Journal of Mathematics and Statistics 5 (3):146-151, 2009
ISSN 1549-3644


# Statistical Analysis of the Risk Factors of the Major Epidemic Disease Among Residents of Sekondi-Takoradi Metropolis (STMA)

A.J. Turkson and X.J. Wang
Department of Mathematics, Huazhong University of Science and Technology,
430074 Wuhan-Hubei, China

**Abstract: Problem statement:** This study focused on statistical analysis of risk factors of the major disease among residents of STMA. The objectives of the study were to find the predominant diseases affecting residents in the catchments area. The study was also meant to identify risk factors associated with the incidence of the predominant disease. **Approach:** Purposive sampling technique was used to select residents from the target population. A structured questionnaire was used to obtain information from the sampling units. The data was analyzed using measures of risk, principal component analysis and chi-square tests. **Results:** Malaria (68%) topped the list of predominant diseases. It was also found that (44%) of those who had been diagnosed of malaria disease lived around swampy areas. The absolute risk value (78%) suggested that people living around the swampy area were at risk of infection by malaria parasites. The absolute risk value of 62% for dwellers of non-swampy areas indicated that aside the environmental factors there are other factors which were associated with the incidence of malaria. The principal component analysis results showed that; diets, social and educational status affect the incidence of malaria. The p-value of the chi-square test on two risk factors (environment and ones family history) were also significant, indicating that there was an association between the incidence of malaria in the catchments area and the identified risk factors. This result confirmed earlier researches that sought to investigate the relationship between diseases and environmental factors. **Conclusion/Recommendations:** It was revealed that apart from environmental factors and family history; diets, social status and educational background play an important role in the health of the residents. It was recommended that STMA takes the appropriate measures to prevent 6 people from the risk factor at any case.

**Key words:** Risk, swampy, environment, malaria

## INTRODUCTION

For centuries, diseases have been the main cause of death around the world[1]. Life expectancy has reduced by uncontrolled epidemics. By the third millennium, diseases appeared sweeping the entire globe, with an increasing trend in developing countries. Most diseases are the result of more than one factor and determining the specific cause of a disease can be difficult. Nonetheless, given the extensive use of individual level health data by national policy makers, it has become important to understand the factors that affect one's well being[6]. The report of the Ghana Statistical Service (GSS), asserted that over 80% of common diseases such as malaria, diarrhea, cholera and hook worm are a result of poor environmental hygiene. One wonders why these diseases are still common and fatal in spite of the availability of highly potent drugs, vaccinations, methods of prevention, health educations and clean up exercises organized periodically. Some assertions have been made to the possible risk factors, prevalent among them are:

Table 1: No. of residents who have been diagnosed of the predominant diseases in the Catchments area

| Type of disease | No. diagnosed | Percentage of residents diagnosed |
|---|---|---|
| Malaria | 230 | 68.0 |
| Asthma | 23 | 6.8 |
| Stroke | 19 | 5.6 |
| Hypertension | 19 | 5.6 |
| Anaemia | 16 | 4.7 |
| Diarrhoea | 15 | 4.4 |
| Tuberculosis | 13 | 3.9 |
| Diabetes | 9 | 2.7 |
| Heart Disease | 8 | 2.4 |
| Cancer | 5 | 1.5 |
| Pneumonia | 5 | 1.5 |
| Hepatitis | 2 | 0.6 |
| None of the above | 41 | 13.9 |

**Corresponding Author:** A.J. Turkson, Department of Mathematics, Huazhong University of Science and Technology, 430074 Wuhan-Hubei, China Tel: 008613477073279

**Age:** Older people suffer more than any other age group, thus the older one becomes the more he is prone to various diseases[8].

**Educational background:** Illiteracy is a disease, thus ones level of education affects his lifestyle and how he manages his environment. Education provides people with the knowledge and skills that can lead to a better quality of life.

**Unemployment:** The unemployed is poor and thus does not on his own volition get the means to subscribe to the cost of good health care which might be high[8].

**Family history:** Some diseases are transmitted from parents to the children through the blood, such sicknesses are hereditary. The GSS report affirmed that there is a moderate relationship between the anemia status of mothers and children.

**Environmental conditions:** A report submitted by STMA indicated that, a good proportion of houses in the Metropolis do not have toilet facilities; it further stated that there are indiscriminate defecations and urination[3].

**Lack of exercises:** Exercises are essential for the shedding of weights which goes a long way to prevent heart diseases and obese [8,].

**Brief survey of the Sekondi-Takoradi Metropolis (STMA)-A district in the western region of Ghana:** The Metropolis is the richest in the region; it abounds in natural resources like oil, timber, gold, cocoa, bauxite, manganese and rubber.

**Characteristic of the metropolis:** STMA is a densely populated district. It also has the largest share (19.2%) of the region's total population of 1,924,577. The central part of the Metropolis is low lying and occupied by muddy lagoons, it also has an equatorial type of climate. Temperatures are high with an average of 22° Celsius. It has a mean annual rainfall of 2.350 mm, which is experienced heavily in May and June with the minor rains occurring between September and October.

**Waste disposal and management:** The collection, transportation and disposal of solid and liquid waste are the sole responsibility of the STMA. The tonnage of waste generated in the Metropolis has been increasing over the years as the population increases. The actual collection in 2005 was 58,690 tonnes, a collection rate of about 70%. The Metropolis has a high proportion of organic waste and low proportions of recyclables[3].

**Objectives of the study:** The objectives of the study are to:

- Assess the kind of diseases predominant among the residents in the metropolis
- Assess the risk factors that contribute to the predominant disease in the metropolis
- Examine the effects of the environment, social and other health hazards on the health of the residents

## MATERIALS AND METHODS

**Data collection techniques:** The target population was all the people living in the STMA at the time the research was conducted. Both adults and children were included in the study. Data was collected from five general hospitals, two educational institutions and few selected communities within the area. A structured questionnaire was designed to elicit personal and basic information relevant to the topic. Interview method was used to obtain the needed information from the respondents.

**Sampling techniques:** A combination of techniques was used in obtaining the sample. Purposive sampling technique was used to select the various strata. The population was stratified according to those who were on admission at the hospital, those at the Out-Patient Department (OPD) waiting to be attended to, those who at the time of the research were not sick, those living in the rural areas and those living in the urban and semi-urban areas. Cluster sampling technique was used to select all the people found in two of the strata (those on admission and those at the OPD); finally, quota sampling was employed in selecting the required samples from the remaining strata

**Review of method:** The major statistical tools used in analyzing the data were Measures of risk[4,7], Chi-square test for independence[2] and principal component analysis[5,10].

**Measures of risk:** The relative risk is given by the formula:

$$\text{The relative risk in a study} = \frac{\text{Incidence rate among risk group}}{\text{Incidence rate among non} - \text{risk group}}$$

A relative risk greater than one is clinically significant.

Attributable risk allows us to attribute the difference in the incidences of a disease to a particular risk factor; it is calculated by finding the difference between the incidence rates of those exposed against those not exposed to the risk factor.

Odds are defined as the probability of experiencing an outcome divided by the probability of not experiencing the outcome. An Odds Ratio can be computed by dividing the odds of people exposed to the risk by the people without the risk.

**Chi-square test of independence:** The test deals with two variables, each of which is classified into a number of mutually exclusive classes arranged in a two-way table known simply as a contingency tables. This test essentially tests the independence of two variables. If the two variables are independent then the tests should be significant, thus if the test is significant, we conclude that the two variables are associated[2].

**Test statistics and degrees of freedom:** The test statistic is the chi-square ($\chi^2$) distribution given by:

$$\chi^2 = \frac{\sum\sum(O-E)^2}{E}$$

Where:
O = The observed frequency
E = The expected frequency for a particular category

Large values for $\chi^2$ indicate that there is a large discrepancy between the observed and the expected frequencies which may warrant rejection of the null hypothesis. The number of degrees of freedom (df) is given by df = (R-1) (C-1). Where R is the number of rows and C is the number of columns, the critical region for the test at 5% significant level is therefore given as $\chi^2 \geq \chi^2_{0.05}$ [(R-1) (C-1)].

**Principal Component analysis (PC):** Principal component analysis is a statistical procedure that transforms p correlated variables into a set of p new uncorrelated variables called PC. Each principal component is a linear combination of the original variables. If we let the variables $y_1$, $y_2 \ldots y_p$ represent the linear combinations of the original variables $x_1$, $x_2 \ldots x_p$, then:

$$\begin{array}{lll} y_1 & = & w_{11}\mathbf{x}_1 + w_{12}\mathbf{x}_2 + \ldots + w_{1p}\mathbf{x}_p \\ y_2 & = & w_{21}\mathbf{x}_1 + w_{22}\mathbf{x}_2 + \ldots + w_{2p}\mathbf{x}_p \\ \vdots & \vdots & \qquad\qquad \vdots \\ y_p & = & w_{p1}\mathbf{x}_1 + w_{p2}\mathbf{x}_2 + \ldots + w_{pp}\mathbf{x}_p \end{array}$$

where, $w_{ij}$ is the weight of the $j^{th}$ (j = 1,2,…, p) variable for the $i^{th}$ (I = 1,2, …, p) PC. In matrix notation, this can be written as y = w′x. The weights are estimated such that the first PC accounts for the maximum variance in the data; the second PC accounts for the maximum variances that have not been accounted for by the first. Each succeeding component accounts for as much variance that has not been accounted for by the preceding components. In general, the $p^{th}$ component accounts for the variance that has not been accounted for by the first p-1 components [5].

$$w_{i1}^2 + w_{i2}^2 + \cdots + w_{ip}^2 = 1 \quad (i = 1, 2, \ldots, p)$$

and

$$w_{i1}w_{j1} + w_{i2}w_{j2} + \cdots + w_{ip}w_{jp} \quad (\text{for all } i \neq j)$$

and $w_i$ $(i = 1, 2, \ldots, p)$ are the corresponding eigenvectors of the variance-covariance matrix $\Sigma$ furthermore, $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_p \geq 0$.

Factor loadings measure the simple correlation between the original and the new variables. They give an indication of the extent to which the original variables are important in forming new variables. In this study MINITAB was used to perform the analysis.

## RESULTS

With reference to Table 2 and 3 the following results could be obtained:

**Absolute risk (AbR) for swampy area dwellers:**

$$\frac{101}{130} \times 100 = 78\%$$

**Absolute risk for non-swampy area dwellers:**

$$\frac{128}{206} \times 100 = 62\%$$

While 78% of those who dwell in swampy areas had malaria, 62% of those living in non-swampy environments had malaria.

**Relative Risk for malaria (RR):**

$\frac{0.78}{0.62} = 1.3$ (an RR greater than one is important in clinical evaluation of a patient)

From the results, it could be inferred that relative to people living in non-swampy areas, people living in swampy areas are 1.3 times likely to have malaria.

Table 2: Distribution of malaria status of residents on risk factors

| Nature of environment | Never been diagnosed | Ever been diagnosed |
|---|---|---|
| Swampy area | 29 | 101 |
| Smoky environment | 13 | 21 |
| Saw dust | 6 | 9 |
| Dusty | 8 | 19 |
| Public toilet | 2 | 2 |
| Chemical shops | 5 | 7 |
| Refuse dumps | 44 | 70 |
| Family history | | |
| Parents never had | 78 | 37 |
| Parents ever had | 30 | 192 |

Table 3: A 2×2 contingency table of risk factors on malaria status

| Malaria status of residents in Sekondi-Takoradi Metropolis (STMA) | | | |
|---|---|---|---|
| Risk factor | Ever | Never | Total |
| Swampy | 101 | 29 | 130 |
| Others | 128 | 78 | 206 |

**Attributable Risk (AtR):**

0.78-0.62 = 0.16

$$\text{Attributable risk\%} = \frac{\text{AtR}}{\text{AbR}} \times 100\% = \frac{0.16}{0.78} \times 100 = 20.5\%$$

Therefore 21% of the time the differences in the incidences of malaria between those exposed to the risk factor and those not exposed to the risk factor may be directly attributable to the presence of the risk factor in this study.

**The number of people needed to be prevented from the risk factor[7]:** This is a measure of the prevention efficacy. It is obtained by inverting the value of the attributable risk (1/0.16) which gives us 6. Thus we should prevent 6 people from getting malaria as a result of the swampy environment in one case.

**Odds:** Let P be the probability of experiencing the outcome:

$$\text{Then the odds} = \frac{P}{1-P}$$

**Odds ratio:**

$$\frac{0.78}{0.22} \div \frac{0.62}{0.38} = 2.17$$

The odds for the incidence of malaria for people living in swampy areas are 2.17 times greater than that of people who do not live in swampy environments.

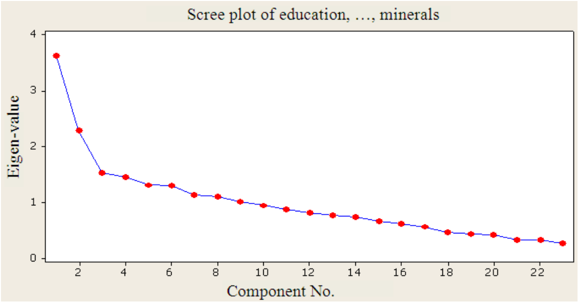


Fig. 1: Scree plot of eigen-values of risk factors on component numbers

**Scree plot:** The Scree Plot is derived by plotting the eigenvalues against the number of components. Typically, the Scree Plot slopes steeply down initially and then asymptotically approaches zero. The point at which the curve first begins to straighten out indicates the maximum number of components to extract[5,10].

**Scree plot:** The scree plot in Fig. 1 shows that 3 principal components can be extracted. An important decision in PCA is determining which principal component loadings are worth considering in the interpretation of each component. Hair et al suggested the use of the rule of thumb frequently employed by researchers. He indicated that factor loadings greater than 0.30 or less than -0.30 are considered significant, loadings greater than 0.40 or less than -0.40 are considered more important and loadings greater than 0.50 or less than -0.50 are considered very significant. This rule of the thumb is cited in Schneider[9]. Accordingly, variables that have absolute loadings of at least 0.32 were considered significant, extracted and interpreted under three main principal components and discussed below.

**Interpretation of the correlation matrix:** The first Principal Component (PC) accounted for 15.7% of the health variation in the data, the second accounted for 10% of the variation and the third accounted for 6.7%. It could also be inferred from the table that the first three (3) PC accounted for 32.4% of the total variation in the health data of the residents.

**Interpretation of the loadings:** The factors identified to have some form of influence on the health of the respondents which could be classified as PC one were the amount of intake of ; egg yolk, sugar, vegetables and cereals, with cereals being the most influential. Together, these factors can be termed the dietary behavior of the residents (Table 4).

Table 4: Principal Component (PC) loadings

| Variable (risk factor) | PC 1 | PC 2 | PC 3 |
|---|---|---|---|
| Education | -0.018 | -0.226 | -0.431 |
| Marital status | 0.096 | 0.388 | -0.099 |
| Location | -0.004 | -0.052 | 0.293 |
| Employment status | 0.098 | 0.501 | -0.170 |
| Financial | 0.093 | 0.472 | 0.099 |
| Exercise | -0.026 | 0.172 | 0.427 |
| Duration | 0.067 | 0.077 | -0.105 |
| Lunch | 0.141 | -0.132 | -0.024 |
| Meal | 0.021 | 0.027 | -0.097 |
| Supper | -0.010 | -0.008 | -0.089 |
| After supper exercise | -0.065 | -0.055 | 0.170 |
| Yolk of egg | 0.328 | -0.050 | 0.313 |
| Sugar | -0.327 | 0.022 | -0.134 |
| Salt | -0.302 | 0.075 | -0.145 |
| Pepper | -0.306 | 0.038 | -0.050 |
| Vegetables | -0.317 | 0.088 | 0.255 |
| Cereal | -0.359 | -0.031 | 0.277 |
| Fruits intake | -0.308 | -0.078 | 0.281 |
| Alcoholic intake | -0.041 | 0.335 | -0.047 |

Factors having further influence on the health of the respondents which can be classified under PC two were marital, employment and financial status of respondents, also included was the level of intake of alcoholic drinks, prevalent among them was the employment status of the residents, together these factors can be classified as social status of the residents (Table 4).

Educational background of the residents, amount of egg yolk intake and physical exercise were factors that contributed positively or negatively towards the health of the residents. These can be classified as PC three; the most influential of them being the educational level of the person. Together these factors can be termed the educational status of the residents (Table 4).

**Significant tests on the risk factors and diseases:** The following inferences were made about the risk factors associated with the incidences of diseases in the Metropolis using the Chi-square test of independence.

The Null hypothesis in each of the cases below was that:

- $H_0$: The two variables; risk factor and disease were independent
- The level of significance for each test is $\alpha = 0.05$
- We Reject $H_0$ if the p-value<0.05

**Testing for the significance of some risk factors on malaria:** The p-values for environment against malaria ($p = 0.012$) and that of one's family history against malaria ($p = 0.040$) show that there is a relationship between the two risk factors (environment and family history) and the incidence of Malaria.

## DISCUSSION

From the onset of the study we sought to assess the predominant diseases, the risk factors associated with the predominant disease, the effects of the environment, social and other health hazards on the health of residents of STMA.

It was found that majority (68%) of the sampled population have had malaria before, it was also found that 84% of those who have had malaria have parents who had been diagnosed of malaria. For those who had been diagnosed of malaria, 44% lived around swampy environments, which is noted for the bleeding of malaria parasites. The moderately high rainfall figures in the region coupled with the waste disposal problems in the region favors the prolific breeding in the sites. The absolute risk value(78%) for people living in swampy areas indicates that swampy area dwellers are at risk of getting malaria, this findings supports the assertion that malaria and other diseases are as a result of poor environmental hygiene[2]. The absolute risk value of 62% for dwellers of non-swampy areas indicates that besides the environmental factors, there are other factors which contribute to the incidence of malaria in the catchments area; some of these factors were identified through the Principal component analysis as dietary behavior, social status and educational level of the people. The relative risk value of 1.3, further confirms that those living in swampy areas are at risk of the incidence of malaria. The chi-square significant test also confirms that environment and family history are determinants when assessing the risk factors of people towards the incidence of malaria.

## CONCLUSION

Malaria Eradication Programs (MEP) have been considered both globally and in Ghana, Globally and nationally there have been some efforts to assess the possible risk factors associated with the incidence of malaria. Preventive measures have also been introduced to eradicate the identified risk factors.

Interestingly, when it comes to the districts in Ghana the same cannot be said about them. This study therefore, narrowed down the assessment of the risk factors to one of the several districts in the region. The study can specifically state that the swampy nature of the environment, ones family history, dietary behaviors, social status and educational background of the individuals, were the determinants of the incidence of malarias It is recommended that STMA takes the appropriate measures to prevent 6 people from the risk factor at any case.

## ACKNOWLEDGEMENT

I wish to express my heart felt gratitude to the many personalities who assisted and directed me to conduct this study. I thank Dr. Mahama Nantogmah of Huazhong University of Science and Technology Medical School who used his medical expertise to provide the medical guidance that I needed and Mr. Eyiah-Bediako Francis of University of Cape Coast - Mathematics and Statistics Department -Ghana, for his assistance in the designing of the questionnaire. My appreciation also goes to Mr. John Laste of the STMA for providing information needed about the Metropolis. Finally, I thank the administrators of the five main hospitals within the Metropolis for the various roles they played.

## REFERENCES

1. Dawood, R.M., 2002. Travellers' Health: How to Stay Healthy Abroad. 4th Edn. Oxford University Press, ISBN: 0192629476, pp: 130.
2. Marques de Sa, J.P., 2007. Applied Statistics Using Spss, Statistica, Mat lab and R. 2nd Edn., Springer Berlin Heidelberg New York, USA., ISBN: 9783540719717, pp: 505.
3. Ministry of Local Government and Rural Development, Ghana, 2006. Economic Characteristics of STMA. http://ghanadistricts.com/region/?r=5&sa=114
4. Hanrahan, E.J. and G. Madupu, 1994. Epidemiology and Biostatistics for USMLE. Prentice Hall International, UK. ISBN: 083850244X, pp: 111.
5. Johnson, R.A and D.W. Wicern, 1992. Applied Multivariate Statistical Analysis. 3rd Edn., Prentice Hall Inc., New Jersey, ISBN: 0130418072, pp: 356
6. Jones, A.M. and O. O'Donnel, 2002. Econometric Analysis of Health Data. John Wiley and Sons, Asia, ISBN: 0470841451, pp: 233.
7. Nordness, R., 2005. Epidemiology and Biostatistics Secrets. 1st Edn. Mosby, UK., ISBN: 10: 0323034063, pp: 288.
8. Puri, A.S., 2001. How to Prevent Common Serious Diseases. 2nd Edn. Sterling PVT Ltd., New Delhi, ISBN: 8120734661, pp: 340.
9. Schneider, B., 2003. Resident Capital outflows: Capital Flight or Normal Flows? A Statistical Interpretation. 2nd Edn., Overseas Development Institute, London, ISBN: 0850036348, pp: 15.
10. Sharma, S., 1995. Applied Multivariate Techniques. 5th Edn. Prentice Hall, New Jersey, ISBN: 0471310646, pp: 512.