

Correlational Analysis of Some Infectious Diseases

Ogoke Uchenna Petronilla¹ and Mbachu Hope Ifeyinwa²

¹Department of Mathematics and Statistics, University of Port Harcourt, Rivers State Nigeria and ² Department of Statistics, Imo State University, Owerri.

Corresponding Author: Ogoke Uchenna Petronilla¹

Abstract: In this study, data were obtained from 684 subjects male and female inclusive, on the spread of five different types of bacteria infectious diseases which are *Staphylococcus*, *Pseudomonas*, *Klebsiella pneumonia*, *Escherichia Coli* and *Proteus*. The data were collected according to their age range which are 10–20yrs, 21–40yrs, 41–60yrs, 61–80yrs and 80yrs and above. Results from the correlational analysis carried out on this data set indicate that *Klebsiella pneumonia* and *Escherichia Coli* has an almost perfect positive relationship with correlation coefficient of 0.976 followed by *Staphylococcus* and *Pseudomonas*, *Staphylococcus* and *Klebsiella pneumonia*, *Staphylococcus* and *Escherichia Coli* which has a substantial positive correlation with correlation coefficient of 0.799, 0.701 and 0.702 respectively. Correlation coefficient between *Klebsiella pneumonia* and *Proteus*, *Klebsiella pneumonia* and *Pseudomonas*, *Escherichia Coli* and *Proteus* indicate that a slight positive relationship exist between them, while that of *Pseudomonas* and *Escherichia Coli* indicates that a fair positive relationship exist between them. For *Pseudomonas* and *Proteus*, *Staphylococcus* and *Proteus*, their correlation coefficient indicates that a negative correlation exists between them.

Date of Submission: 25-01-2018

Date of acceptance: 16-02-2018

I. Introduction

A Disease is a particular abnormal condition, a disorder of a structure or function that affect part or all of an organism. The study of disease is called Pathology which includes the study of cause. Disease is often constructed as a medical condition associated with specific symptoms and signs. It may be caused by external factors such as pathogens, or it may be caused by internal dysfunction particularly of the immune system such as an Immunodeficiency, or a Hypersensitivity. Infection is the invasion of an organism's body tissues by disease-causing agent, their multiplication, and the reaction of host tissues to these organisms and the toxins they produce. Infectious diseases are disorders caused by organisms such as Bacteria, Viruses, Fungi or Parasites. Many organisms live in an on our bodies they are normally harmless or even helpful, but under certain conditions, some organisms may cause disease. Some infectious diseases can be passed from person to person. Some are transmitted by bites from insect or animals. And others are acquired by ingesting contaminated food or water, or being exposed to organisms in the environment. Historical records and research works on infectious diseases exist.

Iraet al (1988) did a study on Statistical Inference for infectious diseases: risk-specific household and community transmission parameters. A statistical model is presented for the analysis of infectious disease data from family studies in the community. The model partitions the sources of infection into those from within the household and those from the community at large. The parameters reflecting these sources of infection are estimated as functions of the risk factors. This new model is used to overcome problems associated with the lack of independence of observations in infectious disease data and negative confounding due to the association of unmeasured exposures and immunity. An example of how this new statistical model is used to provide a clearer and less confounded description of risk factor effects is presented for data from influenza A(H3N2) epidemic seasons in the Tecumseh Respiratory Illness Study. The risk factors examined are age and pre-epidemic season antibody level as measured by the hemagglutination-inhibition test, while the outcome is the infection rate. A standard analysis of the data indicates that the efficacy of protective antibodies is 70% in children and only 47% in adults. However, such an efficacy measurement is negatively confounded by past exposure which is age dependent. By means of the model, the true, unconfounded, efficacy of protective antibodies is shown to be 90% in both adults and children.

Tom and Federica (2014) did a study on statistical inference for infectious diseases using the general stochastic epidemic model for the spread of infectious diseases. They estimated the method for inferring model parameters such as the basic reproduction number R_0 and vaccination coverage V_C assuming different types of data from an outbreak. They concluded by saying that one of the most urgent problem of infectious disease epidemiology over the last decades has been to quickly learn about new diseases.

Emanuele et al (2011) did a study on Joint Modeling of HCV and HIV Co-Infection among Injecting Drug Users in Italy and Spain Using Individual Cross-Sectional Data. The aim of the analysis is to study co-infection with hepatitis C virus (HCV) and human immunodeficiency virus (HIV) in injecting drug users (IDUs) using a joint modeling approach that makes use of multivariate statistical methods for current status data. Using marginal models, they estimate association measures between HCV and HIV infections at individual level, i.e., odds ratios and correlation coefficients, and they regress them against some risk factors, e.g., the length of the injecting career, the age at first injection, the ever sharing of syringes, and the frequency of current injecting. In addition, they fit random-effects models that take into account the individual heterogeneity in the acquisition of the infections. For their analysis, they use cross-sectional data from two independent serological surveys, one carried out in Italy (IT) in 2005 on 856 subjects, and the other in three Spanish (ES) cities, between 2001 and 2003, on 589 subjects. They found that the infections are positively associated within individuals, e.g., in Italy OR=2.56 with 95% confidence interval (CI) (1.43, 6.68) and in Spain OR= 2.42, with 95% CI (1.41, 4.30). They found that the odds ratio and the correlation between HCV and HIV infections increase positively with the length of the injecting career. Moreover, they are found to be significantly positive in case IDUs have never shared syringes or report low injecting frequencies. The variance of the individual random effects is positive, e.g., $\text{Var}(b)=0.34$ (0.14, 0.62), indicating that there is significant individual heterogeneity in the acquisition of the infections. These results show that a significant association between HCV and HIV infections within IDUs is related to significant individual heterogeneity in the acquisition of the infections. Indeed, the association between these infections in IDUs who report ever sharing syringes is not significant, which can be explained by a higher homogeneity in their behaviors and, therefore, in their acquisition of the infections. This research objective is to assess the level of relationship between some types of bacteria infectious diseases and to assess the level of reliability between these bacteria infectious diseases to know if the correlation between this bacteria infectious diseases are slight, fair, moderate, substantial and almost perfect.

II. Method And Analysis

Data were collected from 684 subjects, male and female inclusive on their exposure to these five various kind of bacteria infectious diseases which are; Staphylococcus, Pseudomonas, Klebsiella pneumonia, Escherichia Coli and Proteus. The data were collected according to their age range which are 10–20yrs, 21–40yrs, 41–60yrs, 61–80yrs and 80yrs and above. The Pearson correlation co-efficient was calculated to characterized the various kind of bacteria infectious diseases and the level of reliability of the Pearson correlation will be classified using the characterizations reported by Landis and Koch (1977). These characterizations range from 0.00 to 0.20 (Slight), 0.21 to 0.40 (Fair), 0.41 to 0.60 (moderate), 0.61 to 0.80 (substantial) 0.81 to 1.00 (almost perfect).

Data Presentation

The data presented in table 1 below shows the distribution by age of the five various kind of bacteria infectious diseases consider in this study.

TABLE 1: Distribution of Bacteria by Age

Age (years)	Organism				
	Staphylococcus	Pseudomonas	Klebsiella pneumonia	Escherichia Coli	Proteus
10 – 20	5 (0.90%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
21 – 40	100 (17.10%)	10 (20.8%)	4 (13.30%)	0 (0%)	0 (0%)
41 – 60	201 (34.40%)	22 (45.80%)	6 (20%)	4 (26.7%)	0 (0%)
61 – 80	170 (29.10%)	8 (16.70%)	18 (60%)	10 (66.70%)	2 (28.60%)
≥ 80	28 (4.80%)	8 (16.70%)	2 (6.70%)	1 (6.70%)	5 (71.40%)

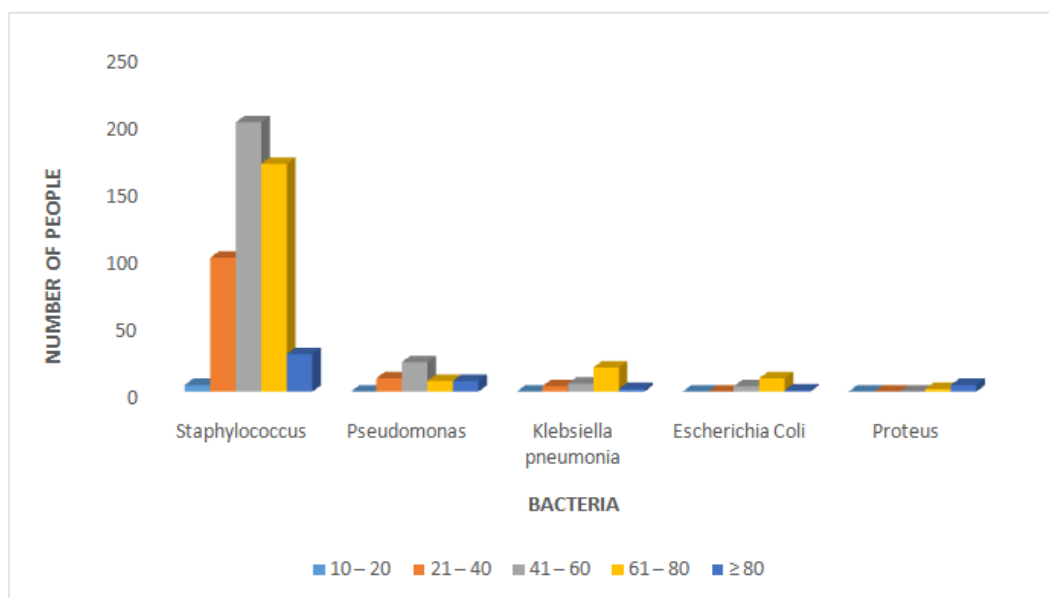


Figure 1 above is a bar chart showing the amount of spread of each bacteria infectious diseases across each age range.

Table 2: Correlation Results

		Staphylococcus	Pseudomonas	Klebsiella pneumonia	Escherichia Coli	Proteus
Staphylococcus	Pearson Correlation	1	0.799	0.701	0.702	-0.301
	Sig. (2-tailed)		0.105	0.187	0.186	0.623
	N	5	5	5	5	5
Pseudomonas	Pearson Correlation	0.799	1	0.196	0.238	-0.161
	Sig. (2-tailed)	0.105		0.752	0.700	0.796
	N	5	5	5	5	5
Klebsiella pneumonia	Pearson Correlation	0.701	0.196	1	0.967**	0.065
	Sig. (2-tailed)	0.187	0.752		0.007	0.918
	N	5	5	5	5	5
Escherichia Coli	Pearson Correlation	0.702	0.238	0.967**	1	0.108
	Sig. (2-tailed)	0.186	0.700	0.007		0.863
	N	5	5	5	5	5
Proteus	Pearson Correlation	-0.301	-0.161	0.065	0.108	1
	Sig. (2-tailed)	0.623	0.796	0.918	0.863	
	N	5	5	5	5	5

** . Correlation is significant at the 0.01 level (2-tailed).

Table 2 above shows the correlation relationship between the various kinds of bacteria infectious diseases.

Table 3: Level of Reliability of Correlation Coefficients.

Bacteria	Correlation Coefficient	Level of Reliability
Staphylococcus/ Pseudomonas	0.799	Substantial/ Positive
Staphylococcus/ Klebsiella pneumonia	0.701	Substantial/ Positive
Staphylococcus/ Escherichia Coli	0.702	Substantial/ Positive
Staphylococcus/ Proteus	-0.301	Moderate/ Negative
Klebsiella pneumonia/ Escherichia Coli	0.967	Almost Perfect/Positive
Klebsiella pneumonia/ Proteus	0.065	Slight/ Positive
Pseudomonas/ Escherichia Coli	0.238	Fair/ Positive
Pseudomonas/ Proteus	-0.161	Slight/ Negative
Klebsiella pneumonia/ Pseudomonas	0.196	Slight/ Positive
Escherichia Coli/ Proteus	0.108	Slight/ Positive

Table 3 shows the level of reliability between the various kinds of bacteria infectious diseases.

III. Discussion of Result

Figure 1 above shows descriptive statistics of the amount of spread of each bacteria infectious diseases across each age range. Results from the analysis shows that, Staphylococcus has its highest amount of spread between age range 41-60yrs, followed by age range 61-80yrs, and has its lowest amount of spread between age range 10-20yrs. Pseudomonas has its highest amount of spread between age range 41-60yrs followed by age range 21-40yrs, and its lowest amount of spread between age range 10-20yrs. Age range 61-80yrs and 80yrs and above is seen to have the same amount of spread of Pseudomonas. Klebsiella pneumonia is seen to have its highest amount of spread between age range 61-80yrs followed by age range 41-60yrs and have its lowest amount of spread between age range 10-20yrs. Escherichia Coli has its highest amount of spread between age range 61-80yrs, followed by age range 41-60yrs. While age range 10-20yrs and 21-40yrs are seen to have the lowest amount of spread Escherichia Coli with equal points. Proteus is seen to have its highest amount of spread between age range 80yrs and above, followed by age range 61-80yrs. While age range 10-20yrs, 21-40yrs and 41-60yrs are seen to have the lowest amount of spread of Proteus with equal points. Table 3 shows the level of Level of Reliability between the various kinds of bacteria. The correlation coefficient of Staphylococcus and Pseudomonas, Staphylococcus and Escherichia Coli, Staphylococcus and Klebsiella pneumonia indicates that there is a strong positive relationship between these bacteria which implies that an increase (decrease) in Staphylococcus implies an increase (decrease) in Pseudomonas, Escherichia Coli, Klebsiella pneumonia. The correlation coefficient between Staphylococcus and Proteus indicates that there is a negative moderate relationship existing between them which implies that an increase (decrease) in Staphylococcus implies a moderate decrease (increase) in Proteus. The correlation coefficient between Klebsiella pneumonia and Escherichia Coli indicates that there is a strong positive relationship existing between the both of them. Which implies that an increase (decrease) in Klebsiella pneumonia implies a strong increase (decrease) in Escherichia Coli. The correlation coefficient between Klebsiella pneumonia and Proteus, Klebsiella pneumonia and Pseudomonas, Escherichia Coli and Proteus indicate that a slight positive relationship exist among them. Which implies that an increase (decrease) in Klebsiella pneumonia implies a slight increase (decrease) in Pseudomonas and Proteus, and an increase (decrease) in Escherichia Coli implies a slight increase (decrease) in Proteus. The correlation coefficient between Pseudomonas and Escherichia Coli indicate that a fair positive relationship exist among them. Which implies that an increase (decrease) in Pseudomonas implies a fair increase (decrease) in Escherichia Coli. The correlation coefficient between Pseudomonas and Proteus indicates that a slight negative relationship exist among them. Which implies that an increase (decrease) in Pseudomonas implies a slight decrease (increase) in Proteus.

IV. Conclusion

Result from the analysis shows that positive correlation relationship exist among most of the bacteria except for Staphylococcus and Proteus, Pseudomonas and Proteus which indicates a negative correlation relationship.

References

- [1]. Emanuele Del Fava, Ziv Shkedy, Niel Hens, Marc Aerts, Barbara Suligoi, Laura Canoni, Fernando Vallejo (2011): Joint Modeling of HCV and HIV Co-Infection Among Injecting Drug User In Italy And Spain Using Individual Cross-Sectional Data.
- [2]. Statistical Communication of Infectious Diseases. Vol.3 (1).
- [3]. Ira M. Longini Jr, James S. Koopman, Michael Haber, George A. Cotsonis (1988): Statistical Inference of Infectious Diseases: Risk-Specific Household and Community Transmission Parameters. American Journal of Epidemiology. Vol.128(4) 845-859.
- [4]. Landis J.R. and Koch G.G. (1977): The measurement of observer agreement for categorical data.
- [5]. Biometrics. 33:159-174.
- [6]. Tom Britton and Federica Giardina (2014): Introduction to Statistical Inference For Infectious Diseases. A Research Report. 3.

Ogoke U. P and Mbachu, H.I "Correlational Analysis of Someinfectious Diseases." IOSR Journal of Mathematics (IOSR-JM) 14.1 (2018):PP54-57.