Spiral Binomial and Related Distributions for Obsession to Abortion

Ramalingam Shanmugam

School of Health Administration, Texas State University-San Marcos, TX 78666, USA

ABSTRACT

Repeated abortion is considered an important health concern by the pregnant women and the governments. Educating the public about the impact of abortion becomes the government's responsibility. For this purpose, the government agencies turn to health and medical researchers for opinion about the impact of abortion. The public health researchers turn to data for clues and seek pertinent data information to answer several questions. One question is: Do repeated pregnancy-abortion data contain clue about a level of obsession? To answer this question based on data-clue, in this article, a new probability distribution is introduced and named it as Spiral Binomial Probability Distribution (SBPD). Various properties of SBPD and its connection to other spiral distributions are established and then utilized to assess whether the estimated obsession to abortion in a given data is significant and how much the obsession impacts likelihood of future pregnancy. When a woman experiences two or more pregnancies, her obsession to abortion is estimated using spiral binomial distribution of this article. The data about pregnancy versus abortion among Romans reveal that under repeated pregnancies (that is, $y = 2, 3, 4$), their obsession level to abortion changed to 0.02, 0.06 and 0.02. Furthermore, the correlation between the number of pregnancies and the number of abortions increased with an increased obsession level.

Keywords: Likelihood Ratio Test, Correlation, Regression, Conditional Probability Distribution, Weighted Sampling, Count Distribution, Marginal Distribution

1. INTERDICTION

Controversies exist about abortion. Nevertheless, abortion is a public health issue. Abortion is natural at times but is induced in other times for the sake of maternal safety, fetus' health, social or financial consequences. When a woman has repeated abortions, does she function with an obsession to abortion? In other words, do repeated abortion data contain a clue about obsession? How should the data analysis be done? These questions become the theme for discussion in this article. To answer these questions, this article first introduces a new probability distribution and names it Spiral Binomial Probability Distribution (SBPD). Various statistical properties of SBPD are derived and utilized to assess whether an estimated level of obsessive abortion in a given data is significant using likelihood ratio approach and to address how much it impacts future pregnancy?

Abortion has a long history. It occurred in ancient China (c. 2700 BCE), Egypt (c. 1550 BCE) and Roman Empire (Paul *et al.*, 2011). Abortion is an elective or voluntary at the request of the pregnant woman for medical or non-medical reasons. An induced abortion is performed to save life of a pregnant woman, to protect a fetus from a premature mortality, disability or other risks to unhealthy life. Clinical studies have proven that there is an association between abortion and breast cancer (Jasen, 2005) According to World Health Organization's report (WHO, 2009), about 42 million abortions occur in each year worldwide and 46% of them are unsafe. The unsafe abortions cause about 70,000 maternal deaths in each year and five million women become disabled in U.S. alone. Abortion related issues cannot be ignored. In many nations, the government spends enormous amount of money on abortion related public health issues. To deal with abortion and related issues, it is vital to configure how obsessive a woman might be to abortion if she has a history of repeated abortions? Sedgh *et al.* (2007) provides worldwide estimates of induced or repeated abortions. In this article, this question is

answered with SBPD, its connection to other spiral count distributions and their statistical properties. The concept and expressions are illustrated using Wilcox and Gladen (1982) pregnancy-abortion data among Romans.

Of course, the abortion is meaningful only when a woman is pregnant. The pregnancy itself is a random phenomenon. The number, Y≥0 of pregnancies a woman undergoes is a factor in discussions about abortion. The random variable, Y≥0 follows a Poisson Probability Distribution (PPD) in (1) below Eq. 1:

$$\begin{aligned} &\text{pr}[Y = y|\lambda] \\ &= e^{-\lambda}\lambda^{y} / y!; \\ &y = 0,1,2,...,\infty; \lambda > 0 \end{aligned} \quad (1)$$

In the absence of pregnancy (that is, Y = 0), the number, X of abortions is meaningless. The event X = 0 occurs with probability one when Y = 0 and this scenario is called *structural zero*. When Y≥1, the occurrence of an event X = 0 is recognized as *sampling zero*. With the PPD (1), the expected number, $E[Y|\lambda]$ and the variance var $[Y|\lambda]$ of pregnancies are equal to the pregnancy rate, λ. The survival function $S[Y \geq m|\lambda]$ of the PPD (1) has been expressed in terms of cumulative chi-squared probability, $S[Y \geq m|\lambda] = \text{Pr}[Y \geq m|\lambda] = \text{CDF}_{\chi^2_{2m df}}(2\lambda)$ with 2m degrees of freedom (df). The chi-squared distribution is extensively tabulated. The probability for the event Y = 0 to occur is Eq. 2:

$$S[Y = 0|\lambda] = 1 - \text{CDF}_{\chi^2_{2df}}(2\lambda) \quad (2)$$

A planned pregnancy usually ends up in a live birth. Under repeated but unplanned pregnancies, a tendency to abort intensifies and it is defined here as obsession to abortion. The usual binomial is not equipped to capture obsession to abortion. The obsession is not directly measurable and hence, it has to be dealt as a parameter. No probability distribution in the literature exists to capture or explain the latent tendency for obsessive abortion. An illustration of obsession is done in this article by induction of a parameter to the binomial probability distribution. Such a strengthened binomial probability distribution is named Spiral Binomial Probability Distribution (SBPD) because the obsession functions a spiral up or down towards a live-birth. The SBPD is versatile enough to illustrate all scenarios including the absence of obsession to abortion. The conditional as well as unconditional statistical distributions for the number of abortions, its expected mean, expected variance and survival function are derived in the article. The correlation between the number of pregnancies and the number of abortions is derived. The Maximum Likelihood Estimate (MLE) of the model parameters and the correlation are derived and applied to analyze pregnancy-abortion data (Wilcox and Gladen, 1982) that were collected for the city of Rome, Italy. A likelihood ratio based procedure, using Wald (1943) likelihood ratio approach, is developed to test whether the MLE of obsession to abortion is statistically significant? The statistical power of the hypothesis testing procedure is investigated and illustrated in this article.

1.1. Main Results: Spiral Binomial and Related Spiral Probability Distributions

Public health workers often ponder over whether the incidence of abortion has been increasing in society generally speaking or in certain socio, economic or religious community in particular? In other words, is there an obsession to abortion among those who have a history of repeated abortions? The aim of this article is to devise an approach to understand an obsession if it exists or its impacts on future pregnancy, as it has not been done, according to a literature search.

To be specific, let p≥0 be a parameter to describe the latent obsession level to abortion. When obsession increases, the odds for a live birth decrease. For a comprehension, let the odds for abortion be $(\frac{\theta}{1-\theta})$ in the absence (that is, p = 0) of obsession, where $0 < 1 - \theta < 1$ refers the chance for a live birth. Suppose that the odds increase to new level $(1+\rho)(\frac{\theta}{1-\theta})$ under an obsessive tendency p≥0 to abortion. Then, the conditional probability distribution of having $0 \leq x \leq y$ abortions and y-x live births in a given y pregnancies is Eq. 3:

$$\begin{aligned} &\text{pr}[X|Y = y, \rho, \theta] \\ &= \frac{(1+\rho x)}{(1+\rho y\theta)}\left(\frac{y!}{x!(y-x)!}\right)\theta^x(1-\theta)^{y-x} \\ &x = 0,1,2,...,y; y \geq 1; 0 < \theta < 1; \rho \geq 0 \end{aligned} \quad (3)$$

The probability mass function (3) is named here Spiral Binomial Probability Distribution (SBPD). The well-known binomial probability distribution is a special case of SBPD (3) with p = 0. The survival function $S[m|Y = y, \rho, \theta] = \text{pr}[X \geq m|Y = y, \rho, \theta]$.

Of the SBPD (3) could be expressed in terms of cumulative probability of F-distribution as in (4). That is Eq. 4:

$$S[m \mid Y = y, \rho, \theta]$$

$$= \frac{\begin{pmatrix} pr\{F_{2m,2(y-m+1)} \le \dfrac{(y-m+1)\theta}{m(1-\theta)}\} \\ +(\rho y \theta)pr \\ \{F_{2(m-1),2(y-m+2)} \le \dfrac{(y-m+2)\theta}{(m-1)(1-\theta)}\} \end{pmatrix}}{(1+\rho y \theta)} \quad (4)$$

The F-distribution is extensively tabulated and expression (4) is a help for computations. The conditional probability for the event $X = 0 \mid Y = y, \rho, \theta$ to occur is Eq. 5:

$$pr[X = 0 \mid Y = y, \rho, \theta]$$

$$= 1 - \frac{pr\{F_{2,2y} \le \dfrac{y\theta}{(1-\theta)}\}}{(1+\rho y \theta)} \quad (5)$$

The conditional expected number, $E[X \mid Y = y, \rho, \theta]$ of abortions, using SBPD (3), is Eq. 6:

$$E[X \mid Y = y, \rho, \theta]$$

$$= \sum_{x=0}^{y} x \frac{(1+\rho x)}{(1+\rho y \theta)} \left(\frac{y!}{x!(y-x)!} \right) \theta^x (1-\theta)^{y-x}$$

$$= \frac{1}{(1+\rho y \theta)} [y\theta \sum_{w=0}^{y-1} \left(\frac{(y-1)!}{w!(y-1-w)!} \right) \theta^w (1-\theta)^{y-1-w}$$

$$+ \rho\{y(y-1)\theta^2 \quad (6)$$

$$\sum_{w=0}^{y-2} \left(\frac{(y-2)!}{w!(y-2-w)!} \right) \theta^w (1-\theta)^{y-2-w}$$

$$+ y\theta \sum_{w=0}^{y-1} \left(\frac{(y-1)!}{w!(y-1-w)!} \right) \theta^w (1-\theta)^{y-1-w}\}]$$

$$= y\theta[1 + \frac{\rho(1-\theta)}{1+y\rho\theta}]$$

$$\approx y\theta[1 + \rho(1-\theta)].$$

The mean, $y\theta$ of the binomial probability distribution is a particular case of (6) with $p = 0$ for the absence of obsession to abortion. The expression (6) provides a clue that the number of pregnancies, y and the expected number of abortions, x increase or decrease together whether or not obsession to abortion exists. However, the obsession level p is catalytic to inflate the number of abortions. It is worth to quantify the correlation r_{xy} between them and it is done later in the article, after deriving necessary results. The conditional variance, var $[X|Y = y, p, \theta]$, using SBPPD (3), is Eq. 7:

$$var[X \mid Y = y, \rho, \theta]$$

$$= E[X(X-1) \mid Y = y, \rho, \theta]$$

$$+ E[X \mid Y = y, \rho, \theta]$$

$$\{1 - E[X \mid Y = y, \rho, \theta]\} \quad (7)$$

$$= y\theta[(1-\theta) + \frac{\rho(1-\theta)}{(1+y\rho\theta)}$$

$$\{(y-2)\theta + (\frac{1+y\rho\theta^2}{1+y\rho\theta})\}]$$

$$\approx y\theta(1-\theta)[1 + \rho\{1 + (y-2)\theta\}]$$

The variance, $y\theta$ (1-θ) of the binomial probability distribution is a particular case of (7) with $p = 0$ for the absence of obsession to abort. The conditional variance (7) increases in proportion to the number y of pregnancies and the obsession level, p to abortion. The obsession level p is seen again to be catalytic for more heterogeneity in abortions.

We now turn to discuss their correlation. Suppose correlation between the number of pregnancies and abortions is r_{xy}. To find it, the joint probability mass function pr [Y = y, X = x] of Y = y and X = x is needed and it is Eq. 8:

$$pr[Y = y, X = x]$$

$$= pr[y \mid \lambda]pr[X \mid Y = y, \rho, \theta]$$

$$= (\frac{(1+\rho x)}{(1+y\rho\theta)}) \frac{e^{-\lambda}\lambda^y \theta^x (1-\theta)^{y-x}}{x!(y-x)!}$$

$$\approx \frac{[1+\rho(x-y\theta)]e^{-\lambda}\lambda^y \theta^x (1-\theta)^{y-x}}{x!(y-x)!}; \quad (8)$$

$$y = x, x+1, \dots, \infty; x = 0, 1, 2, \dots;$$

$$0 < \theta < 1; \lambda > 0; \rho \ge 0$$

The expression (8) is named Here Bivariate Poisson-Spiral-Binomial Probability Distribution (BPSBPD). The product moment, E(XY) is obtained from (8) and it is:

$$E(XY) = \sum_{y=0}^{\infty} \sum_{x=0}^{y} xyPr[X = x, Y = y]$$

$$\approx \lambda(\lambda+1)\theta[1 + \rho(1-\theta)]$$

The covariance, cov (X, Y) between the number of pregnancies and the number of abortions is $cov(X, Y) \approx \lambda\theta[1 + \rho(1-\theta)]$. The marginal variances are var (Y) = λ and Eq. 9:

$$\begin{aligned}
&\text{var}(X) \\
&= E_Y \text{Var}_{X/y}(X/Y = y) \\
&\quad + \text{Var}_Y E_{X/y}(X/Y = y) \\
&= \lambda\theta[\{1 + \rho(1-\theta)\} \\
&\quad \{1 + \rho\theta(1-\theta)\} + \rho\theta(\lambda+\theta)] \\
&\approx \lambda\theta[1 + \lambda\rho\theta]
\end{aligned} \tag{9}$$

Hence, the correlation, r_{xy} in (10) is obtained and it is:

$$\begin{aligned}
r_{xy} &= \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\,\text{var}(Y)}} \\
&\approx \sqrt{\theta[\frac{1+\rho(1-\theta)}{1+\lambda\rho\theta}]}
\end{aligned} \tag{10}$$

Notice that the correlation is $\sqrt{\theta}$ in the absence (that is, $p = 0$) of obsession. In the presence of obsession to abortion, the correlation (10) increases. We now look at the behavior of the marginal probability distribution of the number of abortions.

Using (8), the unconditional probability distribution $\text{pr}[X|\rho,\theta,\lambda]$ for the number of abortions is obtained below and it is named Spiral Poisson Probability Distribution (SPPD). The marginal probability distribution of X is Eq. 11:

$$\begin{aligned}
&\text{pr}[X|\rho,\theta,\lambda] \\
&= \sum_{y=x}^{\infty} \text{pr}[X = x, Y = y] \\
&= [1 + \rho(1-\theta)(x-\lambda\theta)]e^{-\lambda\theta}(\lambda\theta)^x / x!; \\
&x = 0,1,2,...; 0 < \theta < 1; \lambda > 0; \rho \geq 0
\end{aligned} \tag{11}$$

With mean Eq. 12:

$$E[X|\rho,\theta,\lambda] = \lambda\theta\{1 + \rho(1-\theta)\} \tag{12}$$

And variance $\text{var}[X|\rho,\theta,\lambda]$ as specified in (9). The survival function $S[X \geq m|\lambda,\theta]$ of the SPPD (11) is Eq. 13:

$$\begin{aligned}
&\text{Pr}[X \geq m] \\
&= [1 + \rho\lambda\theta(1-\theta)\{CDF_{\chi^2_{2(m-1)df}}(2\lambda\theta) \\
&\quad -CDF_{\chi^2_{2mdf}}(2\lambda\theta)\}]
\end{aligned} \tag{13}$$

where, $CDF_{\chi^2_{2mdf}}(2\lambda\theta)$ denotes the chi-squared distribution function up to $2\lambda\theta$ with 2mdf. The chi-squared

distribution is extensively tabulated and expression (13) makes computations easier. The probability for the event $X = 0$ to occur is $E[X=0|\rho,\theta,\lambda] = \rho\lambda\theta(1-\theta)CDF_{\chi^2_{2df}}(2\lambda\theta)$ marginally speaking when information about the number of pregnancies is non-available.

However, the conditional probability distribution $\text{pr}[Y|X = x, \rho, \theta, \lambda]$ of the number Y of pregnancies for a given number $X = x$ of abortions is Eq. 14:

$$\begin{aligned}
&\text{pr}[Y = y | X = x, \rho, \theta, \lambda] \\
&= \frac{\text{pr}[Y = y, X = x]}{\text{pr}[X = x]} \\
&= [\frac{1 + \rho(x - y\theta)}{1 + \rho(1-\theta)(x - \lambda\theta)}] \\
&\frac{e^{-\lambda(1-\theta)}[\lambda(1-\theta)]^{y-x}}{(y-x)!}; \\
&y = x, x+1,...; x \geq 0; \\
&0 < \theta < 1; \rho \geq 0
\end{aligned} \tag{14}$$

Which is named Spiral Displaced Poisson Probability Distribution (SDPPD). The survival function $S[Y \geq m | X = x, \rho, \theta, \lambda]\pi$ of the SDPPD (14) is Eq. 15:

$$\begin{aligned}
&S[Y \geq m | X = x, \rho, \theta, \lambda] \\
&= [\frac{\begin{array}{l}(1 + \rho x)CDF_{\chi^2_{2mdf}}(2\lambda[1-\theta]) \\ -\rho\theta\{x CDF_{\chi^2_{2(m-x)df}}(2\lambda[1-\theta]) \\ + \lambda(1-\theta)CDF_{\chi^2_{2(m-x-1)df}}(2\lambda[1-\theta])\}\end{array}}{1 + \rho(1-\theta)(x - \lambda\theta)}]
\end{aligned} \tag{15}$$

where, $CDF_{\chi^2_{2mdf}}(2\lambda\theta)$ denotes the chi-squared distribution function up to $2\lambda\theta$ with 2 m df. The probability for the event $Y = 0$ to occur is therefore Eq. 16:

$$\begin{aligned}
&S[Y = 0 | x, \rho, \theta, \lambda] \\
&= 1 - [\frac{(1 + \rho x)CDF_{\chi^2_{2df}}(2\lambda[1-\theta])}{1 + \rho(1-\theta)(x - \lambda\theta)}]
\end{aligned} \tag{16}$$

The conditional expected mean number (17) and variance (18) of pregnancies are respectively:

$$\begin{aligned}
&E[Y | X = x, \rho, \theta, \lambda] \\
&= [x + \lambda(1-\theta)] \\
&[1 - \frac{\lambda\rho\theta(1-\theta)}{1 + \rho\theta(1-\theta)(x-\lambda\theta)}] \\
&\approx [x + \lambda(1-\theta)][1 - \lambda\rho\theta(1-\theta)]
\end{aligned} \tag{17}$$

And:

$$\text{var}[Y|X = x, \rho, \theta, \lambda]$$
$$= \lambda(1 - \theta)[1 - \rho\theta\{1 + \rho\lambda\theta(1 - \theta)$$
$$\{x + \lambda(1 - \theta)\}\}] \tag{18}$$

The correlation (10) captures the linear relation between the number of pregnancies and the number of abortions. With the correlation, one could make a linear projection (19) of number of future pregnancies, $\hat{Y}$ for a known number of abortions, x. Likewise, one could project the number (20) of future abortions, $\hat{X}$ for a given number of pregnancies y. The projection equation for the number of future pregnancies is Eq. 19:

$$\hat{Y} = E[Y|\lambda] + r_{xy}\sqrt{\frac{\text{var}[Y|\lambda]}{\text{var}[X|\rho, \theta, \lambda]}}$$
$$(x - E[X|\rho, \theta, \lambda]) \tag{19}$$
$$\approx a + bx$$

With an initial effect Eq. 20:

$$a = \lambda[1 - \frac{\theta\{1 + \rho(1 - \theta)\}\sqrt{1 + \rho(1 - \theta)}}{(1 + \lambda\rho\theta)}] \tag{20}$$

And the slope Eq. 21:

$$b = \frac{\sqrt{1 + \rho(1 - \theta)}}{(1 + \lambda\rho\theta)} \tag{21}$$

The projection equation for the number of future abortions is Eq. 22:

$$\hat{X} = E[X|\rho, \theta, \lambda] + r_{xy}\sqrt{\frac{\text{var}[X|\rho, \theta, \lambda]}{\text{var}[Y|\lambda]}}$$
$$(y - E[Y|\lambda]) \tag{22}$$
$$\approx c + dy$$

With an initial effect Eq. 23:

$$c = \lambda\theta[1 + \rho(1 - \theta) + \sqrt{1 + \rho(1 - \theta)}] \tag{23}$$

And the slope Eq. 24:

$$d = \theta\sqrt{1 + \rho(1 - \theta)} \tag{24}$$

Now, we discuss to estimate the parameters of the joint probability mass function BPSBPD (8). For this purpose, suppose that a bivariate random sample $(x_1, y_1), (x_2, y_2)....., (x_n, y_n)$ is available. The Maximum Likelihood Estimate (MLE) is selected because of its invariance property. The MLE of λ is $\hat{\lambda} = \bar{y}$. To derive other MLEs, we consider the conditional log likelihood function Eq. 25:

$$\ln L[X|Y = y, \rho, \theta]$$
$$= \sum_{i=1}^{n} \ln(1 + \rho x_i) - n\ln(1 + \rho y\theta)$$
$$+ \sum_{i=1}^{n} \ln\left(\frac{y!}{x_i!(y - x_i)!}\right) \tag{25}$$
$$+ n\bar{x}\ln\theta + n(y - \bar{x})\ln(1 - \theta)$$
$$\approx n(\bar{x} - y\theta)(\rho - 1)$$

By solving the estimating equations $\partial_\theta \ln L[X|Y = y, \rho, \theta] = 0$ and $\partial_\rho \ln L[X|Y = y, \rho, \theta] = 0$, using their Taylor series expansions, the MLE (26) and (27) are obtained, where ∂_z denotes the derivative with respect to z. Note that Eq. 26 and 27:

$$\hat{\theta} = \frac{\bar{x}}{y}[1 + (\bar{x} - 1)\hat{\rho}]$$
$$= \frac{\bar{x}}{y}[1 + (\bar{x} - 1)\{\bar{x}(1 - \frac{\bar{x}}{y}) - s^2\}] \tag{26}$$

And:

$$\hat{\rho} = \left|\bar{x}(1 - \frac{\bar{x}}{y}) - s^2\right| \tag{27}$$

When $s^2 = \bar{x}(1 - \frac{\bar{x}}{y})$, notice that the MLE (27) and (26) reduce to $\hat{\rho} = 0$ and $\hat{\theta}_0 = \frac{\bar{x}}{y}$ which is the MLE of binomial parameter under no obsessive abortion. In a particular maternity case, one might wonder: Is the MLE $\hat{\rho}$ of obsession to abortion negligible? An answer to this question requires a hypothesis testing methodology. For this purpose, the Wald (1943) likelihood ratio approach is resorted. The Wald's approach is a powerful methodology. In our context, to test the null hypothesis $H_o : \rho = 0$ against an alternative hypothesis $H_1 : \rho = \rho^* \neq 0$ using log-likelihood ratio test statistic Eq. 28:

$$-\ln\Lambda_{\rho=0}$$
$$= \ln L(\hat{\rho},\hat{\theta}) - \ln L(\rho=0,\hat{\theta}_0)$$
$$= n\overline{x}(\overline{x}-1) \qquad (28)$$
$$\left|\overline{x}(1-\frac{\overline{x}}{y})-s^2\right|\left|\overline{x}(1-\frac{\overline{x}}{y})-s^2-1\right|$$

which follows a non-central chi-squared probability distribution with one degrees of freedom (df) and the non-centrality parameter $\hat{\delta}_{\rho=0} = \hat{\rho}/\hat{v}\hat{a}r(\hat{\rho})$, where $\hat{v}\hat{a}r(\hat{\rho})$ is the MLE of diagonal element in the inverse of the variance-covariance matrix of the MLEs. Stuart and Ord (1994) for properties of the non-central chi squared probability distribution. Recall that the MLE of variance-covariance matrix which is the inverse of:

$$\hat{I} = \begin{bmatrix} \hat{a} & \hat{b} \\ \hat{b} & \hat{c} \end{bmatrix} = \begin{bmatrix} -\hat{E}(\partial^2_{\theta\theta}\ln L) & -\hat{E}(\partial^2_{\rho\theta}\ln L) \\ -\hat{E}(\partial^2_{\theta\rho}\ln L) & -\hat{E}(\partial^2_{\rho\rho}\ln L) \end{bmatrix}$$

where, $\hat{a} \approx \dfrac{ny[1+\rho(1-\theta)]}{\theta(1-\theta)}$, $\hat{b} \approx ny[1-y\rho\theta]]$ and

$\hat{c} \approx ny\theta(1-\theta)$. The determinant of the matrix $\hat{I}$ is $\hat{D} = |\hat{I}| = \hat{a}\hat{c} - \hat{b}^2 \approx n^2\rho y^3\theta$. Note that $var(\hat{\rho}) \approx \dfrac{[1+\rho(1-\theta)]}{n\rho y^2\theta}$.

Hence, under null hypothesis, $H_o : \rho = 0$, note that $\hat{\theta} = \dfrac{\overline{x}}{y}$ and:

$$\hat{\delta}_{\rho=0} = \hat{\rho}/\hat{v}\hat{a}r(\hat{\rho}|H_0)$$
$$\approx \frac{[1+\hat{\rho}(1-\hat{\theta}_0)]}{ny^2\hat{\theta}_0}$$
$$\approx \frac{[y+\hat{\rho}(y-\overline{x})]}{ny^2\overline{x}}$$

It is known that the non-central chi squared probability distribution with one df and non-centrality parameter δ approximately follows $(1+\dfrac{\delta}{1+\delta})$ times a central chi squared probability distribution with $\dfrac{(1+\delta)^2}{(1+2\delta)}$ df (Stuart and Ord, 1994) for details of this equivalence).

This suggests that the null hypothesis $H_o : \rho = 0$ could be rejected in favor of an alternative hypothesis $H_o : \rho > 0$ if $-\ln\Lambda_{\rho=0} > (1+\dfrac{\hat{\delta}_{\rho=0}}{1+\hat{\delta}_{\rho=0}})\chi^2_{\frac{(1+\hat{\delta}_{\rho=0})^2}{(1+2\hat{\delta}_{\rho=0})}df,\alpha}$ where the

right side is the critical value based on the $100(1-\alpha)^{th}$ percentile of the central chi squared probability distribution with $\dfrac{(1+\hat{\delta}_{\rho=0})^2}{(1+2\hat{\delta}_{\rho=0})}$ df and a significance level $\alpha \in (0,1)$. We now write the p-value (29) for rejecting the null hypothesis in favor of an alternative hypothesis and it is Eq. 29:

$$p - value$$
$$= Pr[\chi^2_{\frac{(1+\hat{\delta}_{\rho=0})^2}{(1+2\hat{\delta}_{\rho=0})}df} < \frac{-\ln\Lambda_{\rho=0}}{(1+\frac{\hat{\delta}_{\rho=0}}{1+\hat{\delta}_{\rho=0}})}] \qquad (29)$$

The statistical power of the test statistic (28) can be calculated with a selection of a specific value for ρ^* in the alternative hypothesis. The statistical power is the probability of rejecting the null hypothesis $H_o : \rho = 0$ in favor of an alternative hypothesis $H_1 : \rho = \rho^* \neq 0$. Under the alternative hypothesis, the minus log likelihood ratio Eq. 30:

$$-\ln\Lambda_{\rho^*}$$
$$= \ln L(\hat{\rho},\hat{\theta}) - \ln L(\rho=\rho^*,\hat{\theta}_{\rho=\rho^*})$$
$$= n\overline{x}(\overline{x}-1) \qquad (30)$$
$$\{\left|\overline{x}(1-\frac{\overline{x}}{y})-s^2\right|\left|\overline{x}(1-\frac{\overline{x}}{y})-s^2-1\right|$$
$$+\rho^*(\rho^*-1)\}$$

where, $\hat{\theta}_{\rho=\rho^*} = \hat{\theta}_1 = \dfrac{\overline{x}}{y}[1+\rho^*(\overline{x}-1)]$. The result (30) follows a non-central chi-squared probability distribution with one df and non-centrality parameter $\hat{\delta}_{\rho^*} = \hat{\rho}/\hat{v}\hat{a}r(\hat{\rho}|H_1)$

$$\approx \frac{[1+\hat{\rho}(1-\hat{\theta}_1)]}{ny^2\hat{\theta}_1}.$$
$$\approx \frac{y(1+\hat{\rho}) - \hat{\rho}\hat{x}\{1+\rho^*(\overline{x}-1)\}}{ny^2\overline{x}\{1+\rho^*(\overline{x}-1)\}}$$

The non-central chi squared probability distribution with one df and non-centrality parameter $\hat{\delta}_{\rho^*}$ is approximately $(1+\dfrac{\hat{\delta}_{\rho^*}}{1+\hat{\delta}_{\rho^*}})$ times a central chi squared score with $\dfrac{(1+\hat{\delta}_{\rho^*})^2}{(1+2\hat{\delta}_{\rho^*})}$ df. The power (31) is the probability

of accepting a *true* alternative hypothesis $H_1 : \rho = \rho^*$ That is Eq. 31:

$$
\text{Power}
= \Pr[\chi^2_{\frac{(1+\hat{\delta}_{\rho^*})^2}{(1+2\hat{\delta}_{\rho^*})}df}
$$
(31)

$$
< \frac{(1+\frac{\hat{\delta}_{\rho=0}}{1+\hat{\delta}_{\rho=0}})(-\ln \Lambda_{\rho^*})\chi^2_{\frac{(1+\hat{\delta}_{\rho=0})^2}{(1+2\hat{\delta}_{\rho=0})}df,\alpha}}{(1+\frac{\hat{\delta}_{\rho^*}}{1+\hat{\delta}_{\rho^*}})(-\ln \Lambda_{\rho=0})}].
$$

The next section illustrates the results using Wilcox and Gladen (1982) data on Roman's abortion.

1.2. Illustration using Pregnancy and Abortion Data among Romans

The **Table 1** provides data on number of pregnancies and number of abortions among Romans as reported in Wilcox and Gladen (1982). There were 13433 cases that had one or more pregnancies. Out of which, 8198 women who had repeated pregnancies but did not have abortion. The pregnancy rate is $\hat{\lambda} = 2.5$. The observed average number of abortions is $\bar{x} = 0.48$. The abortion rate is not less. The chance for an abortion is $\hat{\theta} = 0.12$ which is quite alarming. Their average obsession to abortion is $\hat{\rho} = 0.03$. The likelihood ratio test statistic (28) is 96 and it is statistically significant as its p-value 0.001 is small. When the p-value is small, the null hypothesis $H_o : \rho = 0$ is rejected in favor of an alternative hypothesis $H_1 : \rho \neq 0$. The power (that is the probability of accepting a true alternative hypothesis $H_1 : \rho^* = 0.5$) is 0.99. The correlation between pregnancy and abortion would have been $\sqrt{\theta} = 0.34$ in the absence (that is, p = 0) of obsession but the correlation has increased to 0.36 due to obsession to abort. With an increasing obsession, the risk for abortion does increase (**Fig. 1**).

The projection of future number, $\hat{Y}$ of pregnancies is done using $\hat{Y} \approx 2.2 + x$ for a given number, x of abortions. The projection of future number, $\hat{X}$ of abortions is done using $\hat{X} \approx 0.6 + 0.12y$ for a given number, x of abortions. The chance for an abortion to occur is 0.08 without knowing about the number of pregnancies but is 0.01, 0.03, 0.05 or 0.08 corresponding to one, two, three or four pregnancies.

Table 1. Y= # of pregnancies versus X = # of abortions among Romans

$\frac{x \rightarrow}{y \downarrow}$	0	1	2	3	4	Sum	$\Pr(x=0\|y)$
1	2651	417				3068	0.01
2	2029	1112	106			3246	0.03
3	2559	1599	279	42		4479	0.05
4	960	1092	496	72	20	2640	0.08
Sum	8198	4220	881	114	20	13433	
$\Pr[y=0\|x=0]$	1	0.11	0.22	0.3	0.4		
Summary	$\bar{x} = 0.48$	$s_x^2 = 0.45$	$\bar{y} = 2.5$	$s_y^2 = 1.1$			
Estimates	$\hat{\rho} = 0.03$	$\hat{\theta} = 0.12$	$\hat{r} = 0.35$	$a = 2.2$	$b = 1$	$c = 0.6$	
	$d = 0.12$	$-\ln\Lambda_{\rho=1} 96$	$-\ln\Lambda_{\rho^*} = 742$	$\hat{\delta}_{\rho=0} = 0$	$\hat{\delta}_{\rho^*=1} = 0$	pValue = 0.001	
	power = 0.999	$\Pr(x=0) = 0.082$	$PR(x=0\|Y=0) = 1$				

Table 2. Obsession, p-value, power, unconditional and conditional probability (X = # abortions, Y = # pregnancies)

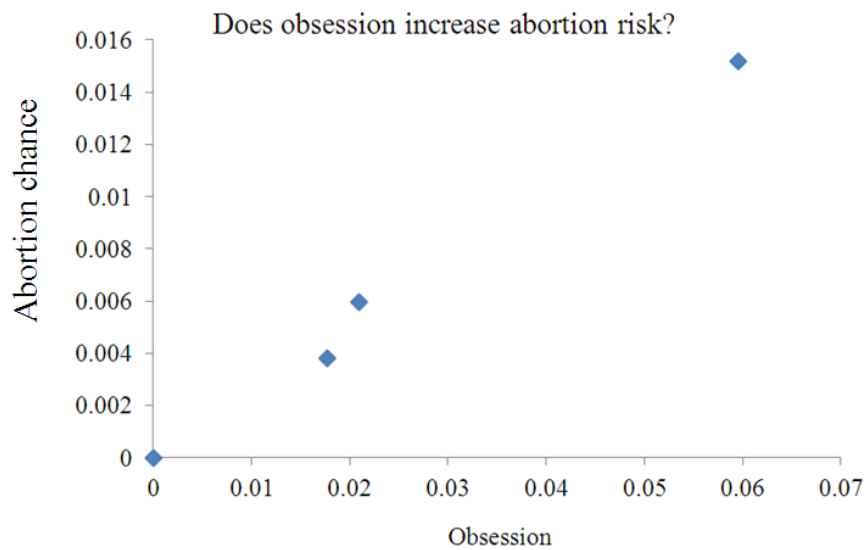| y	$\bar{x}\|y$	$s_{x\|y}^2$	$\hat{\rho}$	$\hat{\theta}$	Corr	$E(x\|y)$	pValue	power	$\Pr(x=0)$	$\Pr(x=0\|y)$
1	0.14	0.12	0	0.14	0.37	1	0.91	0.99	0.000	0.127
2	0.41	0.31	0.02	0.2	0.46	2.02	0.00	0.99	0.004	0.340
3	0.53	0.38	0.06	0.17	0.41	3.11	0.00	0.99	0.015	0.423
4	0.9	0.72	0.02	0.22	0.47	4.05	0.03	0.99	0.006	0.607

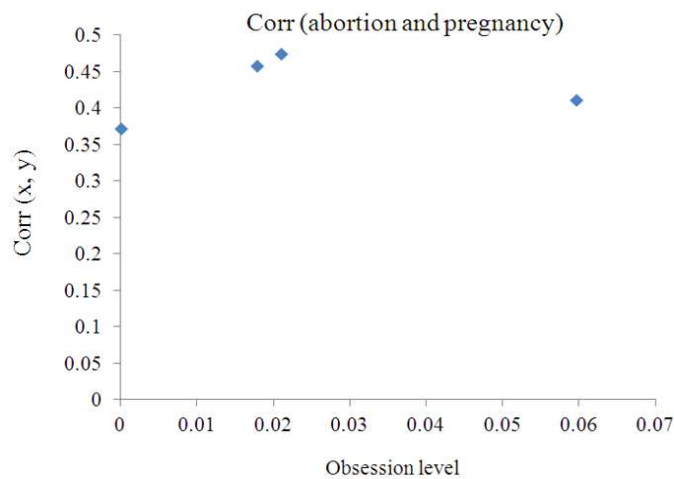**Fig. 1.** Obsession versus abortion risk

Fig. 2. Sketch of r_{xy}

The scenarios are different in a micro level according to the information in **Table 2**. When there is one pregnancy, the information is not enough to capture or articulate the obsession to abortion and hence, the obsession level is zero by definition. Under repeated pregnancies (that is, y = 2, 3, 4), the obsession level increased correspondingly to 0.02, 0.06 and 0.02. This is further substantiated by the consistent increase (**Fig. 2**) in the correlation between the pregnancy and abortion with an increase in obsession.

2. CONCLUSION

We point out that with an increasing obsession to abortion, the chance for a pregnancy to end up in a live birth decreases. Also, under repeated pregnancies, the obsession to abortion increases. With the likelihood ratio test statistic in this article, it is possible to test whether the MLE of obsession in a group of cases is statistically significant. A future research work is necessary to bring in the regression concepts and tools to sort out which are significant causal co-factors in building up the obsession.

New data with covariates proving information about those factors are needed and the public health professionals should collect data for future research work.

3. REFERENCES

Jasen, P., 2005. Breast cancer and the politics of abortion in the United States. Medical Hist., 49: 423-44. PMID: 16562329

Paul, M., S. Lichtenberg, L. Borgatta, D.A. Grimes and P.G.S. Creinin *et al*., 2011. Management of Unintended and Abnormal Pregnancy. 1st Edn., John Wiley and Sons, I Oxford, United Kingdom, SBN-10: 1444358472, pp: 392.

Sedgh, G., S. Henshaw, S. Singh, E. Åhman and I. Shah, 2007. Induced abortion: Estimated rates and trends worldwide. Lancet, 370: 1338-4135. PMID: 17933648

Stuart, A. and J.K. Ord, 1994. Kendall's Advanced Theory of Statistics. 6th Edn., Hodder Arnold, London, ISBN-10: 0340614307, pp: 704.

Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans. Am. Math. Soc., 54: 426-482.

WHO, 2009. The world health report 2008-primary health care (now more than ever). World Health Organization, Geneva.

Wilcox, A.J. and B.C. Gladen, 1982. Spontaneous abortion: The role of heterogeneous risk and selective fertility. Early Hum. Dev., 7: 165-178. PMID: 7151728