RESEARCH PAPER | Medical Science | Volume : 3 | Issue : 11 | Nov 2013 | ISSN - 2249-555X

# Odds Ratio and Hazard Analysis of Head Neck Cancer by Hpv Status with New Logits and Probability

KEYWORDS: Odds ratios; Hazard statistics; Logits; Probability; Head Neck Cancer

**Manoj B Agravat**

MPH, Data Analyst Predictive Statistician, 20108 Bluff Oak Blvd Tampa, Florida 33647 USA

**ABSTRACT** *The analysis of head neck cancer through multivariate statistical analysis is made possible through new methods for odds ratios and hazard ratios by a probability algorithm and new models. A case control study of HPV16 and oropharyngeal cancer is analyzed with logits that improve the risk statistics for three levels of smoker status. The new logit is derived and accommodates bivariate outcomes. Risk factors and exposures such as smoking, HPV status and E6 and E7 elongation factors are examined for the outcome of head neck cancer by various new methods for typical fixed effects models.*

**Introduction**

Risk factors for head neck cancer include betel nut use and other tobacco products such as pan, gutka, hookah, bidi, plus alcohol (Sharma 2012). Higher incidence makes the fixed effects model limited. A case control study is analyzed for HPV status and head neck cancer for hazard ratios and odds ratios by HPV and oropharyngeal cancer and smoker status in pack years (see table 2).

The probabilities, odds ratios, in the discrete probability algorithm will allow calculations and new methods for predictions in head neck cancer outcomes (D`Souza et.al. 2007 , tables 1 and 3) and survival times. E6 and E7 elongation factors are compared for odds ratio statistics by negative and positive status (see table 2). The AML, or Agravat' matrices and logit method, is proposed to solve beta estimates of count data of oropharyngeal cancer. Unique beta estimates are possible through this new method per strata. The matrix created is a 4x4 matrix multiplied with an identity matrix. The count data is summed up for digits for first row then zeros second row followed by the count per strata then one as shown below.

**Example 1. Sample Matrix: Smoker Status (Strata=1)**

1 0 0 0
0 1 0 0
0 0 1 0
0 0 0 1

3 0 44 1
2 1 56 1
2 1 119 1
1 2 81 1

**Example 2. Sample Matrix: E6 and E7 Serology Status (Strata=1)**

3 0 36 1
2 1 64 1
2 1 192 1
1 2 8 1

1 0 0 0
0 1 0 0
0 0 1 0
0 0 0 1

**Methods**

By definition of odds 1=P1 / (1-P1), and odds 2 = P2/ (1-P2). The ratio of OR1 and OR2 gives an odds ratio. However this works well for binomial covariates based on the binomial distribution. The assumption requires homogeneous variance for power to be sufficient. If there were heterogeneous covariates, then random effects are not effectively analyzed or handled in the binomial distribution because of more than two levels of confounders. In these new risk statistics of the author, the linearity assumption is not necessary. The independence assumption however limits the new method proved in Agravat 2011 for the discrete probability algorithm utilized.

$$x = \frac{a}{1-a}$$

$$\frac{x}{x+y} + \frac{y}{x+y} = .5$$

$$-.5 \ln y = -\ln y$$

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

$$\frac{x}{x+y} + \frac{y}{x+y} + \frac{z}{x+z}$$

$$-\ln y + \frac{z}{x+z}$$

$$-\ln y = \frac{\frac{odds_z}{1+odds_z}}{\frac{odds_z}{1+odds_z} - 1 + \frac{odds_z}{1+odds_z}}$$

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

$$\frac{y}{1+y} = \frac{x}{x+y} + \frac{y}{x+y} + \frac{z}{x+z}$$

To begin the new logit derivation, start with the concept of the average of two odds which each represents a variable and its probability as well as the possibility that may be achieved where the exponential of logits can be calculated (Agravat 2012) as in equations derived. Next natural log follows the estimates of the sum of a model. Then the model is substituted into the discrete probability algorithm. The result is that the equation that may help to solve beta for estimates and next odds ratios calculated through the equations for beta estimate of exposure for new logits and equations and however the intercepts are not fixed to 0 calculated by equation which is a relationship of probability derived through inverse equations (Agravat 2009).

The method to calculate beta estimates of the confounder and exposure variables are as follows: 1) exponentiation of the new logit 2) solve for the new relationship of (y)/(1 - y) 3) take the natural log of both sides of the equation of logits 4) solve for By. The variables such as probability of x exposure, y outcome, and z the confounder are algebraically combined to solve for the logit and odds of y. Beta x can be solved by the algorithm below.

$$x = \frac{(z - 2yz)}{y}$$

Several equations are shown to handle and predict head neck cancer

Volume : 3 | Issue : 11 | Nov 2013 | ISSN - 2249-555X

survival times, hazards, beta intercepts for non-normal data, as well as a logit.

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

$$B_0 = -\ln P_{znew} = \frac{-\ln(\frac{1}{HR}) - y}{1 - y}$$

$$OR_e = \exp^{\beta_0 + \beta_z + \beta_x}$$

$$P_{znew} = \frac{\frac{1}{z} - y}{1 - y}$$

$$LogitP_{val} = P_z - \frac{\frac{1}{z}}{1 - y}$$

**Results**

The odds ratios for outcome of oropharyngeal cancer and exposure for the nonsmoker is 3.549 compared to control and first strata. For 1-19 pack years the odds ratios is 5.166 vs 5.262 for 20 plus pack years and statistically significant confidence intervals (see table 1) based on outcome of y and odds of outcome of for beta y at 0 solved by $B_0$. In the case of HPV status and head neck cancer by smoker status, By = -1.3074 for z=1, By = -1.7099 for z=2, and By = -1.4813 for z=3 for nonsmoker, 1-19 pack years and, 20 pack years. The hazard ratios were 0.32, 0.193, and 0.253 for the three respective strata. Nonsmokers had the highest hazard ratio, 0.32, with 20 plus pack years next, hazard ratio of 0.253, next followed by 1-19 pack years and HR of 0.193 due to alcohol possibly being a more important risk for oropharyngeal carcinoma and HPV in the absence of smoking.

**Discussion**

Human Papilloma virus's link to head neck cancer type II or non- keratinizing neck cancer by the World Health Organization(WHO), dictates that HPV has a specific role in the pathogenesis of head neck cancer and squamous cell carcinoma. HPV is involved in 25 percent of the cases of head neck cancer. However, 60 percent of these head neck cancer involve the lingual and palatine cancer. E6 and E7 proteins are known to inactivate TP53 retinoblastoma tumor suppressor gene (Cotran, Kumar, and Robbins 1989). Mutations in mitosis may be affected by TP53 status and elongation factors 6 and 7. There are greater hazard ratios with E 6 and E 7 present, since inactivation occurs of TP 53 and RB suppressor gene, cell regulation goes haywire and apoptosis may result possibly in G1 interphase when P53 increases.

If ulceration is present there may be active hyperplasia (Marur and Forastiere 2008) and more frequent mitosis and possibly deletion or translocations causing apoptosis through malfunction of cell regulation and metabolism. The new method yields an odds ratio of 0.828 for control of negative status of E6 and E7 negative status by HPV 16 marker for oropharyngeal cancer. The odds ratio of 0.894 is also statistically significant (see table 2). The odds of oropharyngeal cancer is 4.497 times that of control for outcome based on odds of serology negative status and 2.794 for positive status. The Shapiro Wilk's P value indicates a non-normal distribution with P < 0.0009 for serology and case, as well as control.

The author states that the odds ratios are 0.828 vs. 0.894 for the two strata. The odds ratios for negative E6 and E7 status for HPV-16 and oropharyngeal cancer are 4.497 times than control for outcome vs. 2.794 which is less and statistically significant with 95 percent confidence intervals (see table 2).

The next method involves the same AML method except that the P (y | x) is the algorithm where y represents the outcome and x the exposure. The ratio of numerator to denominator yields the odds ratios $P(y|x) = \frac{(B_z - B_x)(B_y - B_x)}{(B_y - B_x)^{B_z}}$ where $B_y$ and $B_x$ are exponentiations. The 95 % CI are (5.002, 4.041) and (3.108, 2.511) are statistically significant for the two strata of serology status. Questions may involve E6 E7 and regulation for oncogenes when increase and proliferation stages may occur and result in carcinoma. The probability of the outcome is 0.301 from By = -1.4162 for strata one of negative status and $P_{znew}$ is 3.316. $P_{znew}$ (non-normal probability) is 1.9023 or 0.525 inverse (for normal probability). The difference in probability suggests that mistakes may result with greater harm from the positive status of E6 and E7 serology then negative status sequentially because normal probability is lower for negative serology. The hazard ratios are 0.363 for negative serology and 0.620 for positive serology of E6 and E7 for HPV16 and oropharyngeal cancer.

**Table 1 Oropharyngeal Cancer HPV status by Smoker Status with AML and Probability**

| Smoker Status (Pack years) | Odds Ratio (Confounder) | 95 CI | P Value | Odds Ratio (Outcome) P(y\|x) | 95 CI |
|---|---|---|---|---|---|
| Non-Smoker | 0.650 | (0.692,0.609) | 0.0057 | 3.549 | (3.784,3.330) |
| 1-19 | 0.936 | (0.999,0.880) | 0.0082 | 5.166 | (5.848,5.404) |
| 20+ | 0.738 | (0.768,0.692) | 0.0080 | 5.262 | (5.607,4.938) |

The baseline hazard function is less for negative serology -3.981 than for positive -2.127. Compared to non-normal probability, the baseline functions are -1.2 to –1.118. With respect to baseline hazard function, probability, and survival time, a new statistic (Agravat 2012), is -1.727 vs. –2.081. This statistic shows the baseline hazard with respect to survival probability and probability with slightly worse risks at baseline based on probability and survival for negative serology with respect to E6 and E7. The hazard ratio is more for positive serology 0.620 for hazard ratio of oropharyngeal cancer and positive serology then for negative E6 and E7 strata vs. 0.363 for negative serology of E6 and E7 (see table 3). The death distribution, or cumulative distribution function, for negative serology of E6 and E7 is 0.304 vs. 0.462 for positive.

**Table 2 Serology Status for E6 and E7 and New Odds Ratios with P Values**

| Serology Status | Odds Ratio Confounder | 95 CI | P Value | Odds Ratio (Outcome) P(y\|x) | 95 CI |
|---|---|---|---|---|---|
| -E6 / E7 Elongation Factors | 0.828 | (0.921,0.744) | 0.0064 | 4.497 | (5.000,4.041) |
| +E6 / E7 Elongation Factors | 0.894 | (0.994,0.803) | 0.0046 | 2.794 | (3.108,2.511) |

**Table 3 Serology Status of E6 and E7 and New Hazard Statistics of Survival Distributions**

| Serology | $P_{znew}$ | HR(z) | S(z) | F(z) | -HR'(z) | -HR'(z)/P(z) | -HR'(z)/P(z)/S(z) |
|---|---|---|---|---|---|---|---|
| -E6 / E7 | 3.316 | 0.363 | 0.695 | 0.304 | -3.981 | -1.200 | -1.727 |
| +E6 / E7 | 1.902 | 0.620 | 0.537 | 0.462 | -2.127 | -1.118 | -2.081 |

The risks from smoking indicates that when risks from passive smoke exposure (AgravatA, 2009) are statistically significant (Agravat, 2011), the increased risks (see table 1) from smoking exposure as by HPV status and pack years show that the effect from smoke exposure is more than what is expected or negative confounding is the case for cigarette smoking and possibly its pollution. E6 and E7 serology status indicates that there are risks from exposure which are only slightly different or less than 10 % from $OR_e$ logit equation 0.82 vs. 0.89

Volume : 3 | Issue : 11 | Nov 2013 | ISSN - 2249-555X

per confounder. The suspected risk may involve a mechanism related to malfunction of mitosis and cell regulation linked to TP53 and retinoblastoma suppressor gene causing apoptosis where binding proteins and malfunction results in dysfunctional anaphase proteins.

**Acknowledgement**
Thanks to Dr. Bansidas M Agravat, Gauri B Agravat, (my parents), family and friends for encouragement.

**REFERENCE**
AGRAVAT(A), M. B. (2009) A New Effect Modification P value Test Demonstrated. | http://analytics.ncsu.edu/sesug/2009/SD018.Agravat.pdf | | | AGRAVAT, M. B. (2009) New Method for Calculating Odds Ratios, Relative Risks, and | Confidence Intervals after Controlling for Confounders with Three Levels. (unpublished | submission: American Journal of Epidemiology). | | AGRAVAT, M. B. (2011). Formulas Calculating Risk Estimates and Testing for | Effect Modification and Confounding. PROCEEDINGS, Statistics and Pharmacokinetics. | http://www.lexjansen.com/pharmasug/2011/SP/PharmaSUG-2011-SP03.pdf. | | AGRAVAT, M. B. (2012). Effect Modification, Confounding, Hazard Ratios, | Distribution Analysis, and Probability of Non-normal Data for | Head Neck Cancer. PROCEEDINGS, Statistics and Data Analysis. | http://www.sascommunity.org/wiki/Effect_Modification_Confounding_Hazard_Ratio_Distribution_Analysis_and_Probability_of_Non-normal_Data_for_Head_Neck_Cancer. | | AGRESTI, ALAN (1996). Introduction to Categorical Data Analysis. Wiley Series in | Probability and Statistics 1996. | | COTRAN, KUMAR, ROBBINS (1989). Pathologic Basis of Disease. W.B.Saunders Company | 4th edition. | | MACCOMB, FLETCHER (1971). Cancer of the Head and Neck. MD Anderson Hospital | and Tumor Institute. | | MARUR AND FORASTIERE (2008). Head Neck Cancer: Changing Epidemiology, Diagnosis, | and Treatment. Mayo Clinic Proceedings, Symposium on Solid Tumors. | April 2008; 83(4):489-501. | | SAS INSTITUTE CARY NC VER 9.3 2012. | | SHARMA, Manoj (2012). THE WORLD NO TOBACCO DAY 2012 (WHO). | http://groups.yahoo.com/neo/groups/Quality_of_Medical_Education/conversations/topics/5686 | | | D 'SOUZA ET.AL. (2007). Case-Control Study of Human Papillomavirus and Oropharyngeal Cancer. The New England Journal of Medicine. May 10 2007; 356:1944- |