Statistical Models for Age of Onset of Mental Disorders

Dibyajyoti Bora[1] and Kishore K Das[2]

[1]Department of Statistics, Cotton University
Guwahati 781001, Assam, India
[2]Department of Statistics, Gauhati University
Guwahati 781014, Assam, India

Abstract

Objectives: We have estimated the Age of onset of Mental Disorders using suitable probability model.

Methodology: Data have been collected from patients suffering from Mental Disorders aged 21 years through 60 years. A self prepared and some standard questionnaires are used for data collection. Age of onset of the Disorder has been received from the personal registration file of the patients, recorded by the concerned psychiatrists.

Results: For both genders, the Age of onset of the Disorder has been found positively skewed and the Log normal distribution gives the best fitting model of the onset age.

Conclusion: The probability distribution of Age of onset of the Disorder enables to calculate the projected lifetime risk of the Disorder as well as comprehend the causes and mechanisms of illness and to intervene at an appropriate juncture for primary and secondary prevention.

1. Introduction

1.1 Mental Disorder

A Mental Disorder (MD) also called a mental illness or psychiatric disorder is one aspect of mental health. It occurs when problems with mental health begin. It causes mild to severe disturbances in thought and behaviour of an individual, resulting in an inability to cope with life's ordinary demands and routines. According to the *Diagnostic & Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV-TR)*, a mental disorder is defined by an individual's level of distress, disability and increased risk of harming themselves or others [1].

Mental Health problems have become a major public health concern all over the world [2]. In India, a study by the National Commission on Macro economics and Health (NCMH) 12 years back showed that at least 6.5% of the Indian population has some form of Mental Disorders [3]. In a Survey conducted by The National Mental Health Survey of India, 2015-16, based on uniform and standardized data collection procedures from a nationally representative population, it is estimated that, excluding tobacco use disorders, mental morbidity of individuals above the age of 18 years currently was 10.6% [4]. The 2011 census of India reveals that the mentally ill population constitutes 2.6% of the total population of the country.

1.2 Age of onset of the Disorder:

The Age of onset is a medical term referring to the age at which an individual acquires, develops, or first experiences a condition or symptoms of a disease or disorder [5]. It is a key clinical epidemiological variable, which has only recently become the focus of major study and interest. The study of the age distribution of the onset of the Mental disorders is very important because it enables to calculate the projected lifetime risk of the Disorder as well as comprehend the causes and mechanisms of illness and to intervene at an appropriate juncture for primary and secondary prevention [6].

The study of age of onset of Mental Disorders is technically and conceptually difficult as most of the Mental Disorders often begin with non-specific symptoms. Even sometimes after identifying a disorder based on DSM criteria, it is not clear whether the earlier symptoms were parts of the disorder or not and this may leads to enormous scope for error or bias in measuring the actual age of onset [7].

2. Objectives of the Study:

The main objective of this paper is

- To derive suitable Probability Distributions for Age of onset of Mental Disorders.

3. Methodologies:

3.1 Data collection and Study Area:

For the proposed study, a survey was conducted at Lokopriya Gopinath Bordoloi Regional Institute of Mental Health, (LGBRIMH) [8] Tezpur, Assam, India during the year 2014 - 2015. This is the only Central Institute of Mental Health Care in the entire North-Eastern Region of India. Patients having Mental Disorder are being treated mostly from these areas only. Data are purely of primary nature. Whether a selected patient had Mental Disorder or not was identified using the criteria from the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) [1].

3.2 Study Population

The study population comprised, male and female patients suffering from Mental Disorders mostly from the North Eastern region. The study was a descriptive cross sectional type. The individual patients were interviewed whose ages were from 21 years to 60 years.

3.3 Sample Size Calculation:

The minimum sample size was calculated using the online calculator of National Statistical Service, Australia. [9]. For calculating sample size we have considered confidence level 95%, confidence interval 0.025 and standard error 0.01276. When proportion is not available then it is assumed to be as 0.5. Then the sample size was obtained as 1535.

Alternatively, sample size can be calculated using the formula,

$$n \geq \frac{p(1-p)}{[SE(p)]^2}$$

Where, p stands for proportion. Considering proportion p = 0.5 and $SE\ (p)$ =0.01276, the sample size will be greater than 1535. Thus, for our survey we have considered the sample size to be 1600.

3.4 Sampling Technique:

Patients were registered in the Registration counter to collect their queue number for getting appointment with doctors. They were required to meet doctors in their chambers according to appointment number. This system looks like first in first out (FIFO), i.e., patients met doctors in their chamber according to appointment number on the basis of first in first out. Interviewer collected required information from a patient as per the schedule just after the consultation of the patient with the concerned doctor was over. Time required to fill a schedule was approximately 30 minutes. But the team of doctors checked a number of patients during this period. The interviewer interviewed the next patient who was just come out from the any one of the doctors' chamber. The procedure may be called a systematic sampling without fixed number of gaps.

A total of 1600 number of patients was interviewed out of which 34 patients were found out of the range of the age group 21 - 60. Finally we have the size of the sample was 1566 of which 806 are male and 760 are female.

3.5 Study Limitation

The study has been considered for the age group 21 - 60 years. There may be cases beyond this age group; their information has not been registered.

3.6 Ethical Consideration:

Ethical approval was obtained from the Ethics Committee of LGBRIMH, Tezpur, Assam, India. Informed consents were obtained from the patients or 'his/her companion. Authors hereby declare that the study was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

4.1 Gender and Age of onset of the Disorder:

Majority of the Patients had a very early Age of onset of the disorder. More than one fourth (26.2%) patients (male: 27.1%; female: 25.8%) suffered from the Disorder before reaching the age of 20. Male patients were 1.48 times more likely to suffer from the Disorder at age below 30 than that of females (Odds ratio= 1.48). The average onset age for male and female patients were 26.5 years (sd: 9.6 years) and 28.1 years (sd: 10.5 years) respectively. The median onset age for male patients were 25.6 years and that for females was 27.3 years.

Gender and the age of onset of the Disorder was found significantly different (Mann-Whitney U = 282611.5; p = .004). Also, table- 1 and the Odds ratio indicate that male patients had an early onset of the Disorder than females.

Table 1. Gender and Age of onset of the Disorder

Gender	Age of onset of the Disorder (in years)					Total
	Below 20	20 – 30	30 – 40	40 – 50	50 & above	
Male	214 (26.6%)	334 (41.4%)	171 (21.2%)	64 (7.9%)	23 (2.9%)	806 (100.0)
Female	196 (25.8%)	252 (33.2%)	201 (26.4%)	85 (11.2%)	26 (3.4%)	760 (100.0)
Total	410 (26.2%)	586 (37.4%)	372 (23.8%)	149 (9.5%)	49 (3.1%)	1566 (100.0)

4.2 Distribution of Age of Onset of the Disorder

In section 4.1 it is found that gender and Age of onset of the Disorder are significant, i.e., male and female patients had different pattern of distribution of Age of onset. Therefore two

different probability distributions of Age of onset of the Disorder, one for male and the other for female will be developed.

4.3 Test for normality

For large samples, generally the sample data obeys the normality criteria. As such, the first step in any distribution fitting is to check whether the observed data also have some tendency towards normality. Descriptive statistics, especially the *"skewness and kurtosis"* linked to the second, third and fourth moments are useful for checking the normality of the data. Non-zero skewness reveals a lack of symmetry of the empirical distribution, while the kurtosis value quantifies the weight of tails in comparison to the normal distribution for which the kurtosis equals 3.

The skewness and kurtosis and their corresponding unbiased estimators [10] from a random sample $X_1, X_2, ..., X_n$ drawn from a population X with observations $x_1, x_2, ..., x_n$ are given by

$$Sk = \frac{\mu_3}{\mu_2^{\frac{3}{2}}}, \qquad \hat{Sk} = \frac{\sqrt{n(n-1)}}{n-2} \times \frac{m_3}{m_2^{\frac{3}{2}}}$$

and $$Kr = \frac{\mu_4}{\mu_2^{3}}, \qquad K\hat{r} = \frac{n-1}{(n-2)(n-3)} \left[(n+1)\frac{m_4}{m_2^{2}} - 3(n-1)\right] + 3$$

where, $\mu_i = \mathrm{E}\left[X - E(X)\right]^i$; $i = 1, 2, 3, 4$

and m_2, m_3, m_4 denote empirical sample moments, defined by,

$$m_k = \frac{1}{n}\sum_{i=1}^{n}(xi - x)^{\mathrm{k}}$$

Here *Sk* and *Kr* are respectively the skewness and kurtosis of the population and $\hat{Sk}$ and $K\hat{r}$ are corresponding unbiased estimators.

Consequently we get,

$\hat{Sk}$ (male) = 0.7799764 , $K\hat{r}$ (male) = 3.269271

$\hat{Sk}$ (female) = 0.6211662, $K\hat{r}$ (female) = 2.689601.

The positive values of skewness and values of kurtosis somewhat different from 3, definitely point out non normality pattern of the distributions, instead suggests for some right-skewed probability distributions for best fit of the data.

As such the three right-skewed distributions, namely, the Weibul distribution, the Gamma distribution and the Lognormal distribution are proposed to represent the Age of onset of the Disorder for both genders.

4.4 Model Choice:

To pick up the appropriate probability distribution, two methods are applied- firstly, a comparison of all the three distributions will be done with the help of empirical graphs and finally, the best model will be selected using goodness of fit criteria.

The following graphs give a general comparison of these three distributions:

Histogram and Density plot:

Histogram and density plot may be considered as the basic classical goodness of fit plots. Density plot shows that for both genders, Gamma distribution and Log normal distribution have a better fit of the data, compared to Weibul distribution, which over represents both the

tails of the Age of onset of the Disorder. Figure-1 shows that the Gamma distribution and the Lognormal distribution properly fit the distribution of males, while for female, the central part of the distribution is over represented by both the distributions.

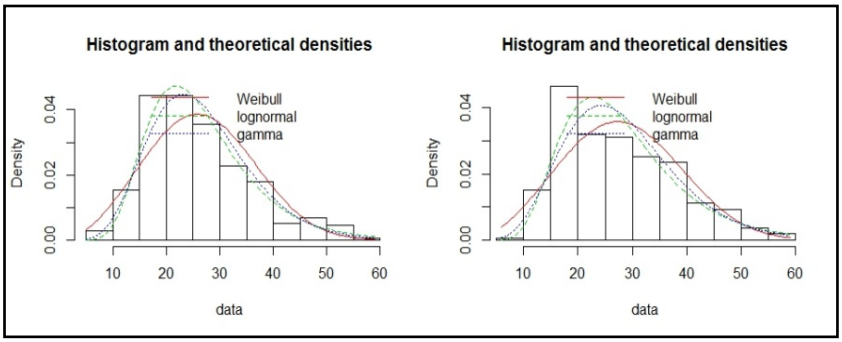

Figure- 1: Histogram and density plot of Age of onset of the Disorder for the three distributions

CDF Plot:

The Cumulative Distribution Function (CDF) plot (figure-2) also clears that for both the distribution of males and females, the Weibul Distribution could not fit better the lower and central part of the age distribution, while the other two distributions properly fit the entire data.

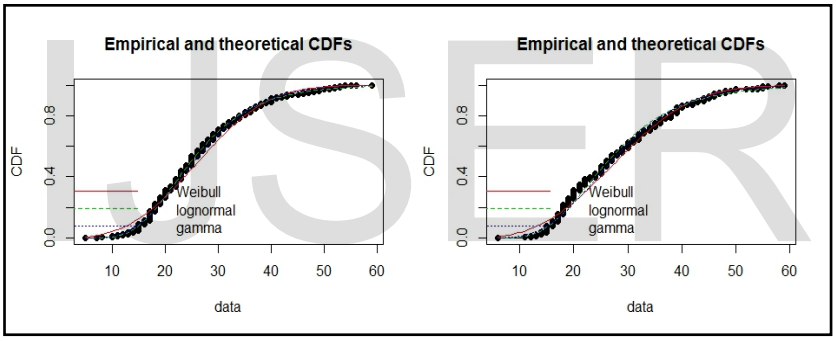

Figure 2: CDF plot of Age of onset of the Disorder for the three distributions

Q-Q plot:

The Q-Q plot emphasizes the lack of fit at the distribution tails. Figure-3 shows that the Weibul Distribution lacks fitting the left tails of the distributions, but fits better the right tails compared to the other two distributions. While Gamma and Lognormal distributions best fit the left tails, but lack fitting the right tails. But, since the right tail of both the distributions of male and female have only a few observations compared to left tails, therefore, overall Q-Q plot also shows that Gamma and Lognormal distributions best fit the tails.

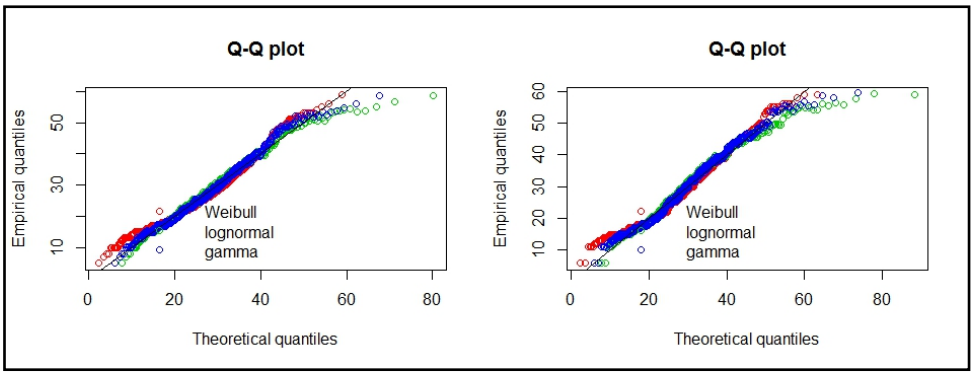

Figure 3: Q-Q plot of Age of onset of the Disorder for the three distributions

It is clear that for both the genders, Gamma distribution and Log normal distribution best fit the distribution of Age of onset of the Disorder, while Weibul Distribution lacks fit.

4.5 Estimation of the parameters:

The parameters of the various probability distributions are estimated by the method of maximum likelihood. Under this method, the likelihood function for estimating the parameters of an empirical probability distribution is,

$$L(\theta) = \prod_{i=1}^{n} f(x_i / \theta) \quad ; \ \theta \ \varepsilon \ \mathrm{R}^{\mathrm{d}}$$

with x_i (i= 1,2,…,n) the n observations of variable X and $f(x_i/\theta)$, the density function of the parametric distribution. To fit the various continuous distributions, the *fitdistrplus* package of statistical software R, version 3.3.2 is used [11]. The estimated values of the parameters are presented in table 2.

Table2: Estimation of the parameters

Gender	Fitted distribution	Estimated values of the parameters
Male	Weibul distribution	Shape: 2.924679 (SE: 0.07613629) Scale: 29.759585 (SE: 0.37992470)
	Gamma distribution	Shape: 7.9363596 (SE: 0.38726918) Rate: 0.2991392 (SE: 0.01506833)
	Lognormal distribution	Mean log: 3.2140353 (SE: 0.012760059) SD log: 0.3622598 (SE: 0.009022415)
Female	Weibul distribution	Shape: 2.87262 (SE: 0.07870211) Scale: 31.62906 (SE: 0.42309315)
	Gamma distribution	Shape: 7.9363596 (SE: 0.38726918) Rate: 0.2991392 (SE: 0.01506833)
	Lognormal distribution	Mean log: 3.2671494 (SE: 0.013702819) SD log: 0.3777607 (SE: 0.009689051)

4.6 Goodness of fit:

To decide if the sample comes from a population with a specific distribution, the Kolmogorov-Smirnov test is used [12]. It is based on a comparison between the empirical cumulative distribution function and the theoretical one defined as,

$$F(x) = \int_{0}^{x} f(y/\theta)dy$$

where $f(y/\theta)$ is the pdf of the distribution to be fitted. If $x_{(1)}, x_{(2)},\ldots, x_{(n)}$ are the order statistics of the n observed values of $x_{(1)}, x_{(2)},\ldots, x_{(n)}$ of the sample, the empirical cumulative distribution function is defined as:

$$F_n(x_i) = \frac{N(i)}{n}$$

Where $F_n(x_i)$ is the cumulative distribution function of i^{th} order statistic $x_{(i)}$; $N(i)$ is the number of observations less than $x_{(i)}$. This is a step function that increases by $\frac{1}{n}$ at the value of each ordered data point. The test statistic used is:

$$D_n = \sup |F(x_i - F_n(x_i)|$$

which is the upper extreme among absolute value differences between Empirical CDF and theoretical CDF. The hypothesis regarding the distributional form is rejected if the test statistic, D_n is greater than the critical value obtained from a table [12]. If the sample size n is greater than 50, the critical value of Kolmogorov- Smirnov test for level of significance 0.05 is calculated using the formula, critical value= $\frac{1.36}{\sqrt{n}}$ [13].

Goodness of fit statistics:

Table 3 exhibits the critical values of the goodness of fit for the three distributions using Kalmogorov- Smirnov test. For both male and female patients the calculated values of Kolmogorov test statistics less than the tabulated values at significant level 0.05 accept the hypothesis that the Lognormal Distribution best fits the Age of onset of the disorder for both male and female patients.

Table 3. Goodness-of-fit statistics

Gender	Critical values of the fitted distributions			Tabulated value
	Weibul	Gamma	Lognormal	(= 1.36/ \sqrt(n))
Male	0.08666359	0.05329846	0.03927248	0.0463
Female	0.08981323	0.08421644	0.04561436	0.04933

If T be the random variable denoting the Age of onset of the Disorder for male patients, then pdf $f_m(t)$ of T is

$$f_m(t) = \frac{1}{t\hat{\sigma}_m\sqrt{2\pi}} e^{-\frac{(\log t - \hat{\mu}_m)^2}{2\hat{\sigma}_m}}; t > 0 \qquad \text{(I)}$$

$$= 0 \qquad \textit{; otherwise}$$

Where, $\hat{\mu}_m = 3.2540353$ and $\hat{\sigma}_m = 0.3622598$

Similarly, the pdf of the Age of onset of the Disoredr $f_f(t)$ for female patients is given by,

$$f_f(t) = \frac{1}{t\hat{\sigma}_f\sqrt{2\pi}} e^{-\frac{(\log t - \hat{\mu}_f)^2}{2\hat{\sigma}_f}}; t > 0 \qquad \text{(II)}$$

$$= 0 \qquad \textit{; otherwise}$$

Where, $\hat{\mu}_f = 3.3071494$ and $\hat{\sigma}_f = 0.3777607$

The estimated average onset age of male patients, given by the fitted probability distribution (I) is 25.9 years and that of females given by the probability distribution (II) is 27.3 years.

Estimates are very close to the average onset ages obtained from the sample (male: 26.5 years, female: 28.1 years).

5 Conclusions:

Majority of the Patients had a very early Age of onset of the disorder. More than one fourth (26.2%) patients (male: 27.1%; female: 25.8%) suffered from the Disorder before reaching the age of 20. Male patients had an early onset of the Disorder than females. The distribution of Age of onset of the Disorder was positively skewed and Lognormal Distribution gives a better fit of the onset age. Further, the probability distribution of Age of onset of the Disorder enables to calculate the projected lifetime risk of the Disorder as well as comprehend the causes and mechanisms of illness and to intervene at an appropriate juncture for primary and secondary prevention.

References

[1] Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition *DSM-IV- TR:* American Psychiatric Association, 2000.

[2] Dalgard, O.S., Mykletun, A., Rognerud, M., Johansen, R. and Zahl, P.H. (2007). Education, sense of mastery and mental health: results from a nationwide health monitoring study in Norway. *BMC Psychiatry* 7:20

[3] NCMH Background Papers: Burden of Disease in India (New Delhi, India), September 2005 Ministry of Health & Family Welfare, Nirman Bhawan, Maulana Azad Road, New Delhi 110011, India

[4] National Mental Health Survey of India, 2015-16 Supported by Ministry of Health and Family Welfare Government of India, Implemented by *National Institute of Mental Health and Neuro Sciences, Bengaluru.*

[5] https://en.wikipedia.org/wiki/Age_of_onset

[6] Kessler, R.C., Amminger, G.P., Aguilar-Gaxiola, S., Alonso, J., Lee, S.and Ustun, T.B. (2007) Age of onset of mental disorders: a review of recent literature. Curr Opin Psychiatry; 20: 359-64.

[7] Jones, P.B. (2013). Adult mental health disorders and their age at onset; *The British Journal of Psychiatry*; 202, s5-s10. doi: 10.1192/bjp.bp.112.119164

[8] Lokopriya Gopinath Bordoloi Regional Institute of Mental Health (LGBRIMH). Kalibari Main Road, Mahabhairab, Tezpur, Assam 784001. *www.lgbrimh.gov.in.*

[9] http://www.nss.gov.au/nss/home.nsf/pages/Sample+ size+calculator.

[10] Casella, G. & Berger, R.L. (2002). Statistical Inference. 2nd edition. Duxbury Thomson Learning.

[11] Laure, M. and Muller, D. (2015). ftdistrplus: An R Package for Fitting Distributions; *Journal of Statistical Software*, published by the American Statistical Association; Volume 64, Issue 4

[12] Ricci, V. (2005). FITTING DISTRIBUTIONS WITH R

[13] http://civil.colorado.edu/%7Ebalajir/CVEN5454/lectures/Ang-n-Tang- Chap3-RV.pdf