International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 7 Issue 7 Ver IV || July 2018 || PP 66-73

Generalized Weibull Pareto Distribution To Analyze The Association Of Estrogen And Progesterone Hormone With Mammographic Density Phenotypes.

*****Sheeja K, ** Dr.S.Lakshmi,
(Research Scholar Mother Teresa Women's University Kodaikanal) Assistant Professor, SIES(Nerul) College Of Arts Science and Commerce, Navi Mumbai.
(Rtd) Principal, Kunthavai Naachiyaar Govt. Arts College for Women (Autonomous), Thanjavur – 7, Tamilnadu, South India
*Corresponding author: *Sheeja K*

Abstract:*A Generalized Weibull-Pareto Distribution is used to analyze the association of estrogen and progesterone hormone with mammographic density phenotypes in premenopausal women. We know that these are the key factor in development of breast cancer in women. The percent mammographic density and absolute mammographic density was obtained from digitized mammograms taken on a scheduled day of the menstrual cycle (day 7-12) among 202 healthy, premenopausal women (Energy Balance and Breast Cancer Aspects Study-I). In this study four parameter life model called Generalized Weibull-Pareto Distribution (GWPD) was proposed using exponentiated distribution. Some properties of the distribution have been discussed including survival, hazard, moments and moments generating function. The proposed distribution was applied to study the behavior of estrogen and progesterone and their association with mammographic density and their survival and hazard function was found and this distribution was found to be the best fit.*
Keywords: *17β-estodiol, mammographic density, progesterone, GWPD, Moments, Moments Generating Function, Survival and Hazard function.Mathematics Subject Classification:60E05, 62F10, 62N05, 62P99*

Date of Submission: 09-07-2018 Date of acceptance: 23-07-2018

I. Introduction:

There are several forms and extensions of the Pareto distribution in the literature. The probability density function (pdf) given as:

$$G(x;\theta,k) = 1 - \left(\frac{\theta}{x}\right)^k \quad \theta \le x < \infty; \theta, k > 0 \qquad (1)$$

Pickands [7] was the first to propose an extension of the Pareto distribution with the generalized Pareto (GP) distribution when analyzing the upper tail of a distribution function. One major problem in the development and application of probability distribution today is that a lot of available data sets do not follow any of the existing and well known standard probability distributions (models) and hence cannot be modeled appropriately. This creates room for developing new distributions which could better describe some of these phenomena and provide greater flexibility in modeling of lifetime data. Vifredo in 1896 defined Pareto distribution using the cumulative distribution function (cdf) as in equation (1) below:

$$g(x;\theta,k) = \frac{k\theta^k}{x^{k+1}} \quad \theta \le x < \infty;\ \theta, k > 0 \qquad (2)$$

Where $\theta > 0$ is a scale parameter, k>0 is a shape parameter and x is a variable.

This study seeks to increase the flexibility of the Pareto Distribution using the Generalized Weibull-G family of distribution proposed by Condeiro [1]. The derived distribution will be used to identify some basic properties which will be used to describe the new distribution (GWPD).

$$F(x) = \int_0^{-\log[1-G(x,\eta)]} \alpha\beta t^{\alpha-1} e^{-\alpha t^{\beta}} dt \qquad (3)$$

Integrating equation (3) yield,

$$F(x) = 1 - \exp\{-\alpha(-\log[1 - G(x;\eta)])^{\beta}\} \qquad (4)$$

Where $G(x;\eta)$ is the cdf of the baseline distribution which depends on a parameter vector η.
The pdf of the corresponding Generalized Weibull-G-family is obtained by differentiating equation (4) with respect to x given by

$$f(x) = \alpha\beta \frac{g(x;\eta)}{[1-G(x;\eta)} (-\log[1 - G(x;\eta])^{\beta-1}\exp\{-\alpha(-\log[1 - G(x;\eta)])^{\beta}\} \quad (5)$$

Where $g(x;\eta)$ and $G(x;\eta)$ are the pdf and cdf of the baseline distribution respectively which depend on parameter vector $\eta, \alpha > 0$ and $\beta > 0$ are the scale and shape parameters.

Generalized Weibull Distribution:

Let $G(x;\eta)$ be the Pareto distribution function from equation (1) with parameter θ and k, then equation (4) yields the Generalized Weibull-Pareto (GWPD) distribution function for $x \geq \theta$;

$$F(x;\alpha,\beta,\theta,k) = 1 - \exp\left\{-\alpha\left(-\log[1 - \left(1 - \left(\frac{\theta}{x}\right)^k\right)]\right)^{\beta}\right\} \quad (6)$$

Where $\theta > 0$ is a scale parameter and the other positive parameters k and β are the shape parameters. The corresponding density function is obtained by substituting equation (1) and (2) in (5) as:

$$f(x) = \frac{\alpha\beta k\theta^k}{\left[1-\left(1-\left(\frac{\theta}{x}\right)^k\right)\right]x^{k+1}}\left(-\log[1 - \left(1 - \left(\frac{\theta}{x}\right)^k\right)]\right)^{\beta-1}\exp\left\{-\alpha\left(-\log[1 - \left(1 - \left(\frac{\theta}{x}\right)^k\right)]\right)^{\beta}\right\} \quad (7)$$

1. Moments and Moments Generating Function:

The moments and generating function of the GWPD is obtained.

Let $f(x/\alpha,\theta,k)$ be a continuous random variable from the GWPD. The nth moment of x is defined as:

$$E(x^n) = \mu_n = \int_0^{\infty} x^n \left(\sum_{k=1}^{\infty} \frac{(-1)^k\alpha^k}{k!}\eta_r \frac{k\theta^k}{x^{k+1}}\left(1 - \left(\frac{\theta}{x}\right)^k\right)^{r-1}\right)dx \quad (8)$$

Moment Generating Function:

Let x be a continuous random variable, the moment generating function of x is defined as:

$$M_x(t) = E(e^{tx}) = \int_0^{\infty} e^{tx} f(x)dx \quad (9)$$

2. **Mathematical Model**:

Survival Function:

Survival function is the probability that a system or an individual will survive beyond a given time. Mathematically, the survival function is defined by:

S(x)=1-F(x) (10)

Where F(x) is the cdf of the GWPD

The survival function of the GWPD is given by:

$$S(x) = 1 - \left(1 - \exp\left\{-\alpha\left(-\log\left[1 - \left(1 - \left(\frac{\theta}{x}\right)^k\right)\right]\right)^{\beta}\right\}\right) \quad (11)$$

$$S(x) = \exp\left\{-\alpha\left(-\log\left[1 - \left(1 - \left(\frac{\theta}{x}\right)^k\right)\right]\right)^{\beta}\right\} \quad (12)$$

The survival function indicates the probability that a component or individual (y) will survive at time (x) for the GWPD.

Hazard function:

Hazard function is also called the failure or risk function and is the probability that a component will fail or die for an interval of time. The hazard function is defined as:

$$h(x) = \frac{f(x)}{1-F(x)} \quad (13)$$

where f(x) and F(x) are the pdf and cdf of the GWPD.

The hazard function of the GWPD is given by:

$$h(x) = \frac{\alpha\beta k\theta^k}{\left[\left(1-\left(\frac{\theta}{x}\right)^k\right)\right]x^{k+1}}\left(-\log[1 - \left(1 - \left(\frac{\theta}{x}\right)^k\right)]\right)^{\beta-1} \quad (14)$$

3. Application:

Women with higher levels of mammographically measured breast density have a significantly increased risk of developing breast cancer [6]. Absolute mammographic density reflects dense areas of the breast, hypothesized to be composed of epithelial and stromal tissues. Both percent mammographic density and absolute mammographic density are positively correlated with the number of epithelial cells at risk for malignant transformation. Although a high percent mammographic density is associated with a three-to-six-fold increase in the risk of breast cancer compared with a low percent mammographic density [6], the absolute dense area is considered to represent the actual target tissue for tumor development as ductal carcinoma in situ and invasive breast cancer more often occur in dense areas [9,8,4].

4.1 **Discussion**:

A total of 202 women aged 25-35 years were included for this study. Three scheduled visit was conducted over the course of one menstrual cycle. First visit days 1-5, early follicular phase. Second visit days 7-12, late follicular phase and third visit days 21-25 late luteal phase. Fasting serum samples were measured at three scheduled visits during menstrual cycle. Concentrations of estradiol and progesterone were measured using a direct immunometric assay. The women collected daily morning saliva samples at home for one entire menstrual cycle, and sampling started on the first day of menstrual bleeding [5,3,2]. Levels of 17β-estradiol were measured in daily saliva samples from 20 days (reverse cycle day -5 to -24) and levels of progesterone from 14 days (reverse cycle days -1 to -14).

4.2 Medical curves:

Figure 1:

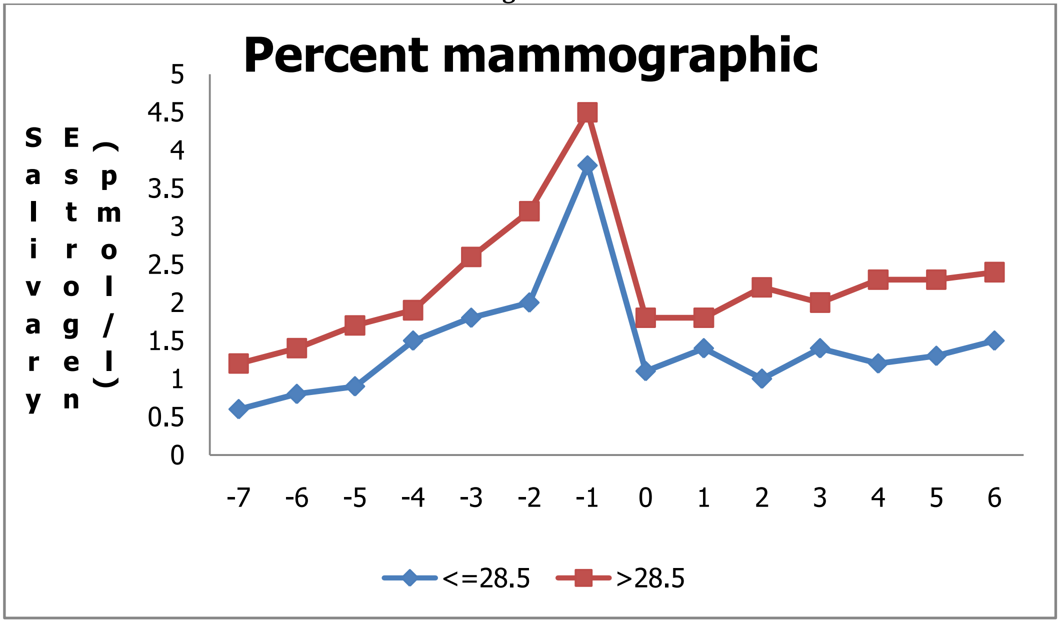


Figure 2:

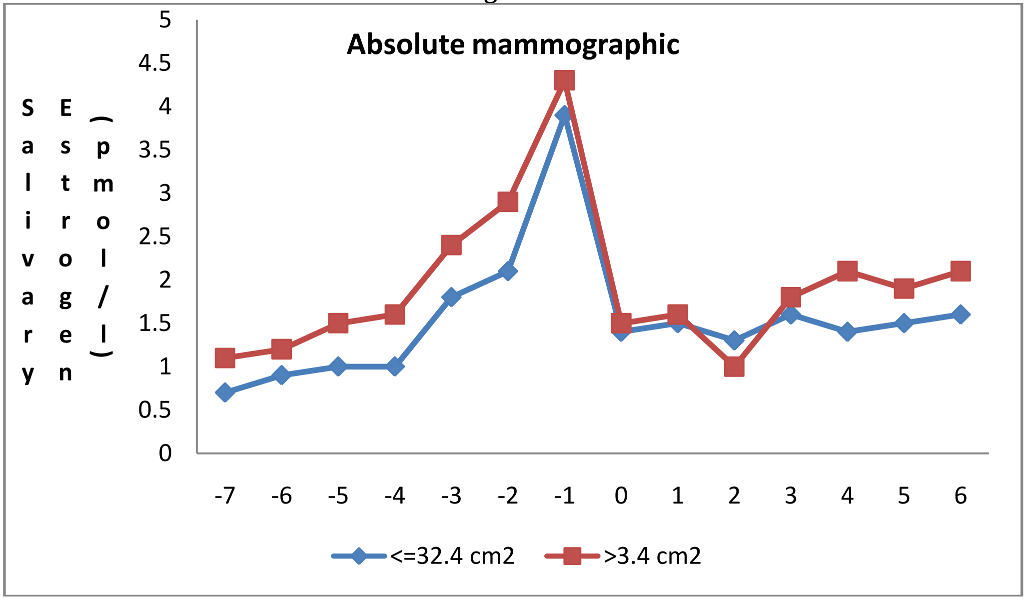


Figure 3:

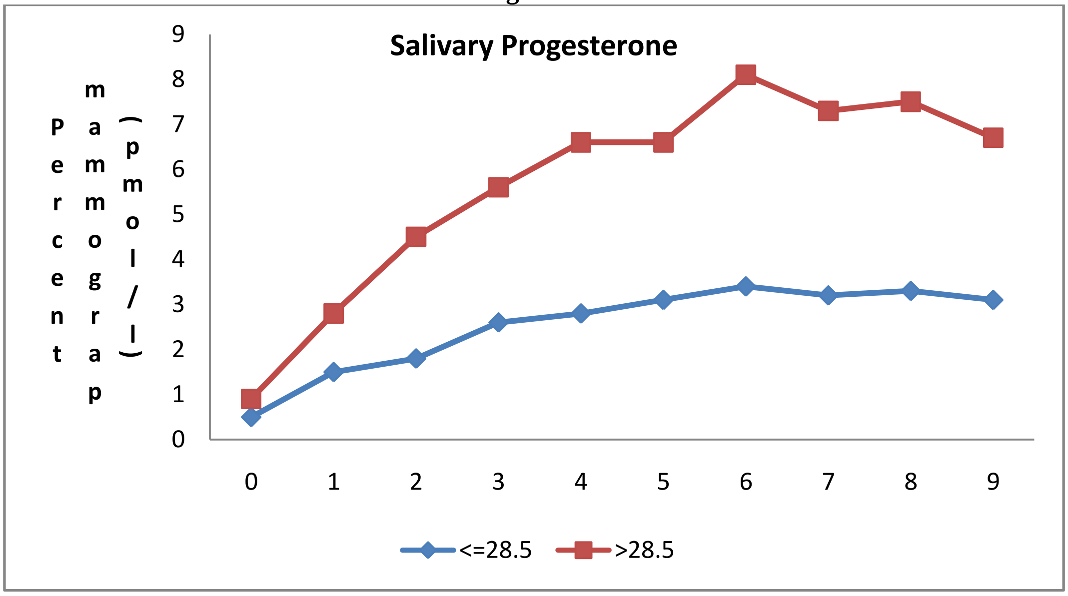


Figure 4:

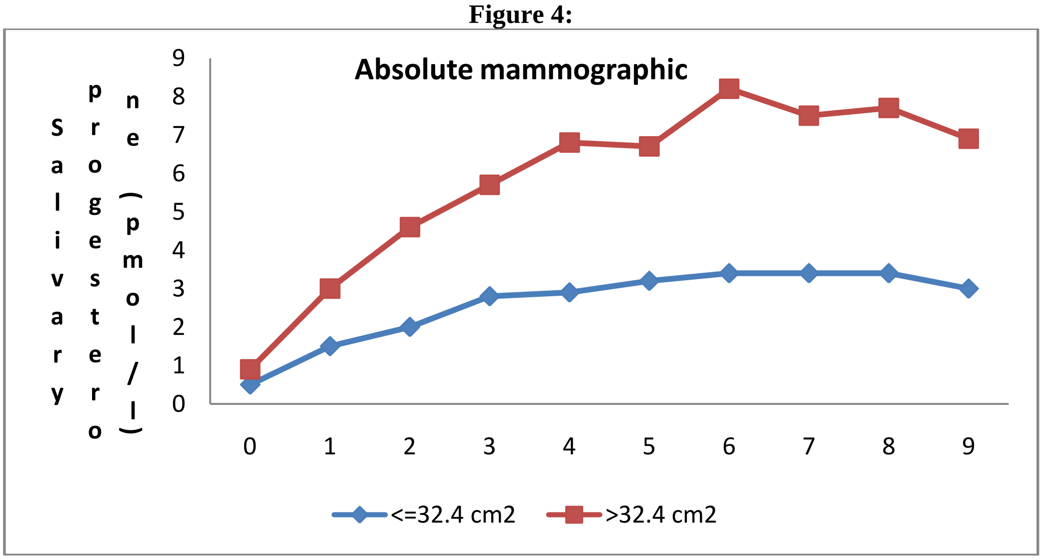


Figure1-4: Mean salivary ovarian hormones by high (red line) and low (blue line) percent mammagraphic density (Figure 1 and 3) absolute mammographic density (Figure 2 and 4) among 202 premenopausal women.

4.3 Mammographic density phenotypes:

The total breast area was defined on the mammographic image using a special outlining tool. The region of interest (ROI) was then outlined. Absolute mammographic density represents the number of the tinted pixels within ROI. Percent mammographic density is the ratio of absolute mammographic density to the total breast area multiplied by 100.

In this study of premenopausal women, positive associations were observed between salivary and serum estradiol and progesterone, and percent mammographic density, but no clear associations were observed between these hormones and other mammographic density phenotypes. Percent mammographic density was a stronger risk factor for breast cancer than absolute mammagraphic density.

5. Mathematical Results:

Figure 5. Estrogen (% mammograph and absolute mammograph)>=28.5% (Survival)

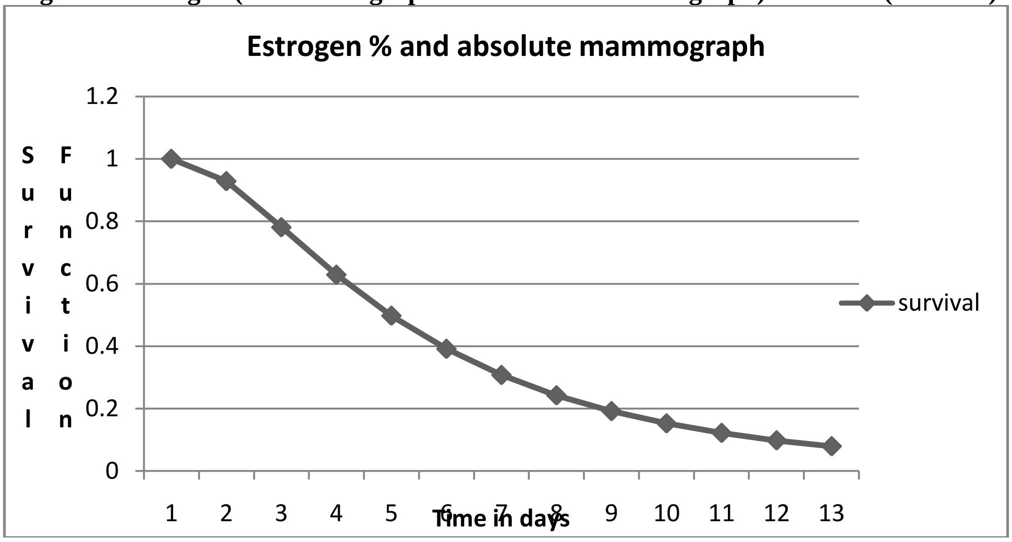


Figure 6. Estrogen (% mammograph and absolute mammograph)<=28.5% (Survival)

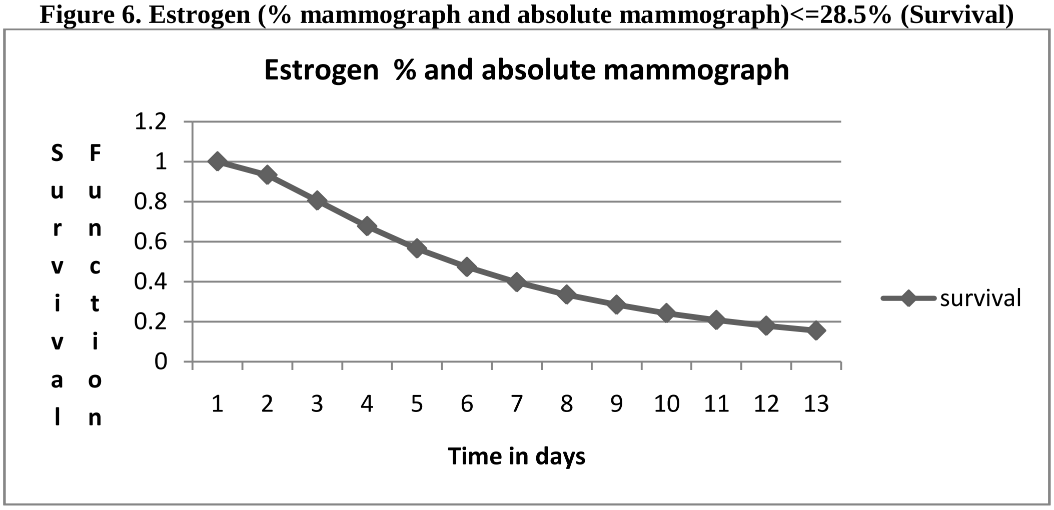


Figure 7. Progesterone (% mammograph and absolute mammograph)>=32.4% (Survival)

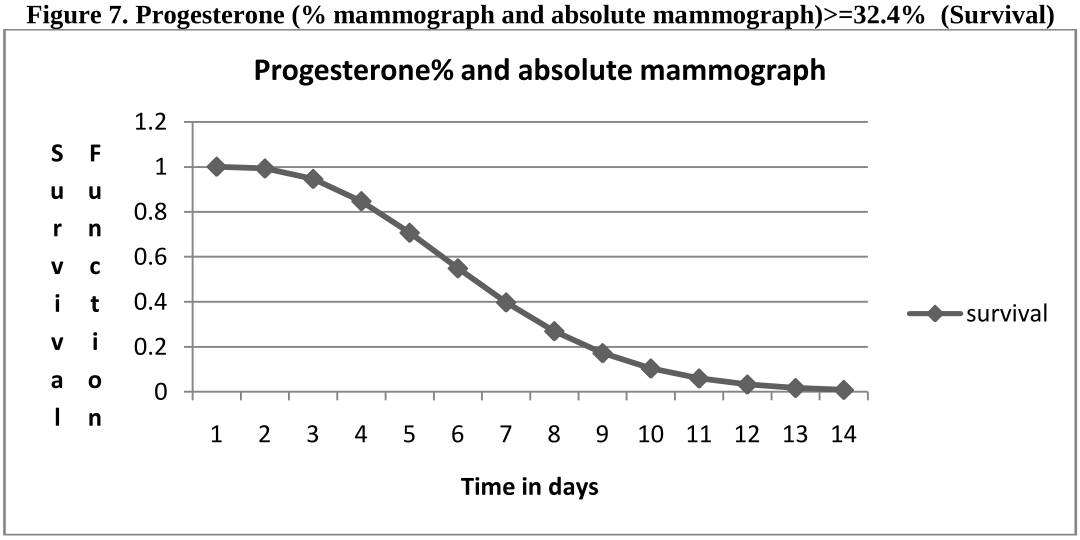


Figure 8. Progesterone (% mammograph and absolute mammograph)<=32.4% (Survival)

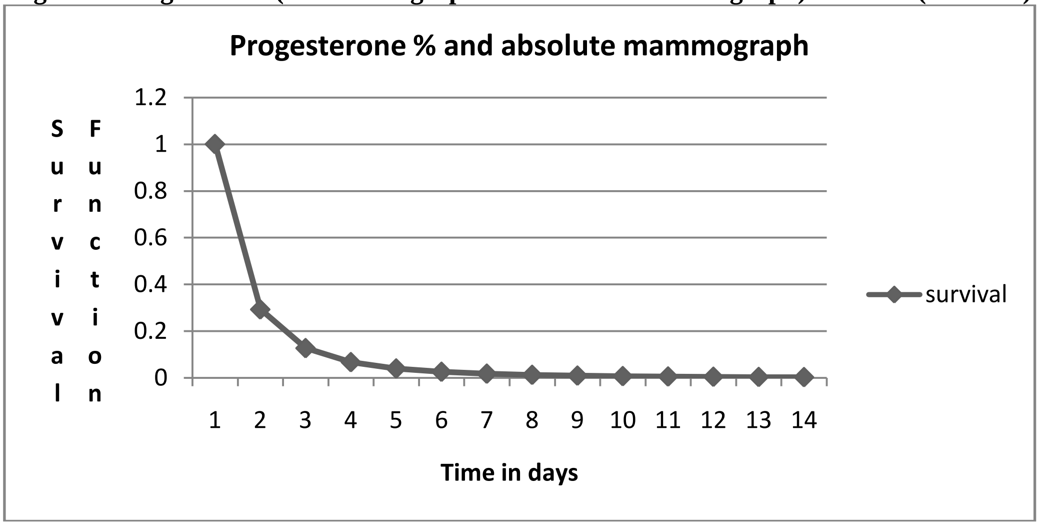


Figure 9. Estrogen (% mammograph and absolute mammograph)>=28.5% (Hazard)

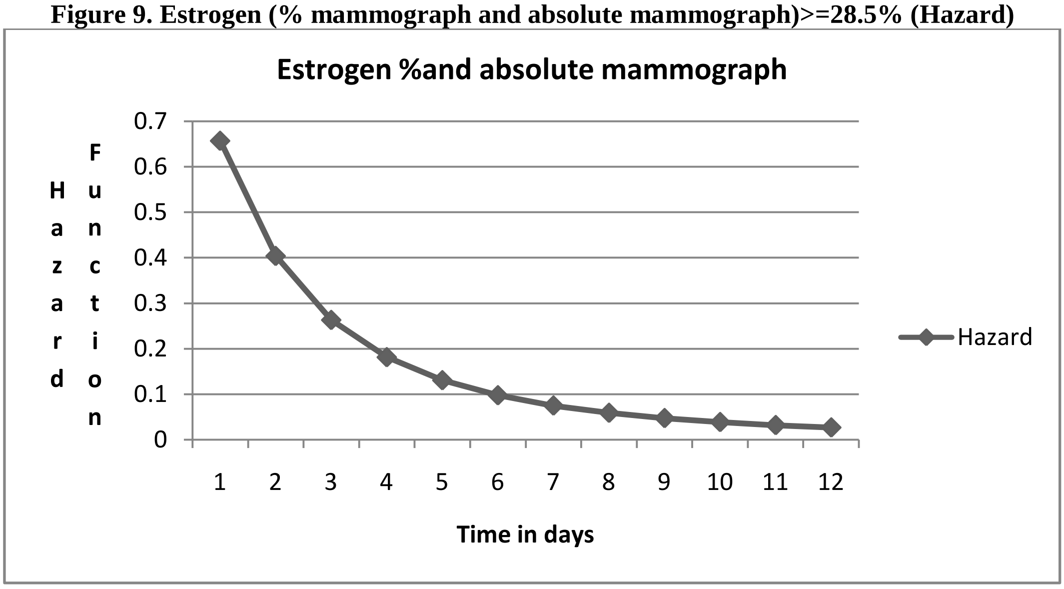


Figure 10. Estrogen (% mammograph and absolute mammograph)<=28.5% (Hazard)

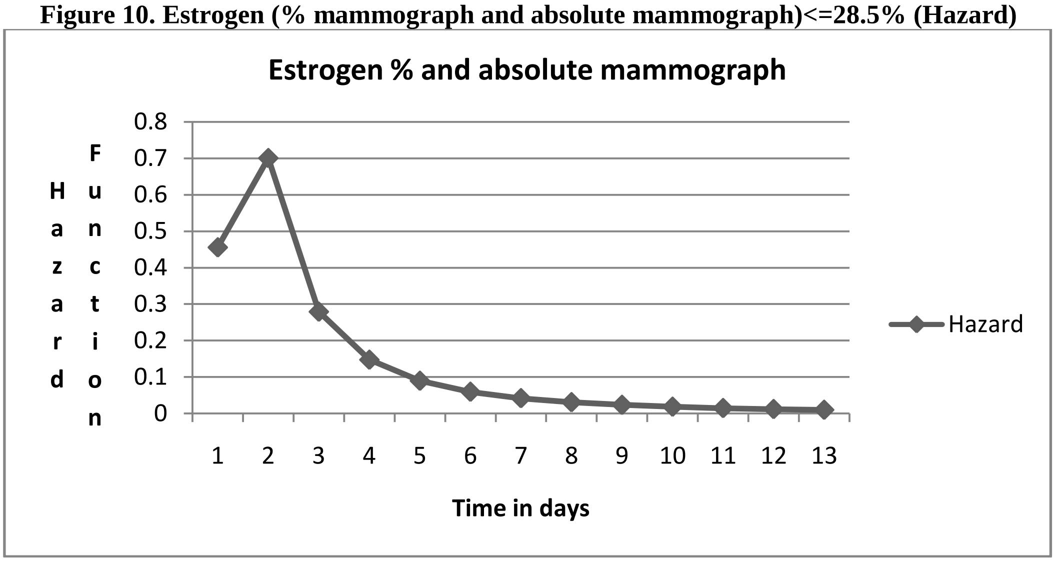


Figure 11. Progesterone (% mammograph and absolute mammograph)>=32.4%

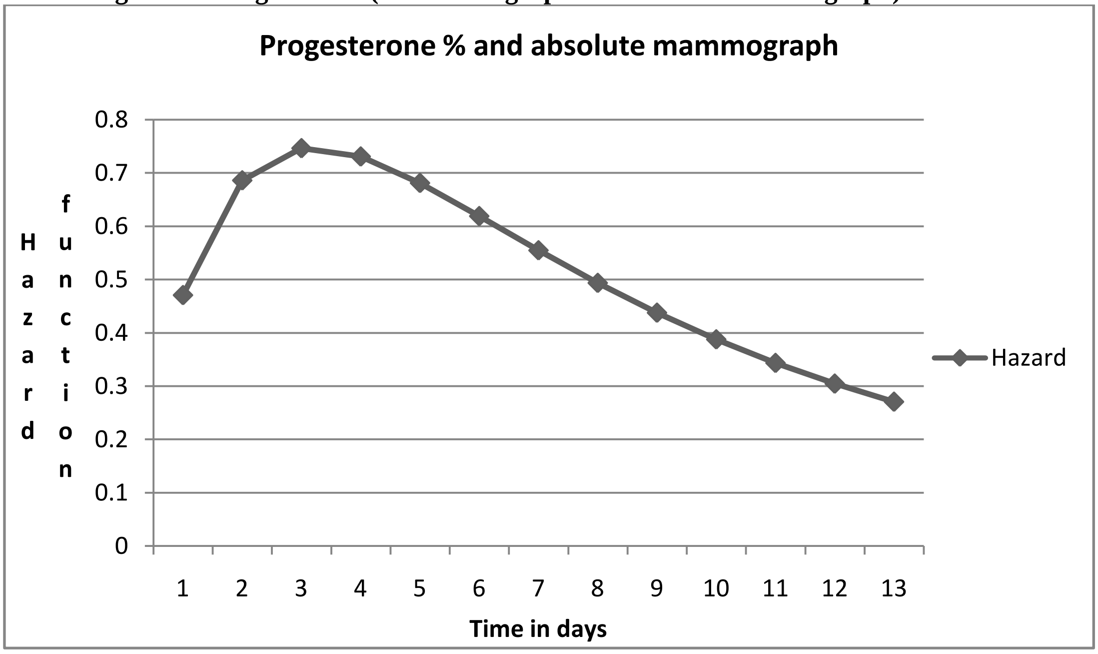
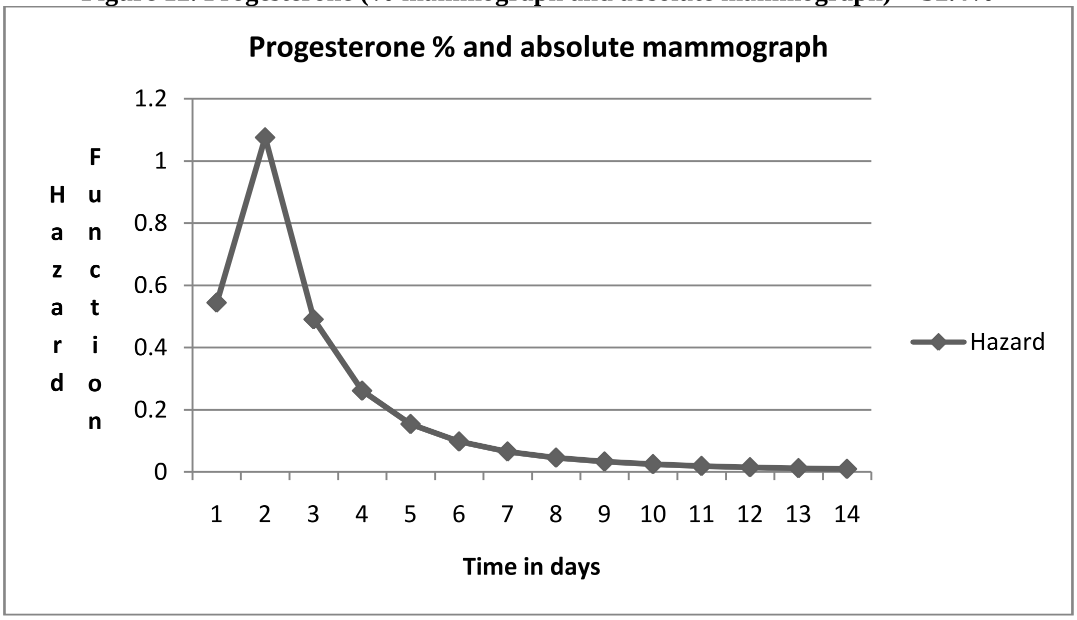


Figure 12. Progesterone (% mammograph and absolute mammograph)<=32.4%



II. Conclusion:

The present observations are biologically plausible, and may be of potential clinical interest. However, larger studies including estrogen and progesterone across the menstrual cycle in various populations are needed to define the clinical implications of these findings. The mathematical curves show a clear picture of studies for combined effect of percentage mammography and absolute mammography. They are monotonically decreasing curve for both survival of estrogen and progesterone (Figures 5,6,7,8). And the hazard function (Figures 9,10,11,12) shows a smooth continuous curve at different time points. The flexibility of this mathematical model in real data is seen. This distribution shows its usefulness to fit real data set.

References:

[1]. Codeiro, G.M., Ewin, M.M., Ortega, M. and Thiago G.R. (2015). A new generalized Weibull family of Distributions: Mathematical properties and applications: Journal of Statistical Distributions and Applications (2015) 2:1300110.1186/s40488-015-0036-6.

[2]. Ellison P.T., Lipson S.F., (1999). Salivary estradiol- a viable alternatives? Fertility Sterility 72:951-952.

[3]. Furberg A.S., Jasienska G, Bjurstam N., Torjesan P.A., Emaus A., Lipson S.F., et al.(2005). Metabolic and hormonal profiles: HDL cholesterol as a plausible biomarker of breast cancer risk. The Norwegian EBBA study. Cancer Epidemiol Biomakers Prev 14:33-40.

[4]. Gill J.K. Maskarinec G., Pagano I., Kolonel L.N. (2006).The association of mammographic density with ductal carcinoma in situ of the breast : the multiethnic Cohort. Breast Cancer Research 8:R30.
[5]. Lipson S.F., Ellison P.T. (1989). Development of protocols for the application of salivary steroid analysis to field conditions. American Journal Human Biology 1:249-255.
[6]. Pettersson A, Graff RE, Ursin G, Santos Silva I.D, McCormack V, Baglietto L, et al.(2014). Mammographic density phenotypes and risk of breast cancer:a meta analysis. Journal of National Cancer Institute 106, doi:10.1093/jnci/dju078.
[7]. Pickands, J. (1975). Statistical inference using extreme order statistics. Annals of statistics, 3:119-131.
[8]. Pinto Pereira S.M., McCormack V.A., Hipwell J.H., Record C, Wilkinson L.S.,Moss S.M. et al. (2011). Localized fibroglandular tissue as a predictor of future tumor location within the breast. Cancer Epidemiol Biomarkers Preview 20:1718-1725.
[9]. Ursin G, Hovanesian-Larsen L, Paaarisky Y.R,Pike M.C. Wu A.H. (2005).Greatly increased occurrence of breast cancers in areas of mammographically dense tissueeee. Breast Cancer Research 7:R605-R608.

Sheeja K "Generalized Weibull Pareto Distribution To Analyze The Association Of Estrogen And Progesterone Hormone With Mammographic Density Phenotypes. "International Journal of Engineering Science Invention (IJESI), vol. 07, no. 07, 2018, pp 66-73