

# Copula-based Bivariate Model for Association between Lung Disease and Skin Lesion

**Arabinda Das**

Department of Statistics  
Acharya Prafulla Chandra College  
Kolkata – 700131, West Bengal  
INDIA  
Email: arabinda@apccollege.ac.in

## ABSTRACT

*Chronic lung disease has been reported to occur from chronic arsenic exposure in several studies. Skin lesion has been considered as a hallmarked sign of arsenic exposure. Several studies also found the elevated prevalence of lung disease in presence of skin lesion. The present study examines the association of lung disease with skin lesion considering lung disease and skin lesion as binary and ordinal observations respectively. A bivariate limited dependent model is an appropriate statistical tool for this type of modeling. This paper develops a bivariate model specifying marginal regression models for the responses, and combines them via a copula to form a joint model. The commonly used Gaussian copula has been used to model the association. The maximum likelihood (ML) method has been used to estimate the model. The application shows a significant positive association between lung disease and skin lesion.*

**Keywords:** Bivariate logit, copula function, arsenic exposure

**AMS Classifications:** 62J02, 62F10, 62P10

## 1. INTRODUCTION

Arsenic is a group-I carcinogen and causes human health effects in many parts of the world including in the millions people of Bangladesh and part of India especially in the state of West Bengal. Arsenic exposure from drinking of contaminated ground water and resultant health effects are well documented in numerous studies reporting various health complications including skin manifestations, cardiovascular complications, chronic liver disease and cancer of various organs from chronic arsenic exposure. Characteristic skin lesion (pigmentation, depigmentation and keratosis) is most distinctive sign of chronic arsenic toxicity. Abnormalities in lung function are also known to occur following chronic arsenic exposure. High prevalence of respiratory illness among children with arsenical skin lesions was found in Chile (Zaldiver and Ghai, 1980). The prevalence of chronic bronchitis disease was observed in an elevated level with arsenic-induced skin lesions in a cross-sectional survey in Antofagasta (Borgono et al., 1977) and in Bangladesh (Milton et al., 2001). Symptoms of chronic cough, dyspnoea and evidences of chest signs were observed in elevated numbers in a cross-sectional study on a population of 7,683 from West Bengal (Guha Mazumder et al., 1997). Lung cancer has been reported to be associated with high levels of arsenic in different studies from Taiwan, (Chen and Wang, 1990) and Argentina, (Hopenhayn-Rich et al., 1998) and Chile (Smith et al., 2006; Marshall et al., 2007; Martinez et al., 2012).

We examined the association of skin lesion with respiratory health in a cross-sectional study with people of West Bengal, communities with exposure to inorganic arsenic through drinking water. A bivariate mixed model with binary and ordinal outcomes is developed to measure the association between skin lesion and lung disease where skin lesion was categorized into 4 categories of no skin lesion, mild, moderate and severe skin lesion and lung disease was categorized into two categories of absent and present of lung disease. Joint analysis of such outcomes is one of the most interesting facts in the literature of statistics in health. Such joint models are potentially advantageous in several statistical and practical respects. For example, a bivariate model enables analysts to account for relationships between outcomes and assess at the same time the joint influence of predictors/covariates on them.

Copula functions is a useful tool derive bivariate distribution with non-normal marginal distributions and various dependence structures using Sklar theorem (1959, 1973). Based on Sklar's theorem, there is a unique copula  $C_\alpha: [0,1]^2 \rightarrow [0,1]$  which satisfies  $F(x_1, x_2; \lambda) = C_\alpha(F_1(x_1; \theta_1), F_2(x_2; \theta_2))$  where  $F_i(x_i; \theta_i)$  is the distribution function of  $X_i$ ,  $\theta = (\theta_1, \theta_2)'$  and  $\lambda = (\theta, \alpha)'$ ,  $\alpha \in \Omega$ . The copula approach of modelling gives wide flexibility in modelling varied types of data. The main benefits of the model is that by considering a number of different copulas one can obtain, in a quasi-likelihood sense, a better approximation to the true underlying distribution. There are number of copula families representing wide range of dependence structure among the random variables.

The paper is organized as follows. We introduce a class of copula-based regression models for bivariate binary and ordinal outcomes in section 2 and discuss likelihood-based methods for model estimation. The model is estimated by maximum likelihood method. The data from arsenic affected area are described in the section 3. In section 4, data from arsenic affected area are used to illustrate our methodology. Section 5 concludes the paper.

## 2. REGRESSION MODEL FOR MIXED BIVARIATE NOMINAL AND ORDINAL OUTCOMES

This section develops a bivariate mixed model with binary and ordinal outcomes with lung disease and skin lesion considered as binary and ordinal variables respectively. Bhattacharya et al (2006) used bivariate normal distribution probit marginals for the binary dependent variables. Joint bivariate distribution with various dependence structures and non-normal marginal distributions have been used in recent years (see for details: Kolev and Paiva 2009; Song and Yuan, among others). More recently, Winkelmann (2009) proposed a bivariate probit model with copula function in modeling the effect of an endogenous binary regressor.

Consider two correlated mixed outcomes  $X_i$  and  $Y_i$  obtained from each of  $N$  subjects, where  $X_i$  is nominal and  $Y_i$  is ordinal. The underlying latent counterpart of  $X_i$  holds the relationship of  $X_i^* = z_{1i}'\beta_1 + \varepsilon_{1i}$  with

$$X_i = \begin{cases} 0, & \text{if } X_i^* \in (-\infty, \eta_1) \\ 1, & \text{if } X_i^* \in (\eta_1, \infty) \end{cases}$$

where  $z_{1i}$  denote explanatory variable,  $\beta_1$  is the corresponding parameter,  $\varepsilon_{1i}$  is the random noise and  $\eta_1$  is unknown threshold.

Further, let  $Y_i$  have  $(C+1)$  distinct values, say,  $1, 2, \dots, C$  representing ordinal scores with underlying latent counterpart  $Y_i^* = z_{2i}'\beta_2 + \varepsilon_{2i}$  such that

$$Y_i = \begin{cases} 1, & \text{if } Y_i^* \in (-\infty, \gamma_1) \\ \dots & \dots \\ k, & \text{if } Y_i^* \in (\gamma_k, \gamma_{k+1}) \\ \dots & \dots \\ C, & \text{if } Y_i^* \in (\gamma_C, \infty) \end{cases}$$

where  $z_{2i}$  denote explanatory variable and  $\beta_2$  is the corresponding parameter,  $\varepsilon_{2i}$  is the random noise and  $\gamma_1 < \gamma_2 < \dots < \gamma_C$  are unknown thresholds, with  $\gamma_0 = -\infty$  and  $\gamma_{C+1} = \infty$ .

Assuming  $X_i^* \sim F_{X_i^*}$  and  $Y_i^* \sim F_{Y_i^*}$  and the joint distribution function of  $(X_i^*, Y_i^*)$  as  $F_{X_i^* Y_i^*}(x_i^*, y_i^*)$ , the joint distribution of  $X_i$  and  $Y_i$  is then given by

$$\begin{aligned} P(X_i = 0, Y_i = k) &= P(X_i^* \leq \eta_1, \gamma_k \leq Y_i^* \leq \gamma_{k+1}) \\ &= P(X_i^* \leq \eta_1, Y_i^* \leq \gamma_{k+1}) - P(X_i^* \leq \eta_1, Y_i^* \leq \gamma_k) \\ &= F_{X_i^* Y_i^*}(\eta_1, \gamma_{k+1}) - F_{X_i^* Y_i^*}(\eta_1, \gamma_k) \end{aligned}$$

and

$$\begin{aligned} P(X_i = 1, Y_i = k) &= P(\eta_1 \leq X_i^*, \gamma_k \leq Y_i^* \leq \gamma_{k+1}) \\ &= P(X_i^* \leq \infty, Y_i^* \leq \gamma_{k+1}) - P(X_i^* \leq \infty, Y_i^* \leq \gamma_k) - P(X_i^* \leq \eta_1, Y_i^* \leq \gamma_{k+1}) + P(X_i^* \leq \eta_1, Y_i^* \leq \gamma_k) \\ &= F_{Y_i^*}(\gamma_{k+1}) - F_{Y_i^*}(\gamma_k) - F_{X_i^* Y_i^*}(\infty, \gamma_{k+1}) + F_{X_i^* Y_i^*}(\infty, \gamma_k) \\ &\quad - F_{X_i^* Y_i^*}(\eta_1, \gamma_{k+1}) + F_{X_i^* Y_i^*}(\eta_1, \gamma_k) \end{aligned}$$

Under the copula, the joint distribution function of  $X_i$  and  $Y_i$  with logit marginals is obtained as

$$F_{X_i^* Y_i^*}(x_i^*, y_i^*) = C_\alpha(\Lambda(x_i^*), \Lambda(y_i^*))$$

where  $C_\alpha(\cdot, \cdot)$  is the copula function,  $\Lambda(\cdot)$  is the distribution function of a logistic random variable and is given by  $\Lambda(z_i) = 1/(1 + \exp(-z_i))$ . We consider standard logistics as the marginals as it is based on logistic distribution and its cdf turns out to be mathematically convenient. Also, it is more popular in health sciences as the coefficients can be interpreted in terms of odds ratios.

Therefore, the joint density  $f_{X_i, Y_i}(x, y) = P(X_i = x, Y_i = y)$  of  $X_i$  and  $Y_i$  under the copula is as follows:

$$\begin{aligned} P(X_i = 0, Y_i = k) &= C_\alpha(\Lambda(\zeta_1), \Lambda(\xi_{k+1})) - C_\alpha(\Lambda(\zeta_1), \Lambda(\xi_k)) \\ P(X_i = 1, Y_i = k) &= C_\alpha(1, \Lambda(\xi_{k+1})) - C_\alpha(1, \Lambda(\xi_k)) - C_\alpha(\Lambda(\zeta_1), \Lambda(\xi_{k+1})) + C_\alpha(\Lambda(\zeta_1), \Lambda(\xi_k)) \end{aligned}$$

where

$$\zeta_1 = \eta_1 - z_1' \beta_1$$

$$\xi_i = \gamma_i - z_2' \beta_2$$

Then, we have

$$P(X_i = x, Y_i = y)$$

$$= \begin{cases} C_\alpha(\Lambda(\zeta_1), \Lambda(\xi_1)) & \dots & \text{if } x = 0, y = \gamma_0 \\ C_\alpha(\Lambda(\zeta_1), \Lambda(\xi_{k+1})) - C_\alpha(\Lambda(\zeta_1), \Lambda(\xi_k)) & \dots & \text{if } x = 0, y = \gamma_k \\ C_\alpha(\Lambda(\zeta_1), 1) - C_\alpha(\Lambda(\zeta_1), \Lambda(\xi_C)) & \dots & \text{if } x = 0, y = \gamma_C \\ C_\alpha(1, \Lambda(\xi_1)) - C_\alpha(\Lambda(\zeta_1), \Lambda(\xi_1)) & \dots & \text{if } x = 1, y = \gamma_0 \\ C_\alpha(1, \Lambda(\xi_{k+1})) - C_\alpha(1, \Lambda(\xi_k)) - C_\alpha(\Lambda(\zeta_1), \Lambda(\xi_{k+1})) + C_\alpha(\Lambda(\zeta_1), \Lambda(\xi_k)) & \dots & \text{if } x = 1, y = \gamma_k \\ 1 - C_\alpha(1, \Lambda(\xi_C)) - C_\alpha(\Lambda(\zeta_1), 1) + C_\alpha(\Lambda(\zeta_1), \Lambda(\xi_C)) & \dots & \text{if } x = 1, y = \gamma_C \end{cases}$$

We consider the more popular Gaussian copula, as this copula is comprehensive and allows for either negative or positive dependence. Under Gaussian copula,  $C_\alpha(u_1, u_2) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \alpha)$ , the joint density  $f_{X_i, Y_i}(x, y)$  becomes

$$\begin{aligned} P(X_i = 0, Y_i = k) &= \Phi_2(\Phi^{-1}(\Lambda(\zeta_1)), \Phi^{-1}(\Lambda(\xi_{k+1})); \alpha) - \Phi_2(\Phi^{-1}(\Lambda(\zeta_1)), \Phi^{-1}(\Lambda(\xi_k)); \alpha) \\ P(X_i = 1, Y_i = k) &= \Phi_2(\Phi^{-1}(1), \Phi^{-1}(\Lambda(\xi_{k+1})); \alpha) - \Phi_2(\Phi^{-1}(1), \Phi^{-1}(\Lambda(\xi_k)); \alpha) \\ &\quad - \Phi_2(\Phi^{-1}(\Lambda(\zeta_1)), \Phi^{-1}(\Lambda(\xi_{k+1})); \alpha) + \Phi_2(\Phi^{-1}(\Lambda(\zeta_1)), \Phi^{-1}(\Lambda(\xi_k)); \alpha) \end{aligned}$$

where  $\Phi$  and  $\Phi_2$  are the univariate and bivariate cdfs of the standard normal distribution, and  $\alpha$  is the coefficient of correlation.

These joint probabilities described in earlier section depend on the selected copula as well as parameter vector  $\lambda = (\beta_1, \beta_2, \alpha)'$ . If the true copula is assumed to belong to a parametric family, consistent and

asymptotically normally distributed estimates of the parameter  $\lambda = (\beta_1, \beta_2, \alpha)'$  can be obtained through maximum likelihood method.

Assuming an independent sample of  $n$  observations on  $(x_i, y_i, z_{1i}, z_{2i})$ , the likelihood function is given by

$$L(\lambda; x, y, z_1, z_2) = \prod_i \prod_k \{P(X_i = 1, Y_i = k / z_{1i}, z_{2i})^{x_i d_{ik}} \times P(X_i = 0, Y_i = k / z_{1i}, z_{2i})^{(1-x_i) d_{ik}}\}$$

where  $d_{ik}$  is a variable that equals one if  $y_i = k$ , and zero otherwise.

The corresponding log-likelihood function is then given by:

$$l(\lambda; x, y, z_1, z_2) = \sum_{i=1}^n \sum_{k=1}^c \{x_i d_{ik} \log P(X_i = 1, Y_i = k / z_{1i}, z_{2i}) + (1 - x_i) d_{ik} \log P(X_i = 0, Y_i = k / z_{1i}, z_{2i})\}$$

The maximum likelihood method is applied to the above log-likelihood function to estimate the model parameters using numerical optimization.

### 3. THE DATA DESCRIPTION

The present study describes dietary arsenic exposure and exposure from low level of arsenic (<50 µg/L) (the permissible limit in India) in drinking water and correlation between skin lesion and development of lung disease in 194 people living in arsenic endemic region in West Bengal. The participants were recruited from a population of 900 residents belonging to 212 households, in two arsenic-affected blocks of Chakdah and Haringhata in Nadia district of West Bengal. The households were selected by systematic sampling with a random start procedure after preparing a village map and listing of households. All subjects included in this study gave written consent for their participation. Approval of the study protocol was obtained from the Ethical Committee of the DNGM Research Foundation, fulfilling the Helsinki criteria and recommendation of the Indian Council of Medical Research, Government of India. Information was collected on demographic and social characteristics, occupation, and addiction from each participant. Addiction was assessed by taking into account the history of smoking, drinking and tobacco chewing in field visits. Water samples used for drinking and cooking purpose as well as diet and urine samples were collected from all 194 individual participants, and arsenic levels were estimated. Cumulative arsenic exposures were calculated as the concentration of arsenic in particular well water multiplied by duration of drinking of that water use. Arsenic concentration in urine and hair samples were measured using an atomic absorption spectrophotometer with a flow-injection hydride generation system (Perkin-Elmer A Analyst 400, Waltham, MA, USA (For details, Guha Mazumder et. al, 2012).

Presence of lung disease was ascertained by clinical examination and further abnormalities in lungs and pulmonary arteries were investigated using high-resolution computed tomography (HRCT) and echocardiography. HRCT was performed with 1.5-mm thin sections (Siemens Emotion 6, H-SP-CR, India) to participants reported chronic respiratory symptoms (persisting for at least more than 3 months), done independently by a radiologist and a pulmonologist (experienced in reading and interpreting HRCT chest) on a prescribed format consisting of different possible abnormalities and the distribution areas. Details of the procedures can be found in Bhattacharyya et al. (2014).

#### 4. APPLICATION OF THE MODEL

Table 1a presents median with IQR of age, gender, cumulative arsenic concentration, arsenic concentration in hair and urine in relation to skin lesion and p-values of Kruskal-Wallis test. The median age showed an increasing trend over severity of skin lesion though the differences are not significant ( $p\text{-value}>0.01$ ). Median cumulative arsenic concentration and arsenic concentration in hair had significant difference over skin lesion. However, arsenic concentration in urine had no significant relation to skin lesion. Table 1b presents median with IQR of age, gender, cumulative arsenic concentration, arsenic concentration in hair and urine in relation to lung disease and p values of Mann–Whitney test. Median age has no significant difference on presence of lung disease. Cumulative arsenic concentration and arsenic concentration in hair had significant elevated median values in presence of lung disease. However, there is no such significant difference in median arsenic concentration in urine ( $p\text{-value}>0.01$ ). Arsenic concentration in hair was found elevated with the severity of skin lesion. As a latent period is required to show skin lesion, the arsenic concentration in hair may be higher in these participants. Moreover, increased concentration of arsenic intake through water and diet with increasing skin lesion prevalence also supports this. However, there were no such differences found in arsenic concentration in urine.

Table 1a: Skin lesion in relation to Age, Gender, Cumulative arsenic exposure, Arsenic concentration in urine, Arsenic concentration in hair

|                                  |        | Skin lesion  |              |                 |              | P-value |
|----------------------------------|--------|--------------|--------------|-----------------|--------------|---------|
|                                  |        | No (N=91)    | Mild (N=67)  | Moderate (N=28) | Severe (N=8) |         |
| Age                              |        | 39.81±10.52  | 42.28±9.89   | 46.14±10.72     | 51.67±16.77  | 0.028   |
| Cumulative arsenic concentration |        | 3.44±2.75    | 4.93±3.06    | 6.35±2.26       | 12.83±11.16  | 0.000   |
| Arsenic concentration in urine   |        | 105.29±74.20 | 112.51±79.47 | 109.82±59.72    | 174.31±98.83 | 0.564   |
| Arsenic concentration in hair    |        | 1.001±0.43   | 1.22±0.45    | 1.69±0.70       | 1.99±0.52    | 0.000   |
| Gender                           | Male   | 57           | 40           | 20              | 5            | 0.685   |
|                                  | Female | 39           | 27           | 8               | 3            |         |

Table 1b: Lung disease in relation to Age, Gender, Cumulative arsenic exposure, Arsenic concentration in urine, Arsenic concentration in hair

|                                |        | Lung disease |              | P-value |
|--------------------------------|--------|--------------|--------------|---------|
|                                |        | No (N=153)   | Yes (N=41)   |         |
| Age                            |        | 41.46±10.59  | 42.90 ±10.86 | 0.410   |
| Cumulative arsenic exposure    |        | 4.38±3.42    | 5.06 ±2.96   | 0.143   |
| Arsenic concentration in urine |        | 112.18±77.88 | 99.49±59.85  | 0.570   |
| Arsenic concentration in hair  |        | 1.13± 0.47   | 1.45±0.69    | 0.010   |
| Gender                         | Male   | 88           | 120          | 0.016   |
|                                | Female | 65           | 74           |         |

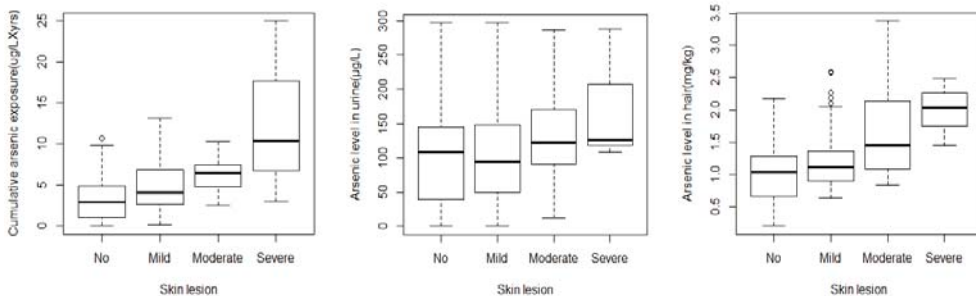


Fig. 1 Box plot of cumulative arsenic exposure, arsenic concentration in urine and arsenic concentration in hair in relation to skin lesion

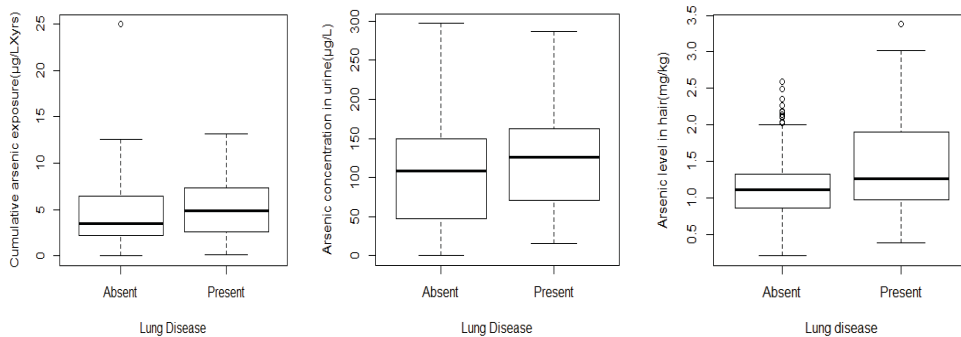


Fig. 2 Box plot of cumulative arsenic exposure, arsenic concentration in urine and arsenic concentration in hair in relation to lung disease

The copula-based bivariate binary-ordinal regression model was then constructed to find the association between two important outcomes of chronic arsenic exposure, skin lesion (ordinal) and lung disease (binary). The results of the model are presented in Table 2. The common covariates and possible risk factors that were taken into account are the participants' age and gender, cumulative arsenic concentration, arsenic concentrations in hair and arsenic concentrations in urine and smoking using the preliminary analysis. The significantly associated variables from the preliminary analysis were further considered for multivariate analysis. Smoking was included as a predictor of lung disease. We did a multivariate analysis using these variables as covariates in a joint response model. The results of our multivariate analysis are reported in Table 2. Odds ratios together with their corresponding standard errors and 95% confidence intervals are presented. A significant association between lung disease and skin lesion was found from the study. The results indicate that higher prevalence of lung disease was associated with male. Smokers had significantly higher prevalence of lung disease than non-smokers (OR = 2.438; 95% CI: 0.995–5.973). The odds of having lung disease was almost 3 fold (OR = 2.989; 95% CI: 1.524–5.861) from arsenic concentration in hair. Cumulative arsenic exposure had higher odds (OR = 1.358; 95% CI: 0.893–1.967) for having lung disease. The odds of having lung disease was higher though not significant (OR = 1.286; 95% CI: 0.990–1.956) from arsenic concentration in urine. Male were associated with higher rates of skin lesion (OR = 1.187; 95% CI: 0.753–4.223) which is not significant. The odds of presence of skin lesion was almost double (OR = 1.924; 95% CI: 1.224–3.066) from arsenic concentration in hair. Cumulative arsenic exposure had higher odds (OR = 1.279; 95% CI:

0.864–1.943) for having skin lesion. A significant association had been found between lung disease and skin lesion (OR = 2.367; CI: 1.127–5.034).

**Table 2:** Estimates of Copula-based Bivariate Probit Model  
Response variables: Lung disease & Skin lesion

| Predictor                      | OR    | SE    | P value | OR             |
|--------------------------------|-------|-------|---------|----------------|
| Equation for Lung Disease      |       |       |         |                |
| Age                            | 0.983 | 0.086 | 0.083   | (0.946, 1.023) |
| Gender                         | 1.778 | 0.749 | 0.139   | (0.686, 4.612) |
| Smoking                        | 2.438 | 0.754 | 0.001   | (0.995, 5.973) |
| Cumulative arsenic exposure    | 1.358 | 0.056 | 0.009   | (0.893, 1.967) |
| Arsenic concentration in hair  | 2.989 | 0.176 | 0.000   | (1.524, 5.861) |
| Arsenic concentration in urine | 1.286 | 0.115 | 0.041   | (0.990, 1.956) |
| Equation for Skin lesion       |       |       |         |                |
| Age                            | 0.995 | 0.016 | 0.298   | (0.912, 1.281) |
| Gender                         | 1.187 | 0.338 | 0.191   | (0.753, 2.186) |
| Cumulative arsenic exposure    | 1.279 | 0.433 | 0.019   | (0.864, 1.743) |
| Arsenic concentration in hair  | 1.974 | 0.176 | 0.000   | (1.124, 3.066) |
| Arsenic concentration in urine | 0.963 | 0.153 | 0.128   | (0.674, 1.358) |
| Disturbance correlation        |       |       |         |                |
| Correlation ( $\alpha$ )       | 2.367 | 0.525 | 0.003   | (1.127, 5.034) |

## 5. DISCUSSION AND CONCLUSION

The association of low level arsenic exposure to restrictive lung disease was established in a recent meta-analysis (Sanchez et al., 2018). Increasing arsenic exposure is associated with coughing and breathing problems (Sanchez et al., 2016). A positive association was found between arsenic exposure and need to stop for breath and cough (Powers et al., 2019). Many studies found the association of arsenic exposure with chronic bronchitis (Dauphine et al., 2011; von Ehrenstein et al., 2019; Milton et al., 2002). Researchers found a significant association between arsenic exposure and chronic bronchitis (Milton & Rahman, 2002; Milton et al., 2003; Guha Mazumder et al., 2005) with high arsenic exposure. In this study we figured the magnitude and sign of association between skin lesion and lung disease in chronic arsenic exposure through a bivariate binary and ordinal model using copulas. We found a significant positive association between skin lesion and lung disease in chronic arsenic exposure. This quantified the association between two important biomarkers from chronic arsenic exposure. The result found that age was not significant covariate of skin lesion and lung disease. However, male showed mildly higher prevalence of developing skin lesion and lung disease. These findings can be associated with the social structure of the society under study area as male generally stay outside for long time and female stay at home. Consuming drinking contaminated water outside made male participants more prone to develop skin lesion and lung disease from arsenic exposure. The result also found that cumulative arsenic exposure found to be highly associated with skin lesion and lung disease. Prevalence of skin lesion and lung disease increased with arsenic concentration in drinking water and the duration of use such water. Arsenic concentration in hair has been considered a chronic arsenic exposure. Prevalence of both skin lesion and lung disease increased with high arsenic concentration in hair i.e. chronic arsenic exposure is highly associated with skin lesion and lung disease. However, arsenic concentration in urine, a recent arsenic exposure biomarker was neither a significant

covariate of skin lesion nor a significant cofactor of lung disease i.e. recent arsenic exposure does not show any relation with skin lesion and lung disease. Most significantly, the result found a moderate but significant positive association between skin lesion and lung disease.

### Acknowledgments

The author would like to thank DNGM Research Foundation for providing the data.

### 6. REFERENCES

- Bhattacharya, J., Goldman, D., McCaffrey D., 2006, Estimating probit models with self-selected treatments. *Statistics in Medicine*. **25**:389–413.
- Bhattacharyya, P., Sen, P., Ghosh, A., Saha, C., Bhattacharya, P.P., Das, A., Majumdar, K., Guha Mazumder, D.N., 2014, Chronic lung disease and detection of pulmonary artery dilatation in high resolution computerized tomography of chest in chronic arsenic exposure. *Journal of Environmental Science and Health, Part A: Toxic/Hazardous Substances and Environmental Engineering*, **49**:13, 1453–1461.
- Borgono, J.M., Vicent, P., Venturino, H., Infante, A., 1977, Arsenic in the drinking water of the city of Antofagasta: epidemiological and clinical study before and after the installation of a treatment plant. *Environ. Health Perspect.* **19**, 103–105.
- Chen, C.J., Wang, C.J., 1990, Ecological correlation between arsenic level in well water and age-adjusted mortality from malignant neoplasms. *Cancer Res.* **50**, 5470–5474.
- Chen, S., Zhou, Y., 2007, Estimating a generalized correlation coefficient for a generalized bivariate probit model. *Journal of Econometrics*. **141**, 1100–1114.
- Dauphine, D.C., Ferreccio, C., Guntur, S., Yuan, Y., Hammond, S.K., Balmes, J., Smith, A.H., Steinmaus, C., 2011, Lung function in adults following in utero and childhood exposure to arsenic in drinking water: preliminary findings. *Int Arch Occup Environ Health*. **84**(6):591–600.
- Guha Mazumder, D.N., Haque, R., Ghosh, N., De, B.K., Santra, A., Chakraborty, D., Smith, A.H., 2000, Arsenic in drinking water and the prevalence of respiratory effects in West Bengal, India. *Int. J. of Epidemiol.* **29**, 1047–1052.
- Guha Mazumder, D.N., Kar-Purkayastha, I., Ghose, A., Mistry, G., Saha, C., Nandy, A.K., Das, A., Majumder, K.K., 2012, Hypertension in chronic arsenic exposure: A case control study in West Bengal. *Journal of Environmental Science and Health, Part A: Toxic/Hazardous Substances and Environmental Engineering*. **47**, 1514–1520.
- Guha Mazumder, D.N., Steinmaus, C., Bhattacharya, P., von Ehrenstein, O.S., Ghosh, N., Gotway, M., Sil, A., Balmes, J.R., Haque, R., Hira-Smith, M.M., Smith, A.H., 2005, Bronchiectasis in persons with skin lesions resulting from Arsenic in drinking water. *Epidemiology*. **16**, 760–765.
- Hopenhayn-Rich, C., Biggs, M.L., Smith, A.H., 1998, Lung and kidney cancer mortality associated with arsenic in drinking water in Cordoba, Argentina. *Int. J. Epidemiol.* **27**, 561–569.
- International Agency of Research on Cancer. Some drinking-water disinfectants and contaminants, including arsenic. In Monographs on the Evaluation of Carcinogenic risks to Humans. World Health Organization: Lyon, France, 2004; 84, 271–441.
- Kolev, N., Paiva, D., 2009, Copula-based Regression Models: A Survey. *Journal of Statistical Planning and Inference*. **139**, 3847–3856.

- Marshall, G., Ferreccio, C., Yuan, Y., Bates, M.N., Steinmaus, C., Selvin, S., Liaw, J., Smith, A.H., 2007, Fifty-year study of lung and bladder cancer mortality in Chile related to arsenic in drinking water. *J. Natl. Cancer Inst.* **99**, 920–928.
- Martinez, V.D., Vucic, E.A., Lam, S., Lam, W.L., 2012, Arsenic and lung cancer in never-smokers: lessons from Chile. *Amer. J. Respir. Crit. Care Med.* **185**(10), 1131–1132.
- Milton, A.H., Hasan, Z., Rahman, A., Rahman, M., 2001, Chronic arsenic poisoning and respiratory effects in Bangladesh. *J. Occup. Health.* **43**, 136–140.
- Milton, A.H., Rahman, M., 2002, Respiratory effects and Arsenic contaminated well water in Bangladesh. *Int. J. Environ. Health Res.* **12**(2), 175–179.
- Milton, A.H., Hasan, Z., Rahman, A., Rahman, M., 2003, Non-cancer effects of chronic arsenicosis in Bangladesh: Preliminary results. *J. Environ. Sci. Health, Part A: Tox. Hazard Subst. Environ. Eng.* **38**, 301–305.
- Powers, M., Sanchez, T.R., Grau-Perez, M., Yeh, F., Francesconi, K.A., Goessler, W., George, C.M., Heaney, C., Best, L.G., Umans, J.G., Brown, R.H., Navas-Acien, A., 2019, Low-moderate arsenic exposure and respiratory in American Indian communities in the Strong Heart Study. *Environmental Health.* **18** (104). <https://doi.org/10.1186/s12940-019-0539-6>
- Sanchez, T.R., Powers, M., Perzanowski, M., George, C.M., Graziano, J.H., Navas-Acien, A., 2018, A meta-analysis of arsenic exposure and lung function: is there evidence of restrictive or obstructive lung disease? *Curr Environ Health Rep.* **5**(2), 244–254.
- Sanchez, T.R., Perzanowski, M., Graziano, J.H., 2016, Inorganic arsenic and respiratory health, from early life exposure to sex-specific effects: A systematic review. *Environ Res.* **147**, 537–555.
- Smith, A.H., Marshall, G., Yuan, Y., Ferreccio, C., Liaw, J., von Ehrenstein, O., Steinmaus, C., Bates, M.N., Selvin, S., 2006, Increased mortality from lung cancer and bronchiectasis in young adults after exposure to arsenic in utero and in early childhood. *Environ. Health Perspect.* **114**, 1293–1296.
- Song, P.X.K., Li, M., Yuan, Y., 2009, Joint Regression Analysis of Correlated Data using Gaussian Copula. *Biometrics*, **65**, 60–68.
- von Ehrenstein, O.S., Guha Mazumder, D.N., Yuan, Y., Samanta, S., Balmes, J., Sil, A., Ghosh, N., Hira-Smith, M., Haque, R., Purushothamam, R., Lahiri, S., Das, S., Smith, A.H., 2005, Decrements in lung function related to arsenic in drinking water in West Bengal, India. *Am J Epidemiol.* **162**(6):533–541.
- Winkelmann, R., 2012, Copula Bivariate Probit Models: With an Application to Medical Expenditures. *Health Economics*, **21**(12), 1444–1455.
- Zaldivar, R., Ghai, G.L., 1980, Clinical epidemiological studies on endemic chronic arsenic poisoning in children and adults, including observations on children with high- and low-intake of dietary arsenic. *Zentralbl Bakteriol [B]*. **170**, 409–421.