including Europe, United States, and Asia. The main outcomes are types of head neck cancer oncology such as squamous cancer, oropharyngeal carcinoma, palatine cancer, pharynx, and larynx as well. The exposure of cigarette smoking and alcohol is important due to cases being 75 % associated with these risks. Non-drinking is at least 100-150 ml glasses of alcohol drinks in the month of beer, vine, and hard liquor. Non-smokers are not users of tobacco as well as cigarette. The people in this study may not be mutually exclusive to avoid bias. 15 case studies were pooled for the study. 5-34 % of head neck cancer are due to genetic heredity perhaps the TP 53 gene and not smoking. Some harmful mutations were known to be on the 3q and 4p site on the 11q13 chromosome. Bias or confounding is the reason the non-drinkers are pooled with the non-smokers as controls with ICD 2 classification. Smoking was limited to more than 100 cigarettes in lifetime in US and Central Europe. Taking snuff and chewing tobacco since they are common in Central Europe is not included in this study.

*Address correspondence to this author at the University of South Florida – COPH, Biostatistics/Epidemiology, 20108 Bluff Oak Blvd Tampa, Florida 33647, USA; Tel: 813-846-0903; Fax: 813-991-5537;
E-mail: Mbagravat00@gmail.com

Variables included for interaction [1-3] are B_x, B_z, B_{yz},$B_{zx,}$ the variable race (B_z), non-drinkers (B_x), and cases with exposure interaction (B_{yx}). These are some of the combination possible for the analysis. Race includes white, black and Hispanic. The hazard ratio HR from HR(z) matches the normal exponential of the hazard ratio estimates. Both SAS 9.3 (including E Miner) and R studio software were utilized for this analysis.

Head neck cancer involves regions of the oral mucosa, larynx, pharynx, and paranasal sinuses, plus sometimes the salivary glands. If it was the oral mucosa it will include 2/3 of the tongue, gums, and the inside cheeks. In some cases, it may affect the vocal cord nodules, papillomas, and squamous cell carcinoma is quite common. Benign tumors are considered polyps in the vocal cords. This type of tumors often occurs in smokers and singers due to chronic abuse of smoking. Laryngeal papilloma is a raspberry like excrescence about 1 cm in size which is often associated with hemoptysis. Papillomas are single in adults but multiple in children. HPV are sometimes non malignant but sometimes are malignant. Epstein Barr virus (EBV) is a factor which is associated with nasopharyngeal carcinoma. 1) squamous cell carcinoma; 2) non-keratinized carcinoma and 3) and undifferentiated carcinoma. Undifferentiated carcinoma is the one closest to EBV. EBV [2] affects B lymphocytes and a reaction of T lymphocytes of people with enlarged lymph nodes. Nasopharyngeal carcinomas spread to cervical lymph nodes. Carcinoma of the larynx is 7:1 more common in

E-ISSN: 1929-6029/18

males than females. It occurs 2 % of the time and after 40 years of age. Carcinoma of the larynx can be affected by smoking, alcohol, and asbestos exposure and represents squamous cell carcinoma normally and are like grey wrinkled plaques and can sometimes be bizarre. 60 percent of these cancers, laryngeal, restrict to the larynx. Some people go through treatment with surgery and, radiation therapy or take drugs with radiotherapy.

Interaction can be evaluated through the author's method which involves data transformation and creation of a P value statistic. The author calculates a new effect modification P value statistic from transformed count data. The beta estimates are calculated from survival analysis representing interaction with explanatory variable, depicting interaction with outcome and effect modifier, (Z is a variable for the effect modifier). The model of the (TM) SAS code for proc logistic class level variables is chosen sometimes or PROC MIXED dependent on whether the outcome converges with all the variables in the model. 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 cases then follows with '1' for a fit variable. This row's '1' means not fitted values. Next, there is a 1 for the effect modifier and 1 for the explanatory variable Agravat (2008, 2009, 2011, 2012). Then, the count or n is from the data set directly. For the next row, the outcome will be same (y = 1) and ones for modifier and explanatory variable followed by the raw count. The next two rows will have the fit values, therefore, both fit variables will be 0 in each row, and the fit values will come from the sequence shown: for the z, or the effect modification variable, the adjusted for count is the: observed (count) = |zx*z| as designed by the author. For the explanatory variable, the new count estimate is: observed (count) =|yz |. The value of count comes from the observed count, and this method is used to calculate a new count. One must alternate it until the symmetric count dataset is created and you must use the absolute value of the beta estimates to adjust the count data. If there is a 0 in the count make it adjusted 1 in the count or 'n' data column. The beta estimates are obtained from using the original count data unadjusted. If the P value is greater than the alpha, then we fail to reject the null and say there is no effect modification. The P values are used to choose which distribution is better for beta; hence, this step is parametric. Interaction terms are used to measure beta for instance B_{zx} for the slope of vector of B_{zx} the interaction between B_z and B_x .

Shapiro Wilks P value is non-normal for the INHANCE head and neck cancer data set with P <0.0003. The effect modification issue involves if there is a 10 % risk difference across strata. The exposure is supposed to be independent of the outcome. This will be evaluated by a P value. A PROC IML and PROC Mixed [14] algorithm can be utilized for this analysis with SAS software for an asymptotic chi square statistic and a F statistic with P values.

To start this procedure the interaction analysis, 2x2 by 2x3 matrices are multiplied as shown in" Formulas Calculating Risk Estimates and Testing for Effect Modification and Confounding" [1]. The means are also calculated in the same way for tables of observed and mean values. Next using the formulas of the O statistic (asymptotic chi-square [2]), calculate the output, through the PROC IML code, calculate the" AEM" variable for the SAS algorithm intended for evaluating confounding with PROC MIXED for effect modification of the head neck cancer data of INHANCE data. The program and algorithm for PROC IML (SAS) is from the author [1,2], and if" AEM" is significant one may conclude that the null of homogeneous null is rejected concluding effect modification exists. The matrix formulas are shown here in the PROC IML code as well as the O statistics. In the" New Effect Modification P Value Test Demonstrated" [15], the cases variable is used in 1, 0, 1, 0 sequence likewise for cases or outcome in this study. This algorithm for effect modification has" fit" set to 1, 1, 0, and 0. In the effect modification algorithm, the technique using O statistics and matrices utilize the observed products from matrix multiplications and mean matrices and the same method of count data transformation [1]. (The procedure for effect modification using PROC IML and PROC MIXED and O statistics is from Agravat (2011 and 2012) [2] Figure **1** (2012) and section 2.3 for PROC IML.

$$\hat{O}_{EM} = \frac{(O - \bar{O})^2}{O}$$

The odds ratio is 3.45 for cigarette smokers [4] and head neck cancer for Europe and South America. India had an odds ratio of 1.17 vs 1.20 for North America. Passive smoking exposure was partly recorded. Pharyngeal cancer risks were from 1-2 drinks per day. There was no control for HPV and head neck cancer.

A case control study, from D 'Szousa *et al* 2007, was shown as having analysis [7] for cancer of the Oropharynx and attributed to HPV as well as smoking

cigarettes and alcohol [4]. The odds ratio, including 3.54 for nonsmokers, is an HPV study. The odds ratio further increased to 5.16 for 1-19 pack years and 5.20 odds ratio for 20 plus pack years. E6/E7 serology was a factor and had an odds ratio of possibly 1.68 for positive serology and 2.75 for negative serology for oropharyngeal cancer based on a probability algorithm. The hazard ratio and probability Pr(z) for negative serology was 0.36 and 3.31 and 0.62 and 1.90. Hence the hazard ratio was greater for positive serology but due to baseline hazard. However, baseline hazard vs probability and survival changed to less risk.

The proof of Cochran Mantel Haenszel Test Θ_{cmh}=1 can be stated by this expression derived.

$$\frac{1}{odds(y)+P(z)-odds(y)*P(z)}=\frac{odds(z)}{odds(y)}$$

$$(odds(z)(odds(y)+P(z)*odds(z)-odds(y)*(Pz)*odds(z))=odds(y)$$

$$odds(y)-odds(y)*odds(z)+odds(z)*odds(y)*P(z)=P(z)*odds(z)$$

$$odds(y)=odds(y)*odds(z)+P(z)*odds(z)-odds(z)*odds(y)*P(z)$$

$$odds(y)-odds(y)*odds(z)+odds(z)*odds(y)*P(z)=P(z)*odds(z)$$

$$odds(y)-odds(y)*odds(z)=0$$

$$odds(z)=1>>>>>z=1>>when....z=1$$

$$odds(y)-odds(y)*odds(z)+odds(z)*odds(y)*P(z)=P(z)*odds(z)$$

$$>>y=1$$

Thus z/y=1/1and this equals 1 the Cochran Mantel Haenszel assumption [6] that the common conditional odds ratio of Independence equals 1 when β_z=0.

New Explanations of probability for linear regression can be obtained from inverse equations proved here.

$$\frac{odds(z)}{odds(y)}=\frac{1}{odds(y)+P(z)-odds(y)*P(z)}=$$

$$odds(z)[odds(y)+P(z)-odds(y)*P(z)]=1$$

$$=[odds(y)+P(z)-odds(y)*P(z)]=\frac{1}{odds(z)}$$

$$P(z)(1-odds(y))=\frac{1}{odds(z)}-odds(y)=$$

$$\hat{P}(z)_{new}=\frac{\frac{1}{odds(z)}-odds(y)}{1-odds(y)}$$

$$P(z)_{new}=\frac{1-\hat{odds}(y)}{1-\hat{odds}(y)}=1;if\rightarrow\beta_z=0;\beta_y=0$$

$$\rightarrow undefined(In-Linear-regression)$$

New equation for confounder and odds $(z)_{new}$ and $P(z)_{new,}$ $Hr(z)_{prob.}$ Subsequently, the hazard ratio expression is derived for calculations (HR(z)).

$$_{new}=\frac{\frac{1}{odds(z)}-odds(y)}{1-odds(y)}=P(z)_{new}*(1-odds(y))+odds(y)=$$

$$\frac{1}{odds(z)}=P(z)new*(1-odds(y))+odds(y);$$

or

$$Hr(z)_{prob}=\frac{1}{P(z)new*(1-\hat{odds}(y))+\hat{odds}(y)}$$

Proof of a logit [7] from a probability algorithm Pr(z) is demonstrated. The hazard ratio of z comes from the odds ratio of z derived in a new manner. With respect to logits however the odds ratios can be found. Probability of z and hazard ratio are then utilized to show the logits relationship to probability whose relationship is the inverse of the probability [2] normally. utilized; hence $Probability=\frac{1}{\Pr(z)}$ and $\Pr(z)=\frac{\frac{1}{z}-y}{1-y}$ while $HR(z)=\frac{1}{\Pr(z)(1-y)+y}$

2. METHODS

(A)

$$HR(z)=\frac{1}{\Pr(z)*(1-y)+y}$$

$$\frac{1}{OR(z)}=\frac{1}{\Pr(z)*(1-y)+y}$$

$$\Pr(z)=\frac{OR(z)}{1-y}-\frac{y}{1-y}$$

$$\frac{y}{1-y}=\frac{OR(z)}{1-y}-\Pr(z)$$

$$\frac{y}{1-y}=\frac{OR(z)}{1-y}-\frac{OR(z)}{1-y}+\frac{y}{1-y}$$

$$\frac{y}{1-y}=\frac{y}{1-y}$$

Then the new hazard ratio can be obtained with proof of the exponential of the beta coefficient. Here one may assume the hazard ratio is the odds ratio due to the logit as an approximate because it started with an inverse odds ratio expression. The probability of z Pr(z) is shown above and to be correlated to the logit expression of y/1-y. First the hazard ratio [8] expression is shown then the logit is tested for odds ratios in this analysis since. The odds ratio expression

is compared and derived for further study [3, 8] from the inverse for confounded and compared to z/y for proving the common conditional odds ratio of Independence.

$$OR = y + \Pr(z) * (1 - y)$$

A sample calculation of odds ratio and hazard ratio is next.

Exp(-1.421) + 1/exp(-1.531)-exp(-1.421)/(1-exp(-1.421) *(1-exp(1-.421))=

0.2414 +4.3813/ 1-.2414=

OR =0.2414 +5.7756 =6.016

1/6.016 =HR z= 0.166

The odds ratio of the original data can come from the 2 x 2 of Cochran Mantel Hansel. However, there is little difference or variation between races.

(B) In addition, there is a new solution as well for probability and logits. This may lead to new solutions.

$$\exp^{\frac{a}{1-a}}$$

$$\frac{\frac{y}{1-y}}{1-\frac{y}{1-y}} = \frac{\frac{y}{1-y}}{\frac{1-2y}{1-y}} = \frac{y}{1-2y}$$

(C) Hazard Ratio and Baseline Hazard Function

The baseline hazard can be rewritten to show its form as an exponential of its beta estimate from probability of z non-normal Pr (z) with I for $HR(i * z) = i * \exp^{z}$. This expression results in a different hazard ratio then normally done by exponentiation but right from HR(z) and Pr(z).

$$HR(z) = \frac{1}{\frac{1}{\frac{z}{1-y}} - y}(1-y) + y$$

$$HR(z) = \frac{1}{\frac{1}{z} - y + y}$$

$$HR(z) = \frac{1}{\frac{1}{z}} = z$$

$$HR(z) = \exp^{z}$$

$$HR(i * z) = i * \exp^{z}$$

The original proof of HR(z) and assumes HR(z) to be OR(z) [6]. With 1/Probability as the terms in the hazard ratio, the new relation of baseline hazard is derived with the quotient rule. 1/Probability is the Pr(z). Previously, the P/1+p term was incorporated. Later, the different baseline hazard with non-normal probability is produced where 1+P/P ~ Prz. The i variable is added to function as variable for strata as shown. In addition, the cumulative distribution's death rate is $F(z) = 1 - S(z)$.

The baseline's hazard function [2] is derived here. A similar relationship is utilized previously [2] where normal probability is there and not the inverse of the normal probability which is equal to Prz.

$$HR(z) = \frac{1}{\frac{1+z}{z}(1-y) + y}$$

$$\frac{\partial HR(z)}{\partial z} = \frac{z}{1 + z(1-y) + yz}$$

$$\frac{\partial HR(z)}{\partial z} = \frac{z}{1 - y + z - yz + yz}$$

$$\frac{\partial HR(z)}{\partial z} = \frac{z}{1 - y + z}$$

$$\frac{\partial HR(z)}{\partial z} = \frac{1(1-y+z) - (1-y) * z}{(1-y+z)^2}$$

$$\frac{\partial HR(z)}{\partial z} = \frac{1 - y + y * i * z}{(1-y+z)^2}$$

$$HR`(z) \approx \frac{1 - y + y * i * z}{(1-y+i*z)^2}$$

Novel Method in R

The novel Survreg method is here. The survreg [1-3,15] is rewritten for count inside the expression and exposure to estimate through a distribution what the beta estimates are. A p value is obtained to determine if the estimates and distribution are statistically significant. Sometimes a distribution that works some may not work for all parameters such as Rayleigh. Strata 1,2, and 3 may work the for P values see Table **9**.

smokA<-survreg(Surv(count,drinkers)~race, data= smoktob, dist="rayleigh")

survreg(formula = Surv(count, drinkers) ~ race, data = smoktob,

dist = "rayleigh")

Coefficients:

(Intercept) race

9.584265 -1.548390

Scale fixed at 0.5

Loglik(model)= -46.3 Loglik(intercept only)= -60.4

Chisq= 28.32 on 1 degrees of freedom, p= 1e-07

n= 12

The data analysis above works for the first strata of the race which is white. The method to vary for the second and third strata is depicted next where is the coefficient of the i variable. B_y =-1.5274 and standard error is 0.150 from R studio software for the tables.

95 % confidence [2] intervals are obtained from the following formula.

$$95\%CI(OR) = \exp^{(\log(OR(z)+-1.96*\sqrt{s.e(y)/n}}$$

For effect modification, please see references [1-3,15].

data nonsmokeraem;

input cases fit zxy xzy aem count;

datalines;

1 1 1 1 2208.8 795

0 1 1 1 1 2586

1 0 301 634.2 1 763

0 0 1735.2 3655 1 4397

1 1 1 1 230.56 111

0 1 1 1 1 233

1 0 24.4 51.5 1 62

0 0 93.9 197.8 1 238

1 1 1 1 250.52 40

0 1 1 1 1 152

1 0 17.7 37.4 1 45

0 0 67.1 141.3 1 170;

run; proc mixed data=nonsmokeraem;

weight count;

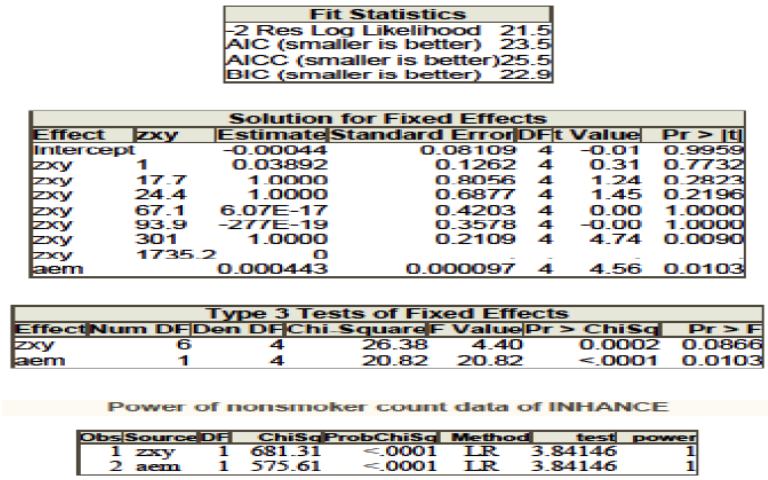
class zxy;

model cases= zxy aem /solution ddfm=satterth covb chisq ; run;

3. RESULTS

The output for effect modification shows a statistically significant relationship for the exposure non-drinking /non-smoking in Figure **1** with P < 0.0001 for chi-square and P<0.01 for F statistics.

The odds ratio of the original data can come from the 2 x 2 of Cochran Mantel Hansel. However, there is little difference or variation between races.



Fit Statistics	
-2 Res Log Likelihood	21.5
AIC (smaller is better)	23.5
AICC (smaller is better)	25.5
BIC (smaller is better)	22.9

Solution for Fixed Effects

Effect	zxy	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		-0.00044	0.08109	4	-0.01	0.9959
zxy	1	0.03892	0.1262	4	0.31	0.7732
zxy	17.7	1.0000	0.8056	4	1.24	0.2823
zxy	24.4	1.0000	0.6877	4	1.45	0.2196
zxy	67.1	6.07E-17	0.4203	4	0.00	1.0000
zxy	93.9	-277E-19	0.3578	4	-0.00	1.0000
zxy	301	1.0000	0.2109	4	4.74	0.0090
zxy	1735.2	0	.	.	.	.
aem		0.000443	0.000097	4	4.56	0.0103

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi Square	F Value	Pr > ChiSq	Pr > F
zxy	6	4	26.38	4.40	0.0002	0.0866
aem	1	4	20.82	20.82	<.0001	0.0103

Power of nonsmoker count data of INHANCE

Obs	Source	DF	ChiSq	ProbChiSq	Method	test	power
1	zxy	1	681.31	<.0001	LR	3.84146	1
2	aem	1	575.61	<.0001	LR	3.84146	1

Figure 1: Effect Modification Output for AEM.

Table 1: Data Analysis for Strata One of White Race for Head Neck Cancer

Beta estimate	OR	95 % CI	HR ($OR_{inverse}$)	HR	HR 95 %CI	Pr(var)	B_0
B_{zy} = -1.341	3.82	4.76,3.07	0.26	0.26	0.26, 0.17	4.60	-1.527
B_{yx} =-1.862	6.43	8.02, 5.16	0.16	0.16	0.19, 0.12	7.94	-2.072
B_x =-.496	**1.64**	2.05,1.32	0.61	0.60	0.76,0.49	1.82	-0.598
$B_{z=}$ -1.548	**4.70**	5.86, 3.77	0.21	0.22	0.33, 0.21	5.72	-1.745
B_{zx} =-1.484	**4.41**	5.49, 3.54	0.22	0.22	0.27, 0.17	5.35	-1.677

Table 2: White Race Stratified

Whites	Cases	No cases
Never Drinkers	795	2586
Never Smokers	767	4397

$OR = \frac{A * D}{B * C}$ = 1.76 and 1/OR=0.56.

Table 3: Black Race Stratified

Blacks	Cases	No cases
Never Drinkers	111	237
Never Smokers	62	238

$OR = \frac{A * D}{B * C}$ = 1.80 and 1/OR = 0.56.

The odds ratio from Cochran Mantel Hansel after its inverse does not equal to a hazard ratio.

Table 4: Hispanic Race Stratified

Hispanics	Cases	No cases
Never Drinkers	40	152
Never Smokers	45	170

$OR = \frac{A * D}{B * C}$ = 0.99 and 1/OR = 1.01.

The variability in the parameter estimate of B_{zy} between z and y come from y and its standard error of 0.1507 produced an approximate decrease from B_z. The increase of B_{yx} comes from the B_y and approximate average of standard errors of B_x (0.1507), B_y (0.1507), and B_z (0.2461). B_{zx} is between the product of z and x whose standard error produce decrease.

The maximum odds ratio of head neck cancer by race from exposure of non-drinking and non-smoking has higher risks for white race. The plot of probability of Hazard ratios for whites and black race shows a curvilinear pattern which is non-normal. Whites have higher probability in the non- normal Pr(z) range and lower hazard ratios for head neck cancer in general in Figure **2**. Whites have lower hazard ratio than blacks on average but higher odds ratio.

Figure **3** shows a linear relationship of odds ratio and probability of Pr(z). Whites have higher non–normal probability than blacks plus higher odds ratios for head neck cancer due to the exposure of non–drinking and non-smoking.

4. DISCUSSION

The results of the analysis from the new method are about non-normal data. This method follows a proof carefully demonstrated to show the relationship of an odds ration expression derived further to include a new probability algorithm and a hazard ratio. Followed through, this method then incorporates a new odds ratio which corresponds to the inverse of the hazard ratio very well. The parameters are then utilized for the development of odds ratios per strata. The parameters can also be explained for terms of interaction, such as

Table 5: Data Analysis of the Black Race (Strata i=2) of Head Neck Cancer

Beta estimate	OR	95 % CI	HR ($OR_{inverse}$)	HRz	HR 95 % CI	Pr(var)	B_0
B_{zy} = -1.341	1.91	2.38,1.53	0.52	0.52	1.65, 0.42	2.16	-0.772
B_{yx} =-1.862	3.21	4.00, 2.58	0.31	0.31	0.39, 0.25	3.83	-1.343
B_x =-.496	**0.82***	1.02,0.65	1.22	1.22	1.52, 0.98	0.77	0.258
$B_{z=}$ -1.548	**2.35**	2.92,1.89	0.42	0.42	0.53, 0.34	2.72	-1.002
B_{zx} = -1.484	**2.20**	2.74,1.77	0.45	0.45	0.56,0.36	2.54	-0.931

*P value not significant for Ralyeigh P < 0.23 (and others not rounded).

Table 6: Data Analysis of the Hispanic Race (Strata i=3) of Head Neck Cancer

Beta estimate	OR	95 % CI	HR ($OR_{inverse}$)	HRz	HR 95 %CI	Pr(var)	B_0
B_{zy} = -1.341	1.27	1.59, 1.02	0.78	0.78	0.98, 0.65	1.35	-0.300
B_{yx} =-1.862	2.15	2.67,1.72	0.47	0.47	0.58, 0.37	2.46	-0.901
B_x =-.496	**0.55**	0.68, 0.44	1.83	1.83	2.27, 1.47	0.42	0.863
$B_{z=}$ -1.548	**1.50**	1.95, 1.25	0.63	0.64	0.79,0.51	1.72	-0.545
B_{zx} =-1.484	**1.47**	5.49,1.18	0.68	0.68	0.85, 0.54	1.60	-0.470

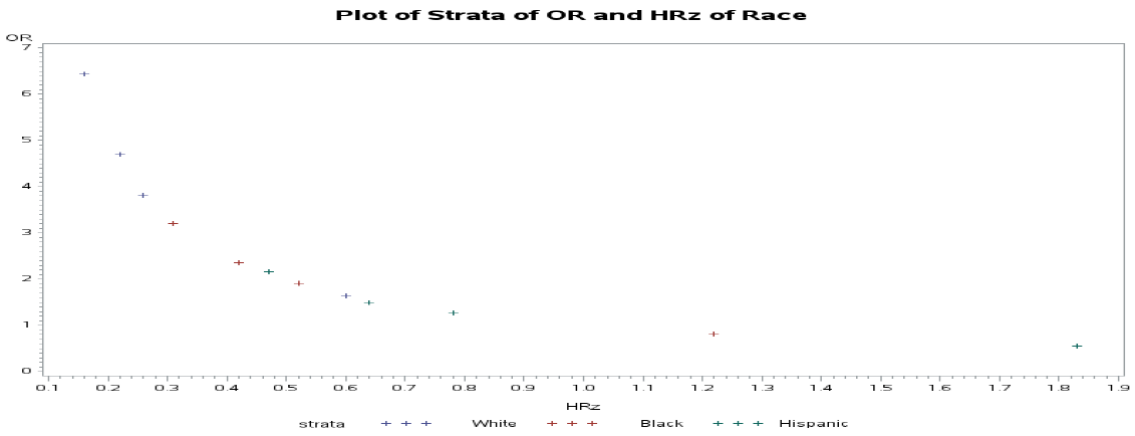


Figure 2: Plot of OR and HRz.

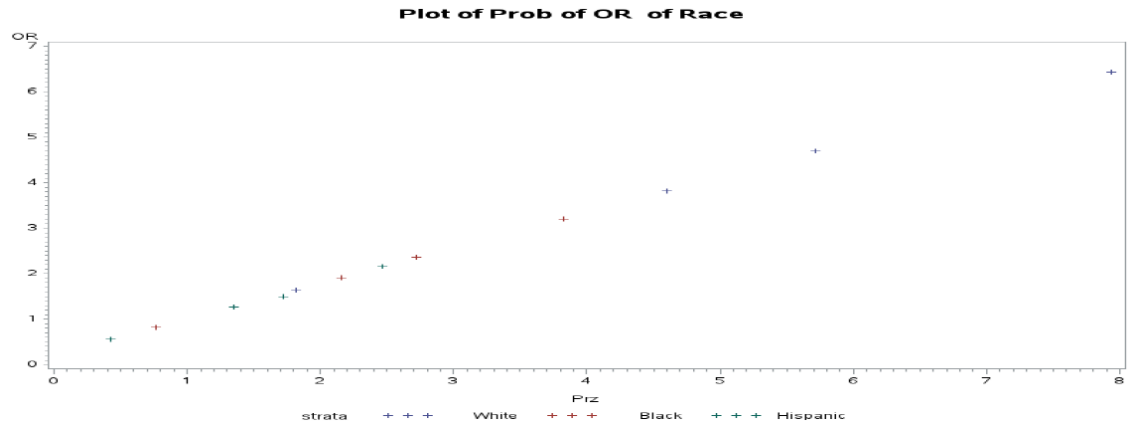


Figure 3: Plot of OR and Probability.

B_{zx} for race and exposure. Next the parameters from the statistical software analysis are compared with expressions involving logits derived to function in calculating odds ratios.

The Shapiro Wilk's distribution shows cases, race, and drinkers are all non-normal with P values of P< 0.0003, P <0.0125, and P <0.0003. Cases were the dependent variable, and drinkers with race were the independent variables. The data from the INHANCE consortium [4] of Europe and involves many parts of the USA and the world.

(1a) $$B_0 + B_1 * X_1 + B_2 * X_2 = \frac{\exp\frac{-y}{1-y}}{B_1 * X_1 + B_2 * X_2}$$

(1b) $$OR = \exp(\exp(\frac{-y}{1-y}) + B_1 * X_1 + B_2 * X_2)$$

(1c) $$OR = y + \Pr(z) * (1 - y)$$

Expression (1a) results in an odds ratio of 1.63 and a HR of 0.61 from inverse. Table **5** shows an odds ratio of 1.91 for race from black race and outcome from exposure of non-drinkers with HR about 0.52. From the statistically significant 95 confidence intervals viewpoint, this may be important as in Table **5**. Logistic regression had parameters of B_0 -1.2806 and B_1 for nondrinkers as -0.1949 and 0.2370 for black race for interaction terms in the model for race and drinkers. From the expression above the B_y is -1.1079 (1a) for intercept and race being black with non-drinkers which is obtained from expression (1a). Then proceed to the expression y/1-y which is exponentiated for an odds ratio of 1.63 with from the expression above with B_y = -1.1079 (1a) for intercept and race being black with non-drinkers. Logistic regression reported an odds ratio of 1.53 for black race. It is only about a -6 % difference. For an interaction term that is significant such as race and nondrinkers the beta estimate is 0.1979 for race and nondrinkers from logistic regression with the interaction term then the expression (1c) which still works since it is statistically significant with P <0.0237. The Pr(x) is then calculated as 0.77 with 1.22 HR(x) and the OR (x) algorithm of 0.82 after B_y -1.5274 is utilized for Pr(z). The E Miner 14.2 shows an odds ratio

of 0.88 for race white vs Hispanic for Beta of 0.1979 in SAS 9.3 which does not yield an estimate in SAS.

Table **1** for white race has an odds ratio of 4.70 for head neck cancer. The exposure plus race has an odds ratio of 4.41 for head neck cancer. Race and outcome has an odds ratio of 3.82 for white race. The exposure of non-drinking/non-smoking has an odds ratio of 1.64 the outcome with exposure has an odds ratio of 6.43.

The HR(z) is then 2.35 of the second strata (the inverse of this gives an odds ratio of 0.42 for the exposure of black race from Pr(z) and HR(z) relationship). The odds ratio of 0.82 from the Table **5** and equation (1c) with B_x of -0.496 with Pr(x) is 0.77 for black race and HR (x) of 1.22 as potential confounder based on non-probability the beta estimate for black race was not statistically significant vs 0.82 from Table **5** and B_x. The relationship here is an inverse for the hazard ratio and odds ratio. The odds ratio for race and exposure for black race is 2.20 from the exposure non–drinking/ non-smoking (see table **5**) for head neck cancer exposure and cases the outcome has an odds ratio of 3.21 for head neck cancer. Race and cases has an odds ratio of 1.91 for black race. The intercept is not fixed to 0 as shown below.

(2) $-\ln \Pr z = B_0$.

(3) solve Prz for y or

(4) Obtain B_y.

(5) $HR = \frac{y}{1-y}$ is the coefficient for hazard ratio.

(6) $Hazardratio\_coef = \exp^{HR}$

(6b) $Hazardratio = \exp(\exp^{HR})$

(7) $OR = \frac{1}{Hazard\_ratio}$

Expression (7) results in a parameter for cases as the outcome as B_y =-1.1079 for race being black with nondrinkers and intercept. Expression (1b) results in an odds ratio of 1.70 with the expression above the B_y is -1.1079 (1a) for intercept and race being black with non-drinkers with B_1 of -0.1949 (nondrinkers) and B_2 of 0.2370 for black race.

For the potential confounder of race being black plus exposure of non-drinkers, the result is different. The exponential of $B_1*X_1*B_2*X_2$ and other parts of the logit must be done as an interaction term.

The odds ratio for race being Hispanic vs whites for confounder and exposure non-drinkers the odds ratio is 1.50 from the Rayleigh distribution. Confounding may be possible by race being Hispanic and non-drinkers vs non-smokers both statistically significant from Table **6**. Hispanic race has an odds ratio of 1.50 for head neck cancer from non drinking-non-smoking. While race and exposure has an odds ratio of 1.47 for head neck cancer for white, black, and Hispanic race. The exposure of nondrinking/non smoking has an odds ratio of 0.55 for Hispanic race for head neck cancer..

PROC CORR of the data shows there is a statistical correlation significant at alpha =0.05 level. Pr(z) and OR are 99.99 % correlated with P <0.0001; while HRz and Pr(z) are -74.77 % correlated with P <0.0052. Strata is not statistically significant for HRz nor hazard ratio HR (P<0.0626 and P< 0.0644). HR is independent with Pr(z) with P values of P<0.0051 at -74.89 % correlation. Strata and HR is also not significant statistically and OR is statistically correlated -68.55 % with P <0.0139; and Pr(z) is also statistically significant -68.30 % with P<0.0143 with strata.

The odds ratio for negative serology from D' Souza and others [5,9] is 2.75 from the expression (1c) and hazard ratio is 0.363 from inverse with Pr(z) 3.31 and B_y =-1.4162 for head neck cancer with [5] HPV cases and E6/E7 serology being positive. Possibly amending the positive serology with this method and an odds ratio of 1.68 with Pr(z) of 1.90 and hazard ratio of 0.60.

The same dataset for gender female (0) and male (1) for cancer of oropharynx [5] from D' Souza and others 2007 yields a B_x (gender) of 16.7437 from Weibull distribution and a B_y of 15.4257. The beta estimate for x is 16.7437 for cancer of the oropahrynx with an odds ratio of 1.64, HR(z) at 0.61 and Pr(z) as 1.82. Hence the odds ratio of male vs female is 1.64 for cancer of the oropharynx with HPV. Ritchie, Smith et. Al 2003 [10] also has an odds ratio of 1.6 for males to females for cancer of oropharynx and HPV infection included.

For the parameters from head neck cancer for white race, however, one expects an increase in OR from 0.82 to 0.55 for races being black and Hispanic from race for head neck cancer from B_x.

PROC GLIMMIX resulted in an odds ratio of 1.77 for white males and 1.83 for black race for interaction race

and drinkers. The white race has beta estimate of 0.5719 and black race an estimate of 0.6036 with odds ratios of 1.73 and 1.76 for whites and black race for head neck cancer. B_y of -1.5274 is utilized. (Then the next step is with expression (1c) but with PROC GLIMMIX please exponentiate it for the odds ratio).

PROC GENMOD has beta estimates: nondrinkers with an estimate of -0.5478; black race an estimate of -0.3005 and an intercept of -1.064 with Pr(z) as 2.8988. Solving for B_y from Pr(z) yields B_y as 0.8138. Expression (7) results in an odds ratio plus hazard ratio of 0.85 and 1.18 with black race and the intercept. For non-drinkers B_y becomes 0.6158 and Pr(z) also 2.8988. Hence of expression (7) the odds ratio is 0.89 and a hazard ratio of 1.12 compared to 0.82 from B_x from Table **5** for black race. The expression (1c), results in an odds ratio of 1.73 for nondrinkers. Black race results in an odds ratio 1.35 for expression (1c) with B_y as -1.5274.

The solution with y/1-2y [2] as a formula for bivariate odds ratio developed is for the coefficient and then exponentiated 2 times. For B_y of 0.6158 non-drinkers, the odds ratio is exp(-0.6850) exponentiated to 1.655. Then for black race and B_y as 0.8138 from Pr(z) and intercept of PROC GENMOD again, the odds ratio is exp(-0.6423) exponentiated to 1.69. The odds ratio from black race vs white form exposure non-drinking and non-smoking is 1.69 times for head neck cancer. For the non-drinkers, the odds ratio is 1.66 for races being white, black, and Hispanic for head neck cancer from the exposure of no exposure. These answers are like PROC GLIMMIX of odds ratios of 1.77 and 1.83 for whites and blacks for head neck cancer and non-drinking and non-smoking. B_x renders an odds ratio of 1.64 for white race from Table **1**.

Another expression for odds ratios [1] yields a method of 2-y/ 1-y where y is the exponential of y which yields again with the same estimates a value of 2.31 for non-drinkers and 3.40 for black race for head neck cancer. B_{yx} and B_{zy} has odds ratios of 3.21 and 1.91 from Table **5** for the black race for cases and exposure plus an estimate of race with cases estimate. B_{zx} has odds ratio of 2.20 and HR of 0.45 from Table **5**.

The white and Hispanic race shows good power upon analysis by a power method for statistics such as for odds ratio, hazard ratio, and probability. The Tables **5** and **6** were included from the analysis. A test on power of multiple regression estimates shows an 83 % power for odds ratio estimate OR. There were excellent results from RANTBL function. For HRz, the power is 80.9 % based on a chi square test with the RANTBL function [11]. Pr(z) has a power of 82.7%.

The results of the effect modification question for this head and neck cancer data can be seen in with P value of P<.0.0001 and F statistics P <0.01 [2]. One can also observe the results more for detail in Figure **2** [2]. The question of effect modification for nondrinking and non-smoking for chi -square statistic has a P <0.0001 which is statistically significant [2] with PROC MIXED [14]. Hence one will have to reject the null of homogeneity and conclude that there is interaction based on the chi square distribution. By the F statistics, the P value is approxiamtely P <0.01 which is also statistically significant as a multivariate statistic for alpha = 0.05. Effect modification exists for this non-normal data and the random effects will allow us to generalize that risk is possible even from the non-drinking and nonsmoking issues due to differences across race being white, black, and Hispanic for head neck cancer. The power of this method is 100 % based on the RANTBL function [2] (see Figure **2**. in Agravat 2012) and the variable 'aem'. The current test fails to find a significant p value P <0.06 in SAS software or the Breslow Day test. Cancers of the oral cavity are serious in terms of risk from smoking cigar or pipes then risk of the oropharynx/hypopharynx (not including Central Europe). Hasibe *et al*. [4] shows risk greater than 10 percent variation for different cancers in this study.

In this case of effect modification, with regards to non-drinking and non-smoking as the exposure, the question is after the P value is shown to be statistically significant with P <0.0001 for chi squares and P <0.01 for F statistics. The next step is to determine if there is a per strata explanation by odds ratios of differences per strata of race being white, black, and Hispanic. Tables **1**, **5**, and **6** show for the parameter of exposure B_x that the odds ratios are for white race 1.64; for black race 0.82, and Hispanic race 0.55. This demonstrates a risk difference of 50 % for white vs. black race. Next, the odds ratio has a 32 % difference with black and Hispanic race. While the white and Hispanic race has 66 % difference which is all clearly greater than 10 % difference required to prove effect modification did occur and report it. The null of homogeneity can be rejected for no difference between strata accordingly for the outcome of head neck cancer by exposure non-drinking/non-smoking from odds ratios perspective.

In addition, the problem of odds ratios and interaction can be found with PROC GLIMMIX and

Table 7: Baseline Hazard Distribution for Head Neck Cancer and Race for New HR`z

Race	Prz	P	HRz	HR`z	HR`z/Prz	S(z)	HR`z/Prz/S(z)	F(z)
White	5.72	0.17	0.22	0.84	0.15	0.80	0.18	0.20
Black	2.72	0.37	0.42	0.60	0.22	0.66	0.33	0.34
Hispanic	1.72	0.58	0.64	0.46	0.27	0.53	0.50	0.47

understood with the understanding of a new logit y/1-2y. Black race has an odds ratio slightly higher of 1.83 after interaction with non-drinkers for head neck cancer than white race whose odds ratio is 1.77. The new logit's solution is a close approximate with odds ratios of 1.66 for race overall and head neck cancer which matched earlier estimate of 1.64 for Ritchie *et al*. The odds ratio for power of the odds ratio of 1.69 with PROC GENMOD had 99.9 % power from proc power.

The risk factors for head neck cancer normally involve heavy smoking, alcohol, HPV, EBV, and mutations to TP53 gene. Other factors are vitamin deficiencies [2, 16] such as Vitamin A and Iron in Plummer Vinson syndrome. HPV is associated with EBV and a non-keratinizing cancer which is type II and type III often occurring in Africa and Asia. Squamous cell cancer is 25 % more frequent with HPV and 60 % with oropahryngeal cancer of the tongue and palatine tonsils. Amplification of CDKN2A is the type of cell regulation occurring. High levels of epidermal growth factor receptors are associated with poor prognosis with tumor factors.

Treatment for head neck cancer can be done with cetuximab, and docetaxol. Patients with squamous cell cancer of the head and neck (SCCHN) at the local advantage [12] stage can be done with docetaxol especially this with distant metastases. This is especially for the head neck cancer of the floor of the mouth, tongue, tonsils, and larynx. SCCHN is associated with tobacco and alcohol abuse. Oncologic supportive care improvements are part of the advances in medical care. The survival increase for stage IVa/b is 35 % to 37 % from docetaxol and stage III for 44 % to 46 %. As EGFR is over expressed, cetumixab which [13] is a monoclonal antibody works on human EGFR on SCCHN. It works to block ligand function and receptor function. Cetumixab works in conjunction to radiotherapy.

This table shows a trend where probability *non-normal* which is originally intended than the first attempt in SAS [2] and hazard ratio for head neck cancer is having less high risk at baseline for Hispanic race followed by black and then white. However, baseline hazard function derivative followed by probability Prz is increasing per strata. The death rate or survival probability is higher for white, than black followed by Hispanics. Next the same parameters shown in Table **7** divided by survival rate is less for white, then black, followed by Hispanics. *The lower the non-normal probability, the lower the baseline hazard derivative for head neck cancer but not in terms of hazard ratio for the exposure non-drinking/non-smoking by race for head neck cancer.* HR`z/Prz/S(z) approaches the Cumulative Distribution Function F(z). The death distribution, for black race, the hazard ratio is 0.42, baseline hazard is 0.60, and is close to survival probability 0.66.There is a 42 % increase in death for black race by exposure nondrinking/non smoking and race. The survival probability is 0.60 for black race close to baseline hazard or derivative. The existence of effect modification can explain this risk partly from nondrinking/nonsmoking with P values P <0.0001 and P <0.01 for chi square and F statistics which can partly imply interaction by race. Hazard ratio analysis also shows effect modification analysis too.

In addition, the odds ratio can then be generalized for $OR = y + \frac{1+z}{z}(1-y)$. Since Pr(z) is ~ 1+z/z, one can then compare the odds ratios in Table **1** for example. The exposure B_x =-1.548 and can then yield: 0.217 +5.70*(0.783) or approximately 4.67 or 4.70 as obtained through Pr(z) for B_y=-1.5274 (R Studio) and B_x=-.3005. $\frac{1+z}{z}$ Is the probability upside down for any event? The exponential of y can be from R software as well. In the Proc Genmod discussion, recalculated odds ratios can be:

1a) $OR = y + \frac{1+z}{z}(1-y) = 2.256 + 2.35 * (-1.2564) = -.6972$

1b) exp (exp (-.6972)) =exp (0.4979)

1c) the odds ratio for black race vs. white is then 1.64 with B_y from Prz and B_x =-0.3005 .

2a) $OR = y + \frac{1+z}{z}(1-y) = 2.256 + 2.729 * (-1.2564) = -1.1732$

2b) exp (exp (-1.1732)) =1.36 for B_x =-0.3005 (SAS) and black race with B_y =-1.5274 (R studio).

2c) For nondrinkers (nondrinkers with an estimate of -0.5478 (Proc Genmod SAS) and B_y=-1.5274) the odds ratio will then be in the general form of 2.36 or the odds of head neck cancer for exposure nondrinkers/ nonsmokers for black race is 2.36. The 2.72 (see Tables **5** and **7**) is the non-normal probability for black race and about matches 1.36 for odds ratio for black race matching Table **5**.

Proc Glimmix which produces the interaction odds ratio like the Cochran Mantel Hansel estimate 1.77 and 1.83 vs 1.76 and 1.80 cannot be considered effective for effect modification. The comparison to the odds ratios for B_x is however acceptable. Another term for odds ratio and effect modification is B_z. If the exposure is independent B_x, and then significant for 10 percent difference across strata one can continue to the race strata (4.70, 2.35, and 1.50 all very different) for head neck cancer where there can be confounding possibly. The possibility of non –passive smoking as a risk is possible for head neck cancer for future discussions through evidence that passive smoke exposure results in effect modification for lung cancer with P <0.0001 [1] despite non drinking/ nonsmoking as exposure. Minor differences across strata for hazard ratio exist for death risks and can be seen in Table **7**.

When logistic regression calculates odds ratios and estimates are shown, there is a similarity to the new method with a twist: $OR = y + \frac{z}{1+z} * (1-y)$ (probability) vs (non-normal probability or flip) $OR = y + \frac{1+z}{z} * (1-y)$ the effect is startling for estimate of 0.2429 (A) for black race and -0.1949 for nondrinkers (without interaction of race and drinkers) and with By =-1.1079 and -0.2739 non-drinkers without interaction (C).

A) $OR = y + \frac{z}{1+z} * (1-y)$ = 0.33025+0.5604*(1-.33025) = 0.70 for black race.

B) $OR = y + \frac{z}{1+z} * (1-y)$ = 0.33025+ 0.4319*(1-.33025) = 0.62 for nondrinkers.

C) $OR = y + \frac{1+z}{z} * (1-y)$ = 0.33025+0.5604^{-1}*(1-.33025) = 1.53 for black race.

Non-drinkers have an odds ratio of 0.58 (SAS) similar to expression C and 0.62. However the true odds ratio will be flip of the probability or 1.88 for nondrinkers and no interaction term. The flip of black race and odds ratio will be 1.53 which is correct for the potential confounder and non-normal probability.

Another trick is from Proc Genmod that is for the exposure and odds ratio from y/1-2y logit: .(y/0.5-y)=2*(B_0+B_1*x1+B_2*x_2) results in B_y=-0.4048 whose exponentials are 0.6671 and 1.94 as odds ratio from Beta estimate of -.3005 and -.5478 for black race and nondrinkers. The normal logit will render 1.36 odds ratio for black race and nondrinkers. Both logistic regression and Proc Genmod will then be similar for nondrinkers 1.88 and 1.94 to Proc Glimmix and 1.83 for black race with the interaction included.

CONCLUSION

In the text, there are several proofs which may find support for why this method can be considered important: Independence, probability, hazard ratio, and then odds ratio proofs. The steps of the proofs are clearly written. The power of the statistics is clear correlation of the correlation is correct. The ease of the method is another reason why it can be utilized than others. The formulas are stepwise and there are also many statistics which can be added to the sample for analysis. In epidemiology and in Social Science, there can be many samples of data where this method can be found practical. One can proceed with this method for better statistics.

ACKNOWLEDGEMENT

The author wishes to acknowledge his friends like Rekha for inspiration words along with family plus a loving wife Kavita. I want to thank the editors and reviewers too plus the support. I wish prayers too.

REFERENCES

[1] Agravat M. Formula's Calculating Risk Estimates and testing for Effect Modification. Pharmasug 2011 [Internet]. 2011 [cited 1 October 2017];. Available from: http://www.lexjansen.com/pharmasug/2011/SP/PharmaSUG-2011-SP03.pdf [

[2] Agravat M. Effect Modification, confounding, hazard ratio, distribution analysis, and probability of non-normal data for head neck cancer. Global SAS 2012 [Internet]. 2012 [cited 1 October 2017];. Available from: http://support.sas.com/resources/papers/proceedings12/315-2012.pdf

[3] Agravat M. Method for Calculating Odds Ratios Relative Risks, And Interaction. University of South Florida Research Day 2008 [Internet]. 2008 [cited 1 October 2017];. Available from: http://health.usf.edu/medicine/research/ABSTRACT_BOOK_.pdf

[4] Hasibe *et al*. M. Alcohol Drinking in Never Users of Tobacco, Cigarette Smoking in Never Drinkers, and the Risk of Head and Neck Cancer: Pooled Analysis in the International Head and Neck Cancer Epidemiology Consortium. Journal of National Cancer Institute. 2007; 99(10): 777-789.

[5] Agravat M. Odds Ratio and Hazard Analysis of Head Neck Cancer by Hpv Status with New Logits and Probability. [Internet]. 2013 [cited 1 January 2017];. Available from: https://www.worldwidejournals.com/indian-journal-of-applied-research-(IJAR)/file.php?val=November_2013_1493100857__122.pdf

[6] Agravat M. Causal Inference. Sciencewise [Internet]. 2016 [cited 21 December 2017];. Available from: https://www.researchgate.net/publication/299423211_Causal_Inference

[7] Agresti A. An introduction to Categorical Data Analysis. Wiley and sons; 1996.

[8] Hosmer D, Lemeshow S. Applied Survival Analysis. Somerset: Wiley; 1999.

[9] Sousa D. Case-Control Study of Human Papillomavirus and Oropharyngeal Cancer. The New England Journal of Medicine 2007; 356: 1944.

[10] Smith R. Human papillomavirus infection as a prognostic factor in carcinomas of the oral cavity and oropharynx. International Journal of Cancer [Internet]. 2003 [cited 21 December 2017]; 336–344. Available from: http://onlinelibrary.wiley.com/doi/10.1002/ijc.10960/pdf

[11] Van Ness P, Alford T, Dubin J. Power Simulation for Categorical Data Using the RANTBL Function. Sugi 30 [Internet]. 2017 [cited 1 October 2017];. Available from: http://www2.sas.com/proceedings/sugi30/207-30.pdf

[12] Rapidis, Sarlis, Lefebre. Docetaxel in the treatment of squamous cell carcinoma of the head and neck. Therapeutics in Clinical Risk Management [Internet]. 2008 [cited 10 October 2017];4(5):865-886. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2621396/

[13] Pol S, Vermorken j. Cetuximab: its unique place in head and neck cancer treatment. Biologics Targets and Therapy [Internet]. 2013 [cited 10 September 2017]; 7: 77-90. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3665438/

[14] Agravat M. A New Effect Modification P Value Test Demonstrated. SESUG 2009 [Internet]. 2009 [cited 10 October 2017];(Statistics and Data Analysis). Available from: http://analytics.ncsu.edu/sesug/2009/SD018.Agravat.pdf

[15] Kumar, Cotran, Robbins. Basic Pathology. WB Saunders; 1999.

Received on 17-11-2017 Accepted on 25-02-2018 Published on 08-05-2018

https://doi.org/10.6000/1929-6029.2018.07.02.3