CANCER, ARSENIC, MORTALITY AND THEIR RELATIONSHIP

a book by

Dr. Virendra Bharti

Executive Summary (1 of 2)

- Drinking water is the primary route of arsenic exposure.
- This study is based on mortality data associated with lung, bladder and liver cancers for the endemic region of Taiwan.
- Initial study done by Tseng et al. (1968) has generated great interest not only in Taiwan but worldwide.

Executive Summary (2 of 2)

- The purpose of the present study is to see the effect of various predictors like Gender, Cancer Type (Lung, Liver and Bladder), Arsenic Concentration (High, Medium and low), Age and Person Years on Mortality.

- The techniques used are Logistic Regression Model, Poisson Regression Model and Negative Binomial Regression Model.

- The effect of all the predictors on mortality have been found significant in the order of Age, Cancer Type, Person Year, Gender and Arsenic Concentration.

Introduction
(1 of 1)

- The most common source of human exposure to inorganic arsenic is through drinking water.
- There are many countries in the world where arsenic in drinking water is more than the recommended limit of 0.01 mg/L.
- The data for the present study is from a population of 42 coastal villages in six south-western townships of Taiwan.
- The purpose of the study is to see the effect of various predictors like Gender, Cancer Type (Lung, Liver and Bladder), Arsenic Concentration (High, Medium and low), Age and Person Years on Mortality.

Materials and Methods (1 of 4)

- The data used in this study is originally reported by Wu et. al (1989)[65].

1. LOGISTIC REGRESSION MODEL

- The best fitting model is evolved by dropping the most non significant predictors one by one sequentially.

- At every step the reduced model is compared with the previous model using chi-square.

- The Odds Ratio (OR) estimate is used to approximate how much more/less likely is the outcome (no death / death) with predictors having one categories compared to other category.

Materials and Methods (2 of 4)

2. POISSON REGRESSION MODEL

- Here the response has been used as a count data in the original form instead two categories.

- Adequacy of the fit is assessed by comparing mean and variance.

- If the mean and variance are very different then the fit is likely to be over/under dispersed.

Materials and Methods (3 of 4)

- Corrective measure in case of over dispersion includes running Negative Binomial Regression Model.

- Significance of parameters is assessed by chi-square and Wald confidence limit.

- Over/Under dispersion is tested with a likelihood ratio test based on the statistics of Poisson and Negative Binomial model.

Materials and Methods (4 of 4)

3. NEGATIVE BINOMIAL REGRESSION MODEL

- The adequacy, analysis, goodness of fit and test procedures are very similar to Poisson Regression Model.

- Here the log of the mean μ is a linear function of predictor variables as in the case of Poisson Regression Model.

- This basically relaxes the assumption about equality of mean and variance, which is a strict property of Poisson distribution.

RESULTS AND DISCUSSION (1 of 5)

1. LOGISTIC REGRESSION MODEL

- All the predictors were included in the multivariate logistic model considering their importance, though 3/5 did not meet the screening criteria of $p<0.25$.

- All the interaction and confounding effects were found non significant (Table 4.2.4.1 and 4.2.4.2).

- The final logistic model after testing of all the assumptions, data noise, collinearity, confounding, interaction and validation testing has been found to have left with one significant predictor namely Age.

RESULTS AND DISCUSSION (2 of 5)

- The findings based on Odds Ratio estimates suggest that (Table 4.2.5.1):

1. Those who are in the age group of 35−40 were 5.4 times more likely to have 1 or more deaths respectively, than those in the age group of 20−35.

2. Those who are in the age group of 40−50 were 27.8 times more likely to have 1 or more deaths respectively, than those in the age group of 20−35.

3. Those who are in the age group of >50 were 114.1 times more likely to have 1 or more deaths respectively, than those in the age group of 20−35.

- The non significance of other predictors except Age leads to further investigation based on actual death count since the non significant predictors are very important from medical perspectives.

RESULTS AND DISCUSSION (3 of 5)

2. POISSON REGRESSION MODEL

- The deviance (1.279) and Chi-square (1.436), > 1 suggests the possibility of over dispersion and hence inadequate fit (Table 4.3.1).

- The chi-square test of over dispersion of Poisson distribution is rejected at 5% level of significance and holds at 1% level of significance.

- This suggests the requirement of more data points or further investigation, which leads to fitting of Negative Binomial model.

RESULTS AND DISCUSSION (4 of 5)

- The overall conclusion is that both models are fine with this dataset.

- **Poisson Model:** All the five predictors were found significant (Table 4.3.3).

- The order of significance of predictors was Age, Cancer Type, Person Year, Gender, and Arsenic Concentration.

- This validates retaining of Age variable and dropping Person Year in the final logistic model earlier.

RESULTS AND DISCUSSION (5 of 5)

3. NEGATIVE BINOMIAL REGRESSION MODEL

- The assessment is similar to Poisson regression model.

- The values of deviance and Pearson chi-square for Negative Binomial model have substantially reduced from Poisson regression model. This suggests improvement and further validation of the model. (Table 4.4.1).

- The overall test of significance of all the predictors in Negative Binomial regression model remains same as in Poisson model though the p value varies (Table 4.4.3).

- The order of importance of predictors remains same as Poisson model (Table 4.4.3).

REFERENCES

Please see the book

?

QUESTIONS PLEASE

THANKS