Bio-Science Research Bulletin
http://www.bpasjournals.com

Original Research Article

Print version ISSN 0970 0889
Online version ISSN 2320 3161
DOI: 10.5958/2320-3161.2020.00006.1
Vol. 36 No. 1, January-June 2020: P.40-49

A Statistical Evaluation of the Data of Ovarian Cancer and Uterine Cancer in Women

V. Sharon Philomena*

Author's Affiliation:
Assistant Professor, P.G. Department of Mathematics, Women's Christian College (Affiliated to University of Madras), Chennai, Tamil Nadu, India

***Corresponding Author:**
V. Sharon Philomena,
Assistant Professor, P.G. Department of Mathematics, Women's Christian College (Affiliated to University of Madras), Chennai, Tamil Nadu, India

E-mail:
sharonphilo2008@gmail.com

Received on 07.02.2019
Accepted on 18.10.2019

ABSTRACT

In this paper, we discuss the emerging trends about women related cancer and detailed information about the incidence of cancers along with the statistical data. Ignorance of cancer among women leads to major risk factors and can be cured with proper treatment. We briefly discuss the statistical methods such as Poisson regression model and one way ANOVA classification in the increasing trends of women related cancer such as Ovarian cancer and Uterine Cancer.

KEYWORDS: Ovarian Cancer, Uterine Cancer, Poisson regression, one way ANOVA.

INTRODUCTION

Cancer is a disease in which cells in the body grow out of control. Cancer is always named for the part of the body where it starts, even if it spreads to other body parts later. There are five main types of cancer that affect a woman's reproductive organs: cervical, ovarian, uterine, vaginal and vulvar. The uterus is the pear-shaped organ in a woman's pelvis - the area below the stomach and in between the hip bones. When cancer starts in the uterus, it is called uterine cancer. It can be treated through proper diagnosis and by creating awareness among women. In this paper we briefly discuss certain mathematical methods to predict the increasing trends of ovarian and uterine cancer among women. These methods enable us to characterize these cancers and promote awareness among women thereby making it possible to treat it in the early times.

OVARIAN CANCER

Ovarian cancer is the most fatal cancer of the female reproductive system. Ovarian cancer is a disease in which, depending on the type and stage of the disease, **malignant (cancerous) cells** are found inside, near, or on the outer layer of the **ovaries**. An ovary is one of the two small, almond-shaped organs located on each side of the uterus that store eggs, or germ cells, and produce female hormones estrogen and progesterone. Ovarian cancer is the **third** most commonly detected cancer amongst **Indian women** and at the **last stage** it is difficult to treat and is often **fatal**.

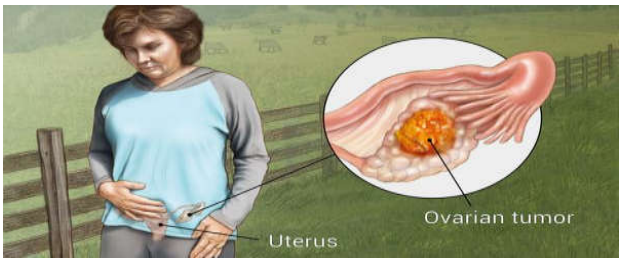

Figure 1: Ovarian Tumor

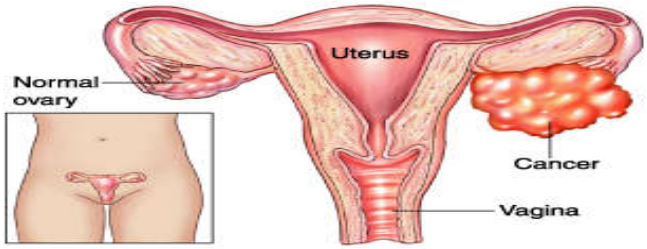

Figure 2: Uterus

Globally Ovarian cancer is the **seventh** most common cancer in the world! A **5-year** rate of survival, amongst ovarian cancer patients, worldwide, is **45**%, if the cancer spreads to other organs in the stomach. If the cancer is restricted to the ovaries and fallopian tube, the overall 5-year rate of survival of patients is **92**%!

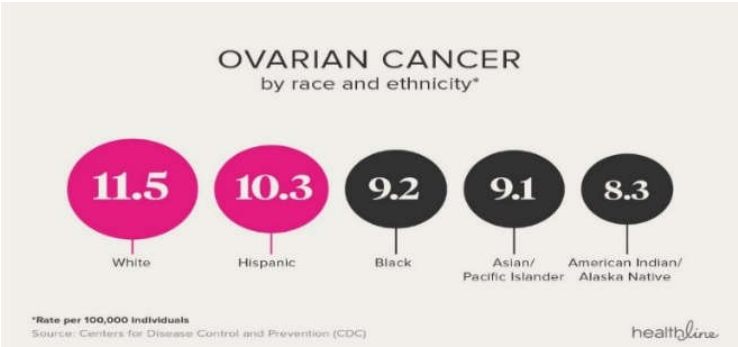

Figure 3: Ovarian Cancer

VICTIMS OF OVARIAN CANCER:

Most women who develop ovarian cancer are diagnosed after menopause, at **age 55** or older, though patients in their 40's and 50's have also been diagnosed with the disease. If a woman has a strong family history of breast or ovarian cancer, then she may also be at an increased risk.

TYPES AND STAGES OF OVARIAN CANCER

There are more than **30** different types of ovarian cancer, which are classified by the type of cell from which they start. Cancerous ovarian tumors start from three common cell types:
Surface Epithelium - cells covering the outer lining of the ovaries,
Germ Cells - cells that are destined to form eggs,
Stromal Cells - Cells that release hormones and connect the different structures of the ovaries.

Figure 4: Ovarian cancer stages

IGNORANCE OF OVARIAN CANCER IN INDIA

Estimates show that in the overall **5-year survival** rate for ovarian cancer, it is seen that in **India**, it is a poor **45 %.** One of the main reasons for such low success rates against this cancer is the fact that diagnosis of ovarian cancer is made only during the advanced stages of the disease**. 56**% of the diagnoses of ovarian cancer are made only in later **state III** and **stage IV** of the cancer. Inherited gene variants of **BRCA1** and **BRCA2** can increase a women's risk of developing ovarian cancer by **65**% and **35** % respectively. BRCA1 and BRCA2 genes are the ones that are mutated in 50% of all ovarian cancers.

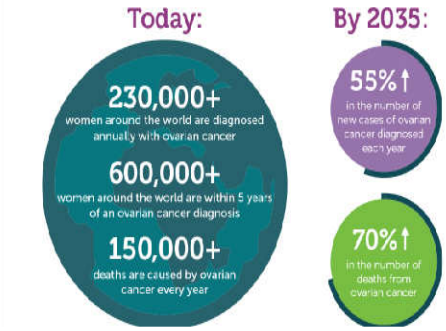

Figure 5: Ovarian Cancer in India

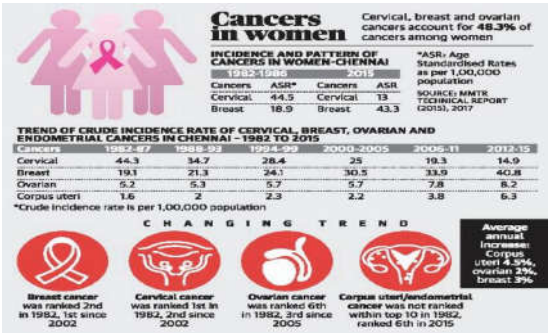

Figure 6: Cancer in women

GLOBAL KEY FINDINGS

It was estimated that in **2012** there were **239,000 cases** and **152,000 deaths** worldwide from **ovarian cancer**, with some **600,000** women living within five years of a diagnosis. It is estimated that by **2035**,

incidence will increase to **371,000** a year (55%) and deaths will increase by **67%** to 254,000. There are major challenges in dealing with global cancer statistics, mainly due to huge variations in registration of cancer incidence and mortality. A country is considered to have high quality data when it records more than 50% of cases. For a considerable number of countries, statistics are estimated rather than based on fact. It is thought this may lead to underestimation of cases and an overestimation of survival rates. China has the largest number of diagnoses per year (34,575), followed by **India (26,834),** then the USA (20,874).

MATHEMATICAL MODEL TO ESTIMATE THE TREND OF OVARIAN CANCER

The Age Specific Incidence Rate (ASIR) for ovarian cancer reveals that the disease slowly increases from 35 years of age and reaches a peak between the ages 55-64. The trend analysis of ovarian cancer by period shows an increasing trend in the incidence rate of ovarian cancer in most cases. Analysis of data by ASIR revealed that the mean annual percentage increase was higher for women in the middle and older age groups in the year 2013-2019. Estimation of annual percent change (EAPC) in ovarian cancer is by Poisson regression model through Maximum Likelihood Estimation (MLE).

The Age Specific Incidence Rate (ASIR) for ovarian cancer revealed that the disease increases from 35 years of age and reaches a peak between the ages 55-70. The trend analysis by period showed an increasing trend in the incidence rate of ovarian cancer in most of the registries, with a mean annual percentage increase in ASR ranged from 0.7% to 2.4 %.

POISSON REGRESSION MODEL

Regression analysis is a form of predictive modeling technique which investigates the relationship between a **dependent** (target) and **independent variable (s)** (predictor). This technique is used for forecasting, time series modeling and finding the causal effect relationship between the variables. For example, relationship between rash driving and number of road accidents by a driver is best studied through regression.

Regression analysis is an important tool for modeling and analyzing data. We fit a curve or line to the data points, in such a manner that the differences between the distances of data points from the curve or line is minimized. There are various kinds of regression techniques. These techniques are mostly driven by three metrics (number of independent variables, type of dependent variables and shape of regression line).

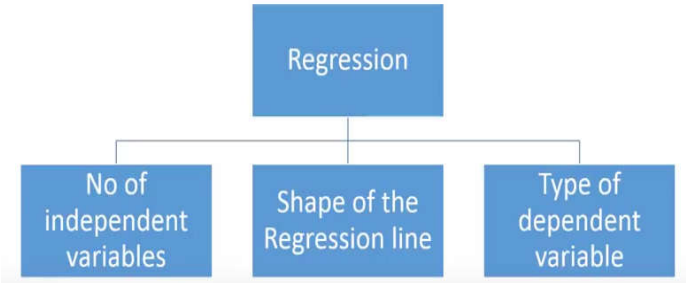

Figure 7: Regression analysis

Poisson linear regression model

It is one of the most widely known modeling techniques. Linear regression is usually among the first few topics which people pick while learning predictive modeling. In this technique, the dependent variable is continuous, independent variable(s) can be continuous or discrete, and nature of regression line is linear. Linear Regression establishes a relationship between **dependent variable** Y and one or more **independent variables** X using a **best fit straight line** (also known as the regression line). It is represented by an equation $Y = a + bX + e$, where a is the intercept, b is the slope

of the line and e is the error term. This equation can be used to predict the value of target variable based on given predictor variable(s).

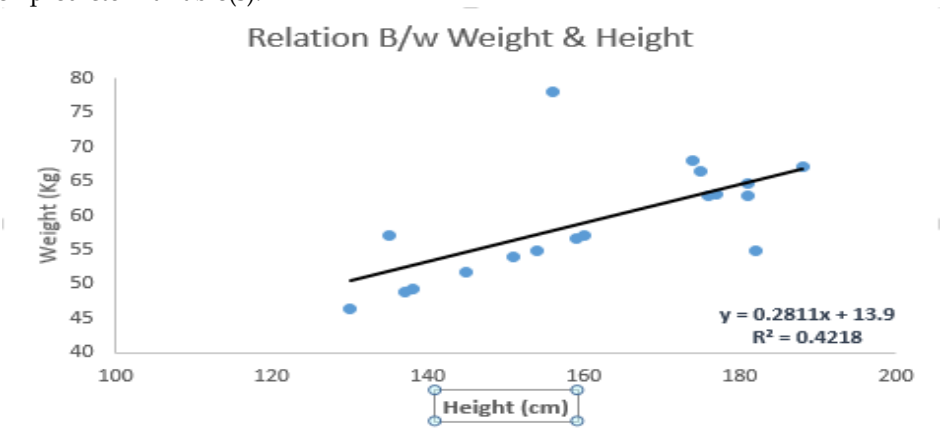

Figure 8: Relation between weight and Height

AGE STANDARDIZATION RATES

Rates are a phenomenal tool for comparing characteristics across various populations and also different segments of a population or the same population over a certain time period.

Crude rates

When we are using rates to analyze unusual events, such as certain crimes or the incidence of rare diseases, they are often expressed as the number of people or occurrences per 1,000 or 100,000 individuals in the population. These rates are often referred to as the crude **rates.** The example here uses only two age groups: often, the characteristic being studied varies considerably across ages and therefore, narrower age categories are required. However, during the same period, the crude rate for each age group decreased. The short answer is that the crude rate for the total population, while it accurately represents the incidence of death due to cancer each year, it is not the correct indicator to use to compare the incidence between years.

Table 1: Cancer deaths and population estimates, Canada, 2000 and 2011

Age group	Characteristic	2000	2011
0 to 39 years	Estimate of population	17,068,876	17,191,850
	Number of deaths	1,345	1,004
	Crude rate	7.9	5.8
40 years and over	Estimate of population	13,616,854	17,150,930
	Number of deaths	61,325	71,472
	Crude rate	450.4	416.7
Total all ages	Estimate of population	30,685,730	34,342,780
	Number of deaths	62,672	72,476
	Crude rate	204.2	211.0

CALCULATION OF THE AGE STANDARDIZATION RATES

The detailed calculation of the age-standardized mortality rate is presented here using the example of deaths due to cancer, and the year 2000 data from Table 1. The rates are standardized to the 1991 population.

To calculate the age standardized mortality rate (ASMR), we must first calculate the **age-specific (mortality) rates** for each age group by dividing the number of deaths by the respective population, and then multiplying the resulting number.

Age -specific rate, 0 to 39 years.
= 1,345 (number of deaths) ÷ 17,068,876 (total population) × 100,000
= 7.9 cancer deaths per 100,000 populations

Age-specific rate, 40+years
= 61,325 (number of deaths) ÷ 13,616,854 (total population) ×100,000
= 450.366 cancer deaths per 100,000 populations

We then multiply each of the age-specific rates by the proportion of the 1991 population belonging to the particular age group (called the standard population weight). In 1991, 61.6% of Canadians were under 40 years of age and 38.4% were of the age 40 or older. The age-standardized rate is obtained by adding the resulting numbers.

ASMR
= (7.9 × 61.6%) + (450.4 × 38.4%)
= 4.9 + 173.0
= 177.9 cancer deaths per 100,000 standard population.

Thus using this we use the age standardization rate for the estimation of annual percentage change in the increasing trend of ovarian cancer between the years **2013-2019.**

Table 2: The increasing trend of ovarian cancer between the years 2013-2019

Date	Female	Population	Death (Approx)	Age Standardization (0-30)	(30-50)	(55-70)
2019	65,29,93,704	1,35,02,76,477	38350	0	11505	26845
2018	64,90,13,567	1,35,98,67,126	36170	0	12660	23511
2017	64,52,21,366	1,33,91,80,127	34650	0	8663	25988
2016	63,78,79,447	1,32,41,71,354	32330	0	14549	17782
2015	63,04,89,688	1,30,90,53,980	29600	0	8880	20720
2014	62,30,63,359	1,29,38,59,294	27300	0	12285	15015
2013	61,55,91,238	1,24,98,17,000	25063	0	6266	18797

Estimation of annual percent change (EAPC) in ovarian cancer by Poisson regression model through Maximum Likelihood Estimation (MLE)

UTERINE CANCER

The uterus, or womb, is part of the female reproductive system, which also includes the ovaries, fallopian tubes, cervix (neck of the uterus) and vagina (birth canal). The uterus is about the size and shape of a hollow, upside-down pear. It sits low in the abdomen between the bladder and rectum, and is held there by muscle. It is joined to the vagina by the cervix, which is the neck of the uterus. The uterus is where a fetus grows during pregnancy.

When cancer starts in the uterus, it is called uterine cancer. The uterus, also called the womb, is where the baby grows when a woman is pregnant. The most common type of uterine cancer is also called **endometrial cancer** because it forms in the lining of the uterus, called the **endometrium**.

The rare types of uterine cancer that begin in the muscle of the uterus include:

- endometrial stromal
- leiomyosarcoma
- undifferentiated sarcoma.

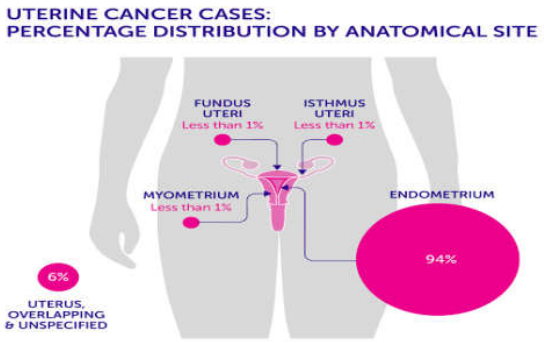

Figure 9: Uterine cancer cases: Percentage distribution by anatomical site

All women are at risk for uterine cancer and the risk increases with age. Most uterine cancers are found in women who are going through, or who have gone through menopause—the time of life when a woman's menstrual periods stop.

Each year, approximately 50,600 women in the United States get uterine cancer. It is the fourth most common cancer in women in the United States and it is the most commonly diagnosed gynecologic cancer.

DETAILED SURVEY OF THE TOTAL NUMBER OF CASES REGISTERED PERCENT OF NEW CASES AND PERCENT OF DEATH BY AGE GROUP

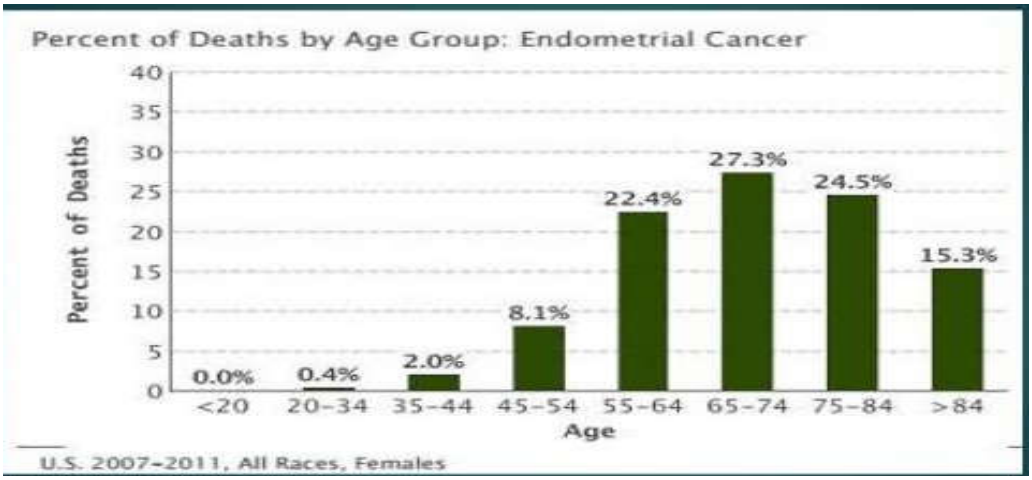

Figure 10: Percent of Death by Age group: Endometrial Cancer

Figure 11: Number of Cases Registered Percent of New Cases and Percent of Death by Age Group

METHODS AND MATERIALS

The t test is commonly used to test the equality of two population means when the data are composed of two random samples. We wish to extend this procedure so that the equality of $r \geq 2$ population means can be tested using r independent samples. Thus the hypothesis and the alternative are

$$H_0 : \mu_1 = \mu_2 = ... = \mu_r$$

$$H_1 : \text{at least two means are not equal}$$

$\mu_j, j = 1,2,...,r$ is the mean of the j^{th} population.

It is not hard to imagine situations in which it is of interest to compare a number of means. For example, 5 varieties of corn are available, and it is to be determined whether or not the average yield from each variety is the same; a company is testing 3 brands of bicycle tires and wants to know if the average life of each brand is the same; 4 teaching methods are being investigated for their effectiveness; an automotive company wants to determine which of the 4 seat-belt designs would provide the best protection in the event of a head-on collision; a drug company would like to compare the effectiveness of 6 different drugs for treating diabetes.

In designing an experiment for a one-way classification, units are assigned at random to any one of the r treatments under investigation. For this reason, the one-way classification is sometimes referred to as a completely randomized design. When an F – test is used to test a hypothesis concerning the means of three or more populations, this technique is called Analysis of Variance (ANOVA). The one-way analysis of variance is used to test the claim that three or more population means are equal. This is an extension of the two independent samples t – test. The response variable is the variable under comparison. The factor variable is the categorical variable being used to define the groups. We will assume k samples (groups).The one-way is because each value is classified in exactly one way. There are some of the Conditions or Assumptions given as

- The data are randomly sampled.
- The variances of each sample are assumed equal.
- The residuals are normally distributed.

The null hypothesis is that the means are all equal. The alternative hypothesis is that at least one of the means is different. Variation is the sum of the squares of the deviations between a value and the mean of the value. Sum of Squares is abbreviated by SS and often followed by a variable in

parentheses such as SS(B) or SS(W) so we know which sum of squares. There are two sources of variation

- the variation between the groups, SS(B), or the variation due to the factor.
- the variation within the groups, SS(W), or the variation that cannot be explained by the factor so it is called the error variation.

Now consider $MS_B = \frac{(\bar{X}_1 - \bar{X}_2)^2}{\frac{1}{n_1}+\frac{1}{n_2}}$, and

$$MS_W = SS_W/(N-r) = SS_W/(n_1+n_2-2) = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = s_p^2$$

Therefore, $F = \frac{MS_B}{MS_W} = \frac{(\bar{X}_1 - \bar{X}_2)^2}{S_p^2(\frac{1}{n_1}+\frac{1}{n_2})}$

Comparing this to the *t* test statistic in (1), we see that, under the null hypothesis, $F = t^2$.

Suppose that $c > 0$ is the critical value of the *F* test at some level α, then $F > c$ iff $t^2 > c$ iff $t < -\sqrt{c}$ or $t > \sqrt{c}$. Thus it makes no difference whether the two-tailed *t* test or the one-tailed *F* test is used, they both produce exactly the same result. This justifies an earlier remark that the *F* test may be regarded as an extension of the *t* test.

It is customary to summarize the results in a table. Here's a general form:

Table 3:

Source of Variation	**SS**	**d.f.**	**MS**	***F***
Between groups	SS_B	*r-1*	$SS_B/(r-1) = MS_B$	$\frac{MS_B}{MS_W}$
Within groups	SS_W	*N-r*	$SS_W/(N-r) = MS_W$	
Total	SS_T	*N-1*		

The partition of the total sum of squares is given by

$$\underbrace{\sum_{j=1}^{r}\cdot\sum_{i=1}^{n_j}(X_{ij}-\bar{X}_i)^2}_{SS_T} = \underbrace{\sum_{j=1}^{r}\cdot\sum_{i=1}^{n_j}(X_j-\bar{X}_i)^2}_{SS_B} = \underbrace{\sum_{j=1}^{r}\cdot\sum_{i=1}^{n_j}(X_{ij}-\bar{X}_j)^2}_{SS_W}$$

The *degrees of freedom* (d.f.) are defined as the total number of terms in an expression minus the number of linear constraints among the observations. The degrees of freedom in each of the three terms in the above partition are:

SS_T has $N = \sum_{j=1}^{r} n_j$ terms and

one linear constraint, $\sum_{i=1}^{n_j}(X_{ij} - \bar{X}_i) = 0$.

Thus, SS_T has *N – 1* degrees of freedom, $SS_B = \sum_{j=1}^{r} n_j\ (\bar{X}_j - \bar{X}_i)^2$ has *r* terms and one linear constraint, $\sum_{i=1}^{n_j}(X_j - \bar{X}_i) = 0$.

Thus, SS_B has *r – 1* degrees of freedom. SS_W has *N* terms and *r* linear constraints: $\sum_{i=1}^{n_j}(X_{ij} - \bar{X}_j) = 0$ (*j = 1,.., r*).

So, SS_W has *N – r* degrees of freedom. Note that the d.f.'s of the three terms are additive.

RESULTS AND DISCUSSION

Data was collected and analyzed from patients affected by uterine cancer who had tested during a period. The ANOVA can be used when variances are only approximately equal if the number of subjects in each group is equal and since we have equal sample sizes for all the groups we can assume that variances are approximately equal and proceed with one-way ANOVA.

Table 4:

Tested	Positive	% Positivity	95% CI
10613	1391	13.11	(0.125—0.138)
19260	3608	18.73	(0.182—0.193)
23345	4589	19.66	(0.192—0.202)
28622	5971	20.86	(0.204—0.213)
32127	4755	14.80	(0.144—0.152)

The ONE WAY ANOVA table for the above mentioned data is given as

Table 5:

Positivity	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3265.724	29	112.611	4.308	0.000
Within Groups	3137.036	120	26.142		
Total	6402.760	149			

Limitations of this study include the study of differences in incidence rates of urban and rural areas as the study is based on the data. There are still many unanswered questions that will require research.

CONCLUSION

Thus we have precisely discussed about the ovarian and uterine cancer. Methods and materials for the cancer are specified and results are drawn by mathematical calculations. It is found that one in 75 women is diagnosed by Ovarian Cancer and one in 60 women are likely to be diagnosed with uterine cancer by the age of 75. The majority of these women are over 50.

REFERENCES

1. Murthy, N.S., Chaudhry, K. and Rath, G.K. (2016) Burden of cancer projections, Indian scenario: Gaps in the availability of radiotherapy treatment facilities. Asian Pac J Cancer, pp 9-67.
2. Pramesh, C.S., Badwe, R.A., Borthakur, B.B., Chandra, M., Raj, E.H., Kannan, T. *et al*. (2014). Delivery of affordable and equitable cancer care in India. Lancet Oncol. pp 23-33.
3. Pramesh, C.S., Badwe, R.A. and Sinha, R.K. (2014). The national cancer grid of India. Indian J. Med. Pediatr. Oncol. pp 35:226-7.
4. International Agency for Research on Cancer, World Health Organization. GLOBOCAN 2012— estimated cancer incidence, mortality and prevalence worldwide.