July-August 2020
ISSN: 0193-4120 Page No. 4247 - 4252

A Cross-Sectional Study of Tuberculosis Patients in Dhemaji Hospital, Assam, India

Achinta Saikia, Bipin Gogoi
Department of Statistics Dibrugarh University, Dibrugarh
Phone-9613516846, 9435390986
Email-achintasaikia373@gmail.com, Bipin_stats@dibru.ac.in

Article Info
Volume 83
Page Number: 4247 - 4252
Publication Issue:
July - August 2020

Article History
***Article Received**: 25 April 2020*
***Revised**: 29 May 2020*
***Accepted**: 20 June 2020*
***Publication**: 10 August 2020*

Abstract

Aims: Tuberculosis (TB) is a major infectious disease among worldwide, affects the lungs. It also may be transmitted to other parts of the body, including the bones, meninges, kidneys and lymph nodes. For control of tuberculosis, it is necessary to find the cases as early as possible, and to ensure that the tuberculosis patients complete their treatment and get cured.

Methods: Being a register-based cross-sectional study, we analyzed the records of 800 tuberculosis patients at Dhemaji Civil Hospital from the TB register booklets, Jan 2017 to Dec. 2018. To test the prevalence of pulmonary Tuberculosis between male and female, Logistic regression was used to estimate the odds ratio for categorizing TB and living area of patients.

Results: From the total of 800 number of registered Tuberculosis patients from which 310 samples are randomly drawn for our sample, we found 217 males and 93 females respectively. Here we have considered pulmonary and extra pulmonary cases and have observed that male pulmonary is 81% and extra pulmonary is 19% whereas among the female74% of the cases are pulmonary and 26% are extra pulmonary respectively. Though male pulmonary is higher than female, there is no significant difference between male pulmonary and female pulmonary. Again as we have found that the patients odds ratio between category of tuberculosis and living area of patients is 2.54, ($Z > Z\ (0.05)$) so we reject null hypothesis. The odds ratio shows that those tuberculosis patients in rural area are 27 times more likely to be pulmonary tuberculosis than those of urban areas and the difference is also found to be statistically significant.

***Keywords:** DOTS, Extra pulmonary, Logistic Regression, Odds Ratio, Pulmonary, RNTCP, Tuberculosis.*

I. INTRODUCTION

Tuberculosis is a disease that infects the lung parenchyma primarily. It also may be transmitted to other parts of the body, including the meninges, kidneys, bones and lymph nodes. It is caused by the bacteria called mycobacterium tuberculosis (Paul et al 2009).

According to WHO, tuberculosis is a disease that cause death worldwide which is closely associated to malnutrition, poverty, overcrowding, inadequate health care and substandard housing. The infection of tuberculosis has been estimated to 1/3 of the world's population and has become the leading cause of death in the world. It is also a major killer of HIV positive patients (WHO 2018).

India is a country where we can find the highest number of patients suffering from TB. It has been estimated that about 40% of India's population is infected with TB bacteria. Among these, the vast majority of patients have latent TB rather than the recent TB disease (RNTCP India 2019). Even today in India, every three minutes, two TB patients demise. Some of the serious challenges to control TB in India include poor primary health care base in rural areas; deregulated private health care which leads to wide spread irrational use of first line and second line anti TB drugs; poverty, HIV infection and above all, corrupt administration (Sandhu 2011). A collaborative effort is being in progress between National Tobacco Control Programme (NTCP) and National Rural Health Mission (NRHM), which might reform and take the initiative to progress primary health care base in rural areas (NRHM 2005). In addition to this, National Tobacco Control Programme (NTCP) has established various initiatives in

July-August 2020
ISSN: 0193-4120 Page No. 4247 - 4252

coordination with the private sector and the India Medical Association (IMA) to progress TB care. People in India are still under wrong conception that TB is a disease of poor people, especially living in slum areas (CDCP 2012; Ayele et al 2004).

1.1 Types of Tuberculosis

a) Pulmonary Tuberculosis (PT)

Pulmonary Tuberculosis has been classified into two: sputum-positive and sputum-negative. Sputum-positive in a patient is to be seen with at least two samples of positive sputum for acid- Fast Bacilli (AFB) and when a patient with all three sample of sputum are found negative for AFB, should after a course of broad spectrum antibiotics repeat the sputum examination (Smith 2003; Zhai et al 2019). If found negative an X-Ray abnormalities are accordant with active pulmonary tuberculosis, it is known as a sputum- negative case. The conclusion of diagnosing and treating such a case will be taken by then physician only (Tomioka 1998; Swai et al 2011).

b) Extra Pulmonary Tuberculosis (EPT)

It is classified as bones and joints, urogenital tract, nervous system, lymph nodes and intestines.

Mostly tuberculosis germs spread through air, especially when cough or sneezes; it is spreaded in the form of tiny droplets. When these droplets are inhaled by a healthy person he is infected with tuberculosis (WHO 2018).

1.2 Specific Objectives

I have chosen the topic "A Cross-Sectional Study of Tuberculosis Patient in Dhemaji Hospital, Assam, India**"**. Study of the disease TB can be done from different angles and statistical analysis can be done accordingly.

So I have set my ultimate objectives as follows.

- To compare the prevalence of pulmonary tuberculosis between male and female
- To estimate and test the Odds ratio for category of tuberculosis and living area of patients

II. Methods

2.1 Material for this Study

I had to use secondary data. I have been using the following information for preparing the report.

- The total number of patients registered during the year 2017 in Unit-I.
- Age of the patient
- Sex of the patient
- Category of TB- Pulmonary or Extra pulmonary
- Type of TB patient with respect to living area

2.2 Method of the Analysis

To achieve the mentioned objectives we can process in different ways. All the different approaches lead to the same or identical conclusion. We have chosen the following statistical techniques for analyzing data.

Table 1: Contingency table

a	b	a+b
c	d	c+d
a+c	b+d	N

$$\chi^2 = \frac{N(ab-bc)^2}{(a+C)(b+d)(a+b)(c+d)} \sim \chi_1^2 \qquad \ldots (1)$$

2.2.1 Test for single proportion (Gupta and Kapoor 2003)

Suppose p_0 is the assumed proportion.

To test H_0: $p = \hat{p}$

Against H_1: $p \neq \hat{p}$

The test statistic is

$$Z = \frac{\hat{p}-p}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1) \qquad \ldots (2)$$

Where

$\hat{p}$ = estimated proportion

p_0=assumed proportion.

N=sample size

2.2.2 INFERENCE BY ODDS RATIO (Kleinbaum 2010)

We have already mentioned some test to draw inference about the relationship between two attributes. The same problem can be approached by odds ratios.

Odds ratio is defines as

$$OR = \frac{P(Y=1,X=1)/P(Y=1,X=0)}{P(Y=0,X=1)/P(Y=0,X=0)}$$

$$= \frac{P_{11}P_{00}}{P_{10}P_{01}} \qquad \ldots (3)$$

There is no association between X and Y, then the odds ratio equal to 1 and the logarithm of odds ratio is 0.

Where P_{00}= Probability that the randomly selected person is does not have the risk factor and does not have the disease

P_{11} = Probability that a randomly selected person has then risk factor and the disease. Similarly P_{10} and P_{01} are defined.

Odds ratio and relative risk can be defined with the help of cross table (Table 2)

July-August 2020
ISSN: 0193-4120 Page No. 4247 - 4252

Table 2: odds ratio and relative risk can be defined with the help of cross table

	Y=1	Y=0	TOTAL
X=1	A	C	A+C
X=0	B	D	B+D
TOTAL	A+B	C+D	N

Odds ratio (OR) is estimated by

$$OR = \frac{AD}{BC} \quad \ldots (4)$$

Estimated variance of OR is

$$\delta^2{}_{logor} = \frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D} \quad \ldots (5)$$

To test the null hypothesis, there is no association between the two attributes the statistics is

$$Z = \frac{(Log_e OR)}{\sqrt{\sigma^2}\, Log_e\, OR} \quad \ldots (6)$$

It is approximately distributed as χ^2 – various with 1 d.f.

The (1-α) x 100% confidence interval of OR is:

$$\widehat{OR} = \text{Antilog } \{\log OR \pm 1.96\{(\sigma_{\log \widehat{OR}})\} \quad \ldots (7)$$

2.2.3 Sample size (n)

The size of the sample considered for any scientific enquiry plays a vital role in the reliability of the result obtained. It should not be very large or very small -it should be simply the appropriate one for the particular study. If we take small n, then the result will be meaningful and thus the whole work will be simply wastage of money and time and other resource. On the other hand, if we take n larger than required, this is also wastage of resource and in case of study involving human being we are unnecessarily harassing people without any gain to anybody. So, utmost care should be taken in deciding the sample size for any scientific study.

So the sample size should be different for different objectives stated by us. I shall need the following formulas for computing n.

$$N=\frac{z^2_{\frac{\alpha}{2}}p(1-p)}{d^2} \quad \ldots (8)$$

Where, p = anticipated value of the proportion in the population

d = required precision for the estimated proportion

Z_k= upper k^{th} percentile from N (0, 1)

$$N = \frac{(z_{\frac{\alpha}{2}}\sqrt{2p(1-p)}+z_\beta\sqrt{p_1(1-p_1)+p_2(1-p_2)}\,)^2}{(p_1-p_2)} \quad \ldots (9)$$

Where, p_1 and p_2 are proportion anticipated in the two populations

p (p_1 + p_2)

α= Level of significance

1-β =Power of the test required

III. Collection of the data

The collection of the data is the most important step in statistical investigation. There for the data must be collected keeping in the view, aim scope of the enquiry and the available resources. In the case of inaccurate unreliable data, the results obtained would be erroneous and misinterpret the whole study. Therefore proper method should be adopted in the collection of the data for studying the true behavior of the variables under study. The collection of data is the foundation upon which the whole superstructure of the statistical analysis is to be raised; therefore the foundation must be properly laid.

In the present study, the purposive sampling is used to select the study area of the Dhemaji Civil Hospital, Dhemaji and simple random technique has been applied to collect the data on the TB patient.

3.1.1 Judgment or purposive sampling

Judgment or purposive sampling, the choices of sample depends upon exclusively on the judgment of the investigator.

3.1.2 Simple random sampling

Simple random sampling, the sampling technique refers to each and every unit of the population has an equal chance of being selected in the sample.

We can use any random numbers using these tables starting any page of it and we can use scientific calculators for selecting a random number.

3.2 Determination of sample size

In the stage of the survey we have to some idea about the approximation sample size. According to WHO report anticipated proportion for male and female are 56% and 44% respectively. The difference will be considered to be significant if at least 20% difference is there and it is assured with 95% guarantee. From preliminary enquiry, it is observed that there is a chance of getting incomplete information in the records which will be around 10%.

Here,

p_1=Anticipated proportion for male =56%=0.56

p_2= Anticipated proportion for female=44%=0.44

For two tailed test,

$$p=\frac{p_1+p_2}{2}=0.5$$

We get n=141

If, 10% non response is incorporated, then

$$n=\frac{110}{100}x141=155.1\cong 155$$

Thus the final sample size required for each group (male and female) is 155 each. So the total sample size needed is (2x155) = 310. So I have decided to take my n is 310.

3.3 Drawing of a sample

I have to take my sample from Unit-I of Dhemaji district and for the year 2016 and 2018. There are total 800 number of registered case and I have to draw a sample of size 310 from this 800 which is my population size N.

For that we have use random number table (column 1-4, 5-8, 9-12,13-16 of Rao, Mitra and Mathai (Rao et al 1966) edited

July-August 2020
ISSN: 0193-4120 Page No. 4247 - 4252

table) or I can generate random numbers between 1 to 800 using calculator.

I have chosen those serial numbers from the TB resister booklets, Jan 2017 to Dec 2018 of Unit-I, Dhemaji medical, Dhemaji district.

The Random number table given (Table 3)

Table 3: Random number table

Year 2017	42	94	149	151	155
	177	181	186	. .	.
Year 2018	43	66	176	196	. .
	.	.	.	.	……..

3.4 Surveyed area

There are 5 TB units under Dhemaji Civil Hospital, Dhemaji district

Unit-I: TU Dhemaji

Unit-II: TU Gogamukh

Unit-III: TU Bengenagarah

Unit-IV: TU Silapathar

Unit-V: TU Jonai

I have purposively selected Unit-I of Dhemaji district for collection data on TB patients for my study. There are two centers under Unit-I.

They are-

1. DMC Dhemaji
2. DMC Civil Hospital

I had to collect secondary data from the Unit-I located at Dhemaji urban centre, Dhemaji, where the records from both subunits are kept.

IV. Analysis of Data

Analysis of data is the most important part of any statistical investigation of the collected data. Analysis should be done with properly scrutinized, edited and tabulated data. After tabulation, an investigator should determine what statistical technique is suitable for analyzing the data keeping in mind different factors such as sample sizes, type of data etc. the findings of the analysis will lead to the conclusion regarding the study. The calculations done by SPSS Version 13.0 software package, some graphics are drawn in Microsoft Excel. Significance level was considered at p-value < 0.05.

4.1 Test for single proportion of Male patient

From WHO data we have seen that the TB is more among male than female. So, we wanted to test with our sample that the situation is same in our population also.

Thus we set our hypothesis as

H(claim): TB is more among male than female.

To test this hypothesis our H_O and H_1 are

H_0: $p = p_O = \frac{1}{2} = 0.5$

H_1: $p > \frac{1}{2}$

Where p=proportion of male TB patient in the population

With α=0.05, the critical region will be as shown in figure 1.

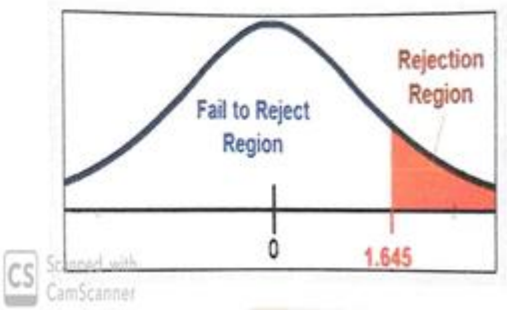

Figure 1: Critical Region

The test statistic to be used here is Z-test by using formula (2)

$\hat{p}$=Estimated proportion of the male TB patient $=\frac{217}{310}$ =0.7

n=310

p_0=assumed proportion=0.5

So,

$$z=\frac{0.7-0.5}{\sqrt{\frac{0.5(1-0.5)}{310}}}=\frac{0.2}{\sqrt{8.064}}$$

=0.07

Since, Z > 0.07, test is significant. So, we can reject H_o on the basis of our sample to accept H_1 and conclude that our null hypothesis is true

4.2 Test of Independence of Attributes Sex and Class of Tuberculosis

Here the class is either pulmonary or extra pulmonary and sex is male and female.

Now our H (claim): there is an association between sex and class of TB.

Ho: Category of the patient is independent of sex

H1: There is an association between sex and class of TB patient

Distribution of patients by sex and category (Table 4)

Table 4: Distribution of patients by sex and category

	Pulmonary	Extra Pulmonary	Total
Male	176	41	217
Female	69	24	93
Total	245	65	310

Using formula (1) we get,

$$\chi^2=\frac{N(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)} = 1.88$$

Critical value of $\chi^2(0.05, 1) = 3.841$

Since $\chi^2 < \chi^2(0.05,1)$ we accept the null hypothesis H_o, which means that our claim is true.

4.3 Comparison between Male and Female Pulmonary Patients

Here, the category is either Pulmonary or Extra Pulmonary and sex is male and female.

Now our H(claim): Male pulmonary patient is more than female pulmonary patient in the population.

Ho: $p_1 = p_2$

H1: $p_1 > p_2$

Where, p_1= Proportion (male pulmonary)

p_2= Proportion (female pulmonary)

Distribution of patients by sex and category (Table 5)

Table 5: Distribution of patients by sex and category

	Pulmonary	Extra pulmonary	Total
Female	69	24	93
Male	176	41	217

The test statistic to be used here is Z-test by using formula (2)

$\hat{p}_1 = \frac{176}{217} = 0.81$

$\hat{p}2 = \frac{69}{93} = 0.741$

$\hat{p} = 0.7893$ (using $\hat{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$)

n=310

So, $Z = \frac{0.81 - 0.741}{\sqrt{0.78(1-0.78)(\frac{1}{93}+\frac{1}{217})}} = 1.378$

Since, Z<1.645, test is not significant. So, cannot reject H_o on the basis of our sample and conclude that our hypothesis is not true.

4.4 Estimation of Proportion of Pulmonary Case

Let, p_1 and p_2 are proportions for female and male pulmonary patients respectively

Let, x represents the class of female TB patients

x=1, if the class is pulmonary

=0, otherwise

Now, x is a Bernoulli variate.

So, ML estimate of p_1 is given by

$p_1 = \frac{69}{93} = 0.74$

And $1 - p_1 = 0.26$

And similarly for male $p_2 = 0.81$, $q_2 = 0.19$

4.5 Tests for Association between Category of Tuberculosis and Living Area of Patients

The Distribution of patients by category and living area given (Table 6)

Table 6: Distribution of patients by category and living area

	Rural	Urban	Total
Extra Pulmonary	215	30	65
Pulmonary	48	17	245
Total	263	47	310

We define,

$OR = \frac{Odds\ of\ pulmonary\ to\ Extra\ Pulmonary\ patient\ in\ Rural\ area}{Odds\ of\ Pulmonary\ to\ Extra\ pulmonary\ patient\ in\ Urban\ area}$

H_o: $\widehat{OR}=1$ ~ H_O: $log_e \widehat{OR}=0$

H_1: $\widehat{OR}>1$ H_1: $log_e \widehat{OR} \neq 0$

Using formula (7) and table 6, we get

$\widehat{OR} = \frac{215 \times 17}{48 \times 30} = 2.54$

Now the natural logarithm of odds ratio is

$log_e \widehat{OR} = 0.93$

$\widehat{\sigma^2}(\log \widehat{OR}) = 0.1176$

The test statistic is

$Z = \frac{0.15}{\sqrt{0.14}} = 2.71$

Since, Z>Z (0.05, 1) so we reject Ho. Thus, we have found that those with rural area TB patients are 15% more likely to be pulmonary TB than those urban areas and the difference is found to be statistically significant.

V. Conclusion

Tuberculosis prevalence is highest in India. To reduce the burden of disease Govt. of India has introduced Directly Observed Treatment Short (DOTS) course under Revised National Tuberculosis Control Programme (RNTCP) program. Under this program trend of incidence and deaths due to this disease has been reducing. In this brief report entitled "A Statistical study on status of Tuberculosis patients in Dhemaji Civil Hospital" we have tried to analyze the data on Pulmonary and Extra Pulmonary TB patient of Dhemaji Civil Hospital, Dhemaji, Assam, and India.

The Unit-I had 800 number of registered TB patient from which 310 samples are randomly drawn for our sample. So, our study is based on a sample of size 310 out of which 217(70%) are male and 93(30%) are female respectively. Here we have considered pulmonary and extra pulmonary TB cases and have observed that male pulmonary is 81% and Extra pulmonary is 19% whereas among the female 74% of the cases are pulmonary and 26% are extra pulmonary respectively. Though male pulmonary is higher than female, we have found that there is no significant difference between male pulmonary and female pulmonary. And the test of association between sex and category also established that there is no significant between them.

AGAIN THE ODDS RATIO SHOWS THAT THOSE TB PATIENTS IN RURAL AREA ARE 15% MORE LIKELY TO BE PULMONARY TB THAN THOSE OF URBAN AREAS AND THE DIFFERENCE IS ALSO FOUND TO BE STATISTICALLY SIGNIFICANT. THE ODDS OF

July-August 2020
ISSN: 0193-4120 Page No. 4247 - 4252

MARRIED TO UNMARRIED TB PATIENT ARE 2.06 TIMES MORE IN PULMONARY GROUP THAN THE EXTRA PULMONARY GROUP.

REFERENCES

1. Ayele W.Y., Neill S.D., Zinsstag J., Weiss M.G., Pavli K.I.(2004); Int J Tuberc Lung Dis. 8(8): 924-937.
2. Centers for Disease Control and Prevention. Sept 1, 2012. http//www.cdc.gov/tb.
3. Global Tuberculosis Report, World Health Organization (WHO), 2018. http//www.who.int/publications/global_report/en/.
4. Global Tuberculosis Report, World Health Organization (WHO), 2018.
5. Gupta SC and Kapoor VK (2003); Fundamental of Mathematical Statistics, 11th Edition, Sultan Chand and Sons Educational Publisher, New Delhi.
6. Kleinbaum D.G. (2010); Introduction to logistic Regression. In: Logistic Regression. Statistics in the Health sciences. Springer New York, NY. Pp-1-38.
7. National Rural Health Mission, Framework for implementation. Ministry of Health and Family Welfare, Govt. of India. 2005-2012.
8. Paul P., Williams B. Brunner and Suddarth textbook of Canadian Medical Surgical Nursing.2009; 5: 598-658.
9. Sandhu G.K.(2011); J Glob. Infect Dis. 3(2): 143-150.
10. Smith I. (2003); Clin Microbiol Rev. Jul. 16(3): 463-496.
11. Swai H.F., Mugusi F.M., and Mbwambo J.K. (2011); BMC Res Notes. 4: 475.
12. TB India Revised National TB Control Programme Annual Report 2019. www.tbcindia.nic.in/.
13. Tomioka H.,(1998); Nihon Rinsho Dec. 56(12): 3001-3007.
14. Rao, Mitra and Mathai (1966) table of random number.
15. Zhai W., Wu F., Zhang Y., Fu Y., Liu Z.(2019); Int J mol sci. 20(2): 340.