

Modeling Time-to-Recurrent Relapse in Patients with Bipolar Disorder

TEMESGEN YAECOB¹ AND MEKONNEN TADESSE*²

¹*Department of Statistics, Wachemo University, Wachemo, Ethiopia*

²*Department of Statistics, Addis Ababa University, Addis Ababa, Ethiopia*

Abstract

One of the main features of bipolar disorder is repetition of relapse overtime. Many studies have focused on time-to-first relapse using the most popular Cox proportional hazard model which discards subsequent information on recurrent relapses. The aim of this study was to identify some risk factors of time-to-recurrent relapses in bipolar disorder inpatients by using appropriate recurrent event model. Data on 206 inpatients, available at Amanuel mental specialized hospital, were collected by reviewing the medical records from September 11, 2013 to March 12, 2019. Different extended cox proportional hazard models including AG, PWP-TT, PWP-GT and semi-parametric shared gamma frailty models were used. R package FrailtyEM package used to fit semi-parametric shared gamma frailty models through EM algorithm. The mean age of the patients was 33.33 years. Within the study time, a total of 418 inpatient admissions (relapses) were registered for 206 inpatients. Among these admissions, about 49.3% of the patients had first relapse and 50.7% of the patients had more than one relapses. The likelihood test results indicated that the appropriate model is the gap-time based semi-parametric shared gamma frailty model and the important risk factors that have effect on time since the end of the most recent relapse to the start of the next relapses are marital status, substance abuse, employment status and residence. Recurrent relapse may be reduced by giving more intensive forms of treatment and creating awareness on each risk factor.

Keywords *bipolar disorder; frailty; recurrent event times; recurrent relapse; risk factors.*

1 Introduction

1.1 Background of the Study

Bipolar disorder is severe mental disorder that causes several problems in a person's psychosocial functioning (Amini et al., 2017). The common types of bipolar disorder are bipolar I and bipolar II characterized by alternating episodes of mania, depression, hypomania and mixtures of them (Bobo, 2017). Patients with bipolar disorder are marked by extreme change in mood, thought, energy, activity level, behavior and ability to carry day-to-day tasks (Biglu and Biglu, 2013). The bipolar I disorder patients have the recurrence episodes of mania and depression while the bipolar II disorder patients never develop severe mania but instead experience milder episodes of hypomania that alternate with depression. Sometimes, patients of bipolar disorder experience more than 4 episodes within 12-months period, called rapid-cycling bipolar disorder (Parikh, 2013).

During depressive episode, the patients are characterized by loss of interest and pleasure in activities enjoyed before; overwhelming sadness; withdrawing from friends and avoiding social

*Corresponding author. Email: mekonnetadesse@yahoo.com or mekonnen.tadesse@aau.edu.et.

activities; ceasing self-care tasks like shopping and showering; changes to appetite and sleep patterns; lack of concentration, extreme tiredness, and feelings of guilt or worthlessness. During manic episodes, they are described by the most severe state of extreme elation and over-activity. The Common symptoms of mania include: elevated mood, increased energy and over-activity, reduced need for sleep, irritability, rapid thinking and speech, recklessness, grandiose plans and beliefs and lack of insight (Jones et al., 2010).

One important feature of Bipolar I disorder is the repetition of relapses over time and more than 90% of bipolar disorder patients have at least one relapse in their life time (Kessler et al., 1994; Bebbington and Ramana, 1995; Merikangas et al., 2007; Grant et al., 2005). Many factors, including demographic and illness characteristic might cause the recurrence of bipolar disorder. Therefore, identifying these factors can help in preventing or reducing the incidence of recurrence of this disease (Vejdani-aram et al., 2017).

Time-to-relapse in bipolar disorder is characterized by a cycling from depression to mania, and back again over time. This type of structure of data is known as recurrent relapse because the relapses happen subsequently on the same patients over time.

Patients with bipolar disorder and history of childhood maltreatment had greater severity of mania, depression and psychosis, higher risk of comorbidity, earlier age of onset, higher risk of rapid cycling, greater number of manic or depressive episodes and higher risk of suicide attempt compared with those bipolar patients without childhood maltreatment (Rowland and Marwaha, 2018).

The recurrent nature of bipolar disorder needs special methods of modeling, analyzing and identifying the influential factors that cause the relapses. Recurrent event is an event in which the event of interest can occur more than once on a subject under the assumption that the same type or different types of events can occur multiple times on the same individual. Examples of such events include recurrence in breast cancer (Rondeau, 2010), asthma attack (Duchateau et al., 2003), lower respiratory tract infection (Amorim and Cai, 2015), sport injury (Ullah et al., 2014), Chronic Renal Insufficiency (Yang et al., 2017) and relapse in bipolar disorder (Kessing et al., 1999).

Bipolar disorder is also known as mood disorder. It is a common chronic recurrence disease with various causes responsible for its recurrence.

Bipolar disorder is a common, chronic and recurrent disease with various causes responsible for its recurrence. Therefore, information on the clinical course of bipolar disorder, including time-to-recurrent relapse provide valuable tool for planning and evaluating the health-outcome results of treatment. However, there is no known exact cause of bipolar disorder. Researchers have yet to find the exact genes that contribute to the disorder or understand exactly how the brain physically changes when the disorder is present (Smith, 2018). Time-to-recurrent relapses of these patients are rarely studied particularly considering heterogeneity between individuals and dependency between event times recurrent relapse (Taheri et al., 2016).

Many studies have focused on the time-to-first relapse in bipolar disorder and used Cox proportional hazard model, but few of them have been carried out to include most appropriate feature of bipolar disorder but have discarded subsequent information on recurrent relapse.

Moreover, ordinary methods in survival analysis implicitly assume that populations are homogenous. But heterogeneity, often referred to as variability is generally recognized as one of the most important sources of variability in medical and biological applications (Wienke, 2010).

A number of studies have failed to account for such heterogeneity assumption, the repeated feature of relapses and within subjects correlation caused by recurrent events. Estimating the individual hazard rate without taking into account the dependence and heterogeneity leads to

underestimate of the hazard function to an increasingly greater extent; as time goes by (Wienke, 2010).

Studies on the factors responsible for the time-to-recurrent relapse in patients with bipolar disorder in Ethiopia are scanty. Hence, this study is an attempt to investigate and identify the most important factors associated with time-to-recurrent relapse in patients with bipolar disorder using multivariate survival recurrent event models.

Furthermore, many studies have also failed to account for recurrent event analyzing techniques of survival analysis in order to study time to relapses, with respect to repeated characteristic of bipolar disorder.

Therefore, this study has mainly attempted to reduce the identified research gap and answer the following basic questions.

1. What are the appropriate models that best fit the bipolar disorder patient data?
2. Which factors have an effect on the time to recurrent relapse in bipolar disorder?
3. Is there any dependency between the recurrent event times within individuals?

2 Data and Methodology

2.1 Data and Study Population

The data on inpatients with bipolar disorder relevant for this study were collected from medical charts available at Amanuel mental specialized hospital, which is the only mental specialized hospital in Ethiopia. Trained psychiatric nurses collected the data using structured questionnaire designed for this purpose.

The patients were diagnosed and treated as bipolar disorder patients based on the criteria of the fifth version of Diagnostic and Statistical manual for Mental disorder (DSM-IV) and the 10th version of the international classification of diseases (ICD 10). All bipolar disorder patients who were clinically admitted from September 11, 2013 to March 12, 2019 for bipolar disorder cases and those who have experienced at least one event of relapse after onset of the disease were included in the study. For each recurrent relapse the start date of relapse and the discharge date of relapse were recorded. Since we used gap-time based approach, it was measured by the duration of the time since the end of the previous event to the start of next events. Data on the dependent variable was recorded as the date of admission and date of discharge for any severe mental disorder for each relapse that occurred within patients.

2.2 Variables of the Study

The dependent variable for our study was time-to-recurrent relapse of bipolar disorder. The periods between two successive relapses of bipolar disorder are called episodes. Episodes, to be considered as two separate episodes, must be separated by at least 8 consecutive weeks without mood of disturbance (Kessing et al., 1999).

Potential factors that have effect on time-to-recurrent relapse in bipolar disorder included in the study are family history of disorder, seasonal based relapse (spring, summer, autumn and winter), having other chronic disease (cardiovascular, diabetes and obesity), relation breakdown, death of significant others, type of treatment, stopping medication or using it erratically, substance abuse, high stressful life events, childhood stress factors, number of history prior relapse, number of relapse within the study time and total number of relapses. In addition,

socio-demographic characteristics of the patients including sex, age at start of first admission, marital status, employment status, religion, place of residence and region were included.

2.3 Sampling Techniques and Sample Size Determination

Since we are interested to test the null hypothesis of homogeneity between the recurrent relapse times, the sample size (total number of patients) required for one sided test necessary to achieve the power of the test, $= 1 - \beta$, is given by

$$n = \frac{(z_{1-\beta} + z_{1-\alpha})^2}{\psi \beta^2},$$

$$\psi = \sum_{m=0}^{\max k_i} (\theta + m) E_{m, T_{m+1}} [C_m(t) - C_m^2(t)],$$

$$C_m(T) = \frac{e^{\beta_0} \left(\sum_{j=1}^m T_j^\gamma + (k - m) f_t^\gamma \right)}{\theta + e^{\beta_0} \left(\sum_{j=1}^m T_j^\gamma + (k - m) f_t^\gamma \right)},$$

$$E_{m, T_{m+1}} [C_m(t) - C_m^2(t)] = \int_0^{f_t} \dots \int_0^{f_t} \int_{f_t}^\infty \dots \int_{f_t}^\infty [C_m(t) - C_m^2(t)]$$

$$f(t_1, \dots, f_k) dt_k \dots dt_{m+1} dt_m \dots dt_1,$$

where f_t is median follow-up time, θ is the variance of frailty, $\max k_i$ is the maximum number of events per patient, $E_{m, T_{m+1}} [C_m(t) - C_m^2(t)]$ is the expectation with t_1, \dots, f_{m+1} , m is the non-censored event from the maximum number of events $\max k_i$. $0 \leq m < \max k_i$. We can evaluate $E_{m, T_{m+1}} [C_m(t) - C_m^2(t)]$ by using the r statistical package called cubature. Using $\alpha = 0.05$ and $\beta = 0.2$, the tabulated values obtained from z table are $z_{0.2} = 0.84$ and $z_{1-0.05} = 1.645$. ψ , is integrated from the r software. However, for moderately correlated events its proposed values range from 0.5 to 0.75. This method of sample size determination for recurrent events in shared frailty models is widely discussed in [Chen et al. \(2014\)](#). Using $\psi = 0.75$, we have determined the sample size to be

$$n = \frac{(0.84 + 1.645)^2}{(0.2)^2 \times 0.75} = \frac{6.18}{0.03} = 206.$$

A list of 822 inpatients that experienced at least one bipolar disorder relapse in all bipolar disorder wards was used as a sampling frame. From this sampling frame, we selected 206 patients using systematic random sampling method.

2.4 Methodology

Modeling time-to-recurrent relapse of inpatients with bipolar disorder have more than one event time for each patient and hence, requires multivariate survival data analysis methods.

2.4.1 General Data Layout for Recurrent Event Times

To represent recurrent event data, we used the following notations; N = total number of patients, T_{ij} denotes the event time with t_{ij0} representing start time and t_{ij1} discharge time of the j^{th} event

for the i^{th} subject. Patients that have not experienced relapse are represented by censored times $C_{ij} = (\delta_{i0}, \delta_{i1}, \dots, \delta_{ir_i})$. If a patient has r_i recurrent events, we observe information on recurrent events for patient i by representing r_i triplets

$$((t_{i10}, t_{i11}, \delta_{i0}), (t_{i11}, t_{i112}, \delta_{i1}), \dots, (t_{ir_i0}, t_{ir_i1}, \delta_{ir_i})),$$

where, δ_{ij} represent the status of event ($1 =$ if the event has occurred, $0 =$ if the event did not occur), each event time has X_{ijk} covariates, where $i = 1, 2, \dots, N, j = 1, 2, \dots, n_i, k = 1, 2, \dots, p$.

In order to make an inference about recurrent relapse in bipolar disorder and for assessing the effects of covariates on the survival times between different recurrent relapses, we use different recurrent event models.

2.4.2 Key Components to Characterize the Recurrent Events Models

Deciding on risk interval, risk set, assumption of baseline hazard and within subject correlation independent increment, conditional, marginal and shared frailty models are proposed for recurrent events (Therneau and Grambsch, 2000).

2.4.3 Anderson and Gill Independent Increment Model

This model, also known as A-G independent increment model, is first proposed by Anderson and Gill (1982). It is particularly useful when interest centers on modelling the overall recurrence rate of the event of interest (Collett, 2015).

The AG model for the hazard of an event occurring at time t for the i^{th} individual is given by

$$h_{ij}(t) = Y_{ij}(t) h_0(t) e^{\beta' X_{ij}(t)},$$

where $Y_{ij}(t)$ denotes whether or not the i^{th} individual is at risk of experiencing an event at time t , β represents the vector of parameters to be estimated, $X_{ij}(t)$ represent time-dependent explanatory variables, and $h_0(t)$ represent a baseline hazard function.

2.4.4 Prentice, William and Peterson (PWP) Conditional Model

This model is proposed by Prentice, William and Peterson (Prentice et al., 1981) and is the most flexible model in which separate strata are defined for each ordered event. PWP models assume that recurrent events within subject are related and baseline hazard vary from event to event and occurrence of subsequent event is affected by previous event (Yadav et al., 2018).

In the PWP model, the hazard of the j^{th} occurrence, $j = 1, 2, \dots$, of an event at time t in the i^{th} of n individuals is

$$h_{ij}(t) = Y_{ij}(t) h_{0j}(t) e^{\beta'_j X_{ij}(t)}.$$

Where $Y_{ij}(t)$ is unity until the $(j - 1)^{th}$ recurrent event and zero otherwise, β'_j is the vector of coefficients of the p explanatory variables for the j^{th} recurrence time, $X_i(t)$ is the vector of the explanatory variables, and $h_{0j}(t)$ is the event specific baseline hazard for the j^{th} recurrence. In this model, the risk set for the j^{th} recurrence is restricted to individuals who have experienced the previous $(j - 1)^{th}$ recurrences.

The PWP model has two forms: PWP-Total time and PWP-Gap time model, the basic interest is knowing the effect of intervention on the outcome variable from the beginning of the study and knowing the effect from previous event respectively (Yadav et al., 2018).

2.4.5 The Frailty Model

The term frailty is first introduced by [Vaupel et al. \(1979\)](#) and it is the convenient way to handle and describe the dependency of the observation within an individual and/or the heterogeneity between individuals in the recurrent events ([Duchateau and Janssen, 2007](#)).

A frailty model in statistical terms is called a random effect model for time-to-event data; this random effect has a multiplicative effect on the baseline hazard function. Its aim is to model and account dependency in cluster (recurrent) survival time through the introduction of a cluster-specific random effect, in which the survival times are conditionally independent given frailty ([Wienke, 2010](#)).

The extended Cox proportional hazard model with frailty term w_i is specified as

$$h_{ij}(t) = h_0(t) e^{X_{ij}\beta' + w_i},$$

where $h_0(t)$ denotes baseline hazard function, $X_{ij}(t)$ denotes the vector of observed covariates, β the vector of regression coefficients and w_i is the frailty term. If $u_i = e^{w_i}$, the frailty model is given by $h_{ij}(t) = h_0(t) u_i e^{X_{ij}\beta'}$.

When modeling time to recurrent relapse, the relapse is clustered within patients and the patients share a common frailty. This common frailty is termed as shared frailty ([Hanagal, 2011](#)).

The conditional multivariate shared frailty survival model for observed covariates $X_{ij1}, X_{ij2}, \dots, X_{inik}$ is given by

$$S(t_{ij1}, t_{ij2}, \dots, t_{ini} | X_{ijk}, u_i) = e^{-u_i \sum_{j=1}^{ni} M_0(t_{ij}) e^{\beta' X_{ij}}},$$

where $M_0(t_{ij}) = \int_{t_{ij0}}^{t_{ij1}} h_0(s) ds$ denotes the cumulative baseline hazard function and $X_{ijk} = X_{ij1}, \dots, X_{inik}$ is the covariate matrix in the i^{th} cluster.

The unconditional multivariate shared frailty survival function is derived by using Laplace transform from conditional survival function, called marginal survival function is given by

$$\begin{aligned} S(t_{ij1}, t_{ij2}, \dots, t_{ini} | X_{ijk}) &= ES(t_{ij1}, t_{ij2}, \dots, t_{ini} | X_{ijk}, u_i) \\ &= E e^{-u_i \sum_{j=1}^{ni} M_0(t_{ij}) e^{\beta' X_{ijk}}} \\ &= \left(L \sum_{j=1}^{ni} M_0(t_{ij}) e^{\beta' X_{ijk}} \right), \end{aligned}$$

where L is the Laplace transform of frailty variable.

The joint survival function for all recurrent event-time data is the product of the survival functions of all the clusters because of the assumption about independence between clusters.

$$S(t_{11}, t_{12}, \dots, t_{in_i} | X_{ij1}, \dots, X_{ijk}) = \prod_i \left(L \sum_{j=1}^{ni} M_0(t_{ij}) e^{\beta' X_{ijk}} \right).$$

The gamma distribution has been widely applied for the choice of frailty distribution in different literature from computational and analytical point of view and availability of software. It is a flexible distribution that takes a variety of shapes. When, $\theta = 1$, it is an exponential distribution and when θ is large, it takes a bell-shape identical to normal distribution ([Wienke, 2010](#)).

Therefore, if we assume $u_i = u_1, u_2, \dots, u_N$ are independent and identically distributed gamma random variables with density function

$$f_{u_i}(u_i, \theta) = \frac{u_i^{\theta-1} e^{-\frac{u_i}{\theta}}}{\theta^{\frac{1}{\theta}} \Gamma} \left(\frac{1}{\theta} \right), \theta > 0.$$

The gamma frailty distribution has $E(u_i) = 1$ and $Var(u_i) = \theta$, where u_i is used to measure degree of heterogeneity among subjects and the dependence (association) within subjects.

If the sample size is large enough, a semi-parametric approach is often preferred because it does not rely on assumptions that are difficult to verify. To adapt this approach to proportional hazard models with frailty, the EM algorithm (Expectation-Maximization algorithm) can be used (Dempster et al., 1977).

An alternative approach to fit semi-parametric gamma frailty models based on penalized partial likelihood maximization leads to the same estimates as the EM algorithm. This technique, however, can be extended to a semi-parametric model with normally distributed random effects (Duchateau and Janssen, 2007).

3 Results

3.1 Descriptive Analysis

Of a total of 822 inpatients with bipolar disorder, a sample of 206 patients who were hospitalized from September 11, 2013 to March 12, 2019 and those who have experienced at least one event of relapse of bipolar disorder were included in the study. To select the sample, Health information management system (HIMS) books available at all bipolar wards of Amanuel mental specialized Hospital were reviewed and 206 inpatients were selected randomly.

The frequency and percentage distribution of bipolar disorders episodes in different relapses are presented in Table 1.

Table 1 shows that there were 206 (49.3%) first admissions, 99 (23.7%) second admissions, 56 (13.4%) third admissions, 31 (7.4%) fourth admissions, 12 (2.9%) fifth admissions, 8 (1.9%) sixth admissions, 5 (1.2%) seventh admissions and 8 (1.9%) eighth admissions. About 50.7 % patients have experienced more than one relapse.

Table 2 shows the distribution of the sample of 206 BD patients included in the study by factors. Among the 206 patients, 124 (60.2 %) patients were female , 150 (72.8%) patients were unmarried, 149 (72.3%) patients were from urban areas, 87 (42.2%) patients were from Addis Ababa, only 2 (1.0%) patients were from Benishangul-Gumuz, 180 (87.4%) were unemployed patients and 172 (83.5%) were uneducated patients as shown in the third and fourth columns of Table 2.

Table 2 also shows that a total of 418 inpatient admissions (relapses) were registered for the 206 inpatients under study within the study time from September 11, 2013 to March 12, 2019. Among these admissions, 161 (38.6%) admissions were for male patients, 333 (79.7%) admissions were for patients from urban areas, 220 (52.6%) admissions were for patients from Addis Ababa, only 2 (0.5%) were for patients from Benishangul-Gumuz, 325 (77.8%) admissions were for unmarried patients, 372 (89%) admissions were for unemployed patients and 364 (87.1%) admissions were registered for uneducated patients as shown in the fifth and sixth columns of Table 2.

The median and mean follow up times were 13.875 months and 20.366 months respectively as shown in Table 3. The average age of the patients at first admission was 33.33 years. The

Table 1: Frequency and percentage of bipolar disorders episodes in different relapses.

Relapse	Frequency	Percent (%)
1 st	206	49.3
2 nd	99	23.7
3 rd	56	13.4
4 th	31	7.4
5 th	12	2.9
6 th	8	1.9
7 th	5	1.2
8 th	1	0.2

Table 2: Frequency and percentage of the sampled patients and bipolar disorders episodes in different relapses by socio-demographic factors.

Variables/ Factors	Categories	Sampled patients		Total admission /episodes	
		Frequency	Percent (%)	Frequency	Percent (%)
Gender	Male	82	39.8	161	38.6
	Female	124	60.2	257	61.4
Marital status	Married	56	27.2	93	22.2
	Unmarried	150	72.8	325	77.8
Religion	Orthodox	112	54.4	227	54.3
	Protestant	40	19.4	75	18.2
	Muslim	54	26.2	115	27.5
Residence	Urban	149	72.3	333	79.7
	Rural	57	27.7	85	20.3
Employment status	Employed	26	12.6	46	11.0
	Unemployed	180	87.4	372	89.0
Education status	Educated	34	16.5	54	12.9
	Uneducated	172	83.5	364	87.1
Type of treatment	Drug	131	63.6	283	67.7
	Therapy	5	2.4	7	1.7
	Drug and therapy	70	34.0	128	30.6
Region	Amhara	28	13.6	41	9.8
	Oromia	56	27.2	98	23.4
	Benishangul-Gumuz	2	1.0	2	0.5
	SNNPR	33	16.0	57	13.6
	Addis Ababa	87	42.2	220	52.6

Table 3: Summary of descriptive statistics.

Summary	Follow-up time	Age	Previous relapse	Current relapse	Total relapse
Minimum	0.250	15	0	1	1
1 st quartile	7.312	27	0	1	1
Median	13.875	32	1	3	4
Mean	20.366	33.33	1.22	3.06	4.268
3 rd quartile	29.188	39	2	4	6
Maximum	66.250	63	7	8	11

study participants had at least one relapse and at most 8 relapses during the study time. The average number of relapses was 3.06. The minimum age of participants at study entry or at baseline was 15 and the maximum age was 63.

3.2 Inferential Analysis

3.2.1 Description of results from proposed models

Results from Anderson and Gill independent increment model (AG), total-time and gap-time based Prentice William and Peterson conditional models (PWP-TT and PWP-GT) and Semi-parametric shared gamma frailty model are displayed in the appendix. Coefficients of covariates, hazard ratios, robust standard errors (for non-frailty recurrent models), adjusted standard errors (for frailty model), z-values and p-values have been computed.

The estimated regression coefficients for residence, religion (Protestant and Muslim), substance abuse, treatment type (therapy and both drug and therapy) and stopping medication or using it erratically were negative for all models. Conversely, the estimated coefficients for marital status, employment status, having history of previous relapse and childhood stressful life events are positive.

3.2.2 Model comparison

The likelihood ratio test (LRT) was used to compare AG, PWP-TT, PWP-GT and semi-parametric shared gamma frailty models and the result in Table 4 show that the semi-parametric shared gamma frailty model with the smallest LRT is the best fitting model to our data.

3.2.3 Results from semi-parametric shared gamma frailty model

As can be seen in the second column of Table 5, the hazard ratio corresponding to each of the variables: age, marital status, region, employment status, previous relapse, childhood stress factors, stressful life events and seasonal case is greater than one while the hazard ratio for each of the remaining variables is below one.

Table 4: Results for model comparison.

Criterion	AG	PWP-TT	PWP-GT	Frailty
LRT	43.7	40.62	30.98	0.994

Table 5: Semi-parametric shared gamma frailty model fit results.

Variables	Category	Coefficients	exp (coefficients)	se (coefficients)	Adjusted se	Z	p
Age		0.005	1.005	0.007	0.007	0.757	0.45
Sex	Female	-0.053	0.948	0.123	0.124	-0.428	0.67
Marital	Unmarried	0.381	1.463	0.155	0.156	2.435	0.01
Region	Oromia	-0.164	0.849	0.219	0.219	-0.746	0.46
	Benishangul- Gumuz	-1.862	0.155	0.780	0.799	-2.331	0.02
	SNNPR	0.276	1.318	0.244	0.248	1.113	0.27
	Addis Ababa	-0.005	0.995	0.217	0.218	-0.023	0.98
Residence	Rural	-0.400	0.670	0.167	0.167	-2.399	0.02
Religion	Protestant	-0.261	0.771	0.161	0.163	-1.600	0.11
	Muslim	-0.128	0.880	0.144	0.145	-0.886	0.38
Employment	Unemployed	0.423	1.527	0.185	0.188	2.251	0.02
Previous relapse	Not have PR	0.066	1.068	0.043	0.043	1.548	0.12
Family history	No FH of disorder	-0.046	0.955	0.210	0.212	-0.217	0.83
Childhood stress factors	No CHSF	0.129	1.137	0.323	0.324	0.398	0.69
Stressful Life Events	No SLE	0.105	1.111	0.150	0.170	0.619	0.54
Substance abuse	Not used substance	-0.267	0.766	0.138	0.138	-1.931	0.05
Stopping medication or using it erratically	Not stop or using it erratically	-0.170	0.844	0.157	0.160	-1.066	0.29
Having other chronic disease	Not have OCD	0.268	0.765	0.255	0.255	-1.051	0.29
Seasonal case	Not SC	0.231	1.260	0.658	0.661	0.349	0.73
Treatment	Therapy	-0.501	0.606	0.434	0.436	-1.147	0.25
	Drug and therapy	-0.097	0.908	0.126	0.126	-0.769	0.44

Estimated distribution: gamma / left truncation: FALSE

Table 6: Frailty summary and confidence intervals based on the likelihood function.

	Estimate	Confidence interval	
		lower 95%	upper 95%
Var [U] or theta	0.079	0.000	0.297
Kendall's tau	0.038	0.000	0.129
Median concordance	0.037	0.000	0.126
E [logU]	-0.040	-0.156	0.000
Var [logU]	0.083	0.000	0.346
Frailty variable (U)	12.605	3.362	Inf

The last column gives p-values for testing the significance of each coefficient. Thus, Marital status, with $p=0.01$; employment status ($p=0.02$), residence ($p\text{-value}=0.02$) and substance abuse ($p\text{-value}=0.05$) have significant effect on the response variable, time-to-recurrent relapse. The hazard ratio for unemployed patients is greater than one ($\exp(\text{coefficient}) = 1.52678$).

3.2.4 Results on measure of dependency and frailty variance

The results in table 6 below describe the estimated measures of dependency between recurrent relapses within patients. The results are part of the estimation of the semi-parametric shared gamma frailty model obtained by using EM algorithm.

Gamma distributed estimated frailty variance; Var [U] is 0.079 with 95% confidence interval [0.00, 0.297]. Unobserved estimated frailty for each individual, 12.605, which is shared by all recurrent observations within the same individual, is significantly greater than one.

Table 6 also presents estimated values of 0.038 and 0.037 for Kendall's tau and Median concordance respectively.

3.2.5 Results for heterogeneity test

The log-likelihood result for frailty (full) and no-frailty (null) models are -2046.511 and -2046.014 respectively. This gives the observed likelihood ratio test value 0.994 presented in Table 4. The corresponding p-value for one-sided likelihood ratio test equals 0.16 while the p-value for the Commenges-Anderson test for heterogeneity is 0.31. Hence, one-sided test of heterogeneity for the given sample size and number of relapses per patient is not supported.

Table 7: Results of the Commenges-Andersen and likelihood ratio tests for heterogeneity.

Fit summary for heterogeneity test	Values
Commenges-Andersen test for heterogeneity	p-value = 0.31
Frailty Log-likelihood	-2046.511
no-frailty Log-likelihood	-2046.014
LRT: $1/2 \times \text{pchisq}(0.992)$	p-value = 0.16

4 Discussion

Before analyzing our data, we placed two different data frames based on total time approach for AG and PWP-total time models and based on gap time for PWP-Gap time and for semi-parametric shared gamma frailty models. This structure for data frame was similarly used by [Ozga et al. \(2018\)](#) for comparison of recurrent event models for composite end points.

Summary of the fit results of the four models are presented in the appendix. The estimated coefficients and other results showed disagreement among the four recurrent event models. As stated by [Amorim and Cai \(2015\)](#), this disagreement is expected.

Based on the results of the LRT, we used semi-parametric shared gamma frailty model to analyze our recurrent relapse data. The same was used by [Duchateau et al. \(2003\)](#) to analyze recurrent asthma event rate. [Rondeau \(2010\)](#) has also used shared frailty model for recurrent breast cancer events.

The fit results of this model revealed that marital status has effect on time since the end of the preceding relapse to the start of the next relapse. Its hazard ratio was 1.46320 (Table 5). This implies that unmarried patients have higher risk of shorter period of time for having the next relapse than married patients on the same frailty level. In other words, unmarried patients have 1.46320 times shorter period of time between two successive relapses than married patients having the same level of frailty controlling for other variables in the model.

Similarly we have identified employment status as one of the main significant factors with hazard ratio 1.52678. This means unemployed patients have 1.52678 times higher risk of a shorter time-to-relapse than employed bipolar disorder patients conditioning on the same frailty.

Another finding of the present study is that rural patients had 0.67013 times less risk of a shorter time of relapse than urban patients at the same level of frailty adjusting for other covariates.

Similarly, patients who did not abuse substance had 0.76585 times lower risk of next relapse than patients with substance use disorder (substance abuse) given the same frailty and controlling for other variables. The result of the present study is consistent with the findings of [Taheri et al. \(2016\)](#). They used Semi-parametric penalized frailty model and found that substance abuse have significant effect on recurrent relapse in bipolar disorder patients. Our finding is also in agreement with the findings by [Shim et al. \(2017\)](#) that marital status and employment status are significant.

The present study also found that age and type of treatments have no significant effect on time-to-recurrent relapse. This result does not correspond with the findings of previous studies by [Vejdani-aram et al. \(2017\)](#).

The present findings about the effect of employment status and residence are also similar to those of previous studies conducted by [Najafi-Vosough et al., \(2016\)](#). But our results on age, sex, marital status, season, family history and type of treatment are not consistent with their results on these variables.

The model used for analyzing our recurrent bipolar disorder data is different from those in the stated studies. For example, [Taheri et al. \(2016\)](#) used semi-parametric penalized frailty model, [Vejdani-aram et al. \(2017\)](#) used fragmented frailty model, [Kessing et al. \(1998\)](#) used cox proportional hazard model and the others used logistic regression to analyze recurrent relapse in bipolar disorder patients.

After identifying the effect of observed covariates, we have presented adjusted parameters for correlation in Table 6 and the frailty variance is 0.079. This value is interpreted as a measure of the correlation between the survival times in the recurrent relapse in patients with bipolar

disorder. This result is expected as suggested by [Wienke \(2010\)](#). The interpretation is different from univariate frailty case, in which the value of the frailty variance indicate the measure of unobserved heterogeneity in the study population. Our model is appropriate to handle the correlation between survival times since the value for frailty variance is not zero.

The additional measure of variance of the frailty variable is dependency. As presented in [table 6](#), the value is greater than zero, suggesting that there is dependency between the survival times of the recurrent relapse in patients with bipolar disorder.

Another value in [Table 6](#) is the value 12.605 for the frailty variable U , which follows gamma frailty distribution. As expected, the value is greater than one meaning that the value of the same frailty variable is shared by all recurrent events in patients.

The frailty variable is responsible for creating dependency between survival times of relapse in patients. We found the dependency between the event times to be 0.038 as shown in [table 6](#). As proposed by [Hougaard \(2012\)](#), the dependency between survival times of recurrent relapse is measured by Kendall's tau. This is also an expected result and we conclude that the event times are dependent since the value is greater than zero.

Finally, the null hypothesis of no heterogeneity between survival times against the alternative that there is heterogeneity is tested by using likelihood ratio test, which follows 50:50 mixture of χ_0^2 and χ_1^2 . The heterogeneity test was not significant at 5% level of significance. Additionally, we used Commenges-Andersen test for heterogeneity, which also follows a mixture of chi-square distribution as likelihood ratio.

We failed to reject the null hypothesis of no heterogeneity based on the Commenges-Andersen test and likelihood ratio test ([Table 7](#)). Both tests indicated that there is no evidence to reject the null hypothesis of no heterogeneity in the recurrent event times between patients. This does not mean that frailty variance is zero and there is no correlation between survival times within patients. This is because in frailtyEM, Commenges-Andersen test is performed before the actual maximization of the likelihood, as it does not depend on the frailty distribution and it does not require the actual estimation of the frailty model ([Balan and Putter, 2019](#)).

The asymptotic behavior Likelihood ratio test converges to mixture of chi-square at point mass zero and with one degree of freedom when the sample size goes to infinity ([Zhi et al., 2005](#)). The reason for the unexpected result may be due to the sample size, number of relapses within patients and magnitudes of frailty variance.

According to the final suggestion made in the above study, the likelihood test for homogeneity is conservative and open problem if the sample size is for recurrent event and the number cluster for other events is finite. [Claeskens et al. \(2008\)](#) discussed that not only likelihood ratio test but also Commenges-Andersen test (score test) is still an interesting open problem to test homogeneity in the null hypothesis for semi-parametric frailty model.

5 Conclusion

The gap-time based semi-parametric shared gamma frailty model was found to be the most appropriate model to analyze time-to-recurrent relapse in inpatients with bipolar disorder. This model was used to identify some important risk factors that are responsible for further relapse. The appropriate multivariate survival recurrent event model used in this study was the gap-time based semi-parametric shared gamma frailty model. Using EM algorithm on this model, the important risk factors identified to have effect on time since the end of last relapse to the start of the next relapses are marital status, substance abuse, employment status and residence. The

value of Kendall's' tau showed that there is dependency between survival times of recurrent relapses.

In many studies, cox proportional model and logistic regression models were used to analyze recurrent event data. These lead to underestimate and losing most important information about latter events that occur after the first event. Based on our findings, we recommend researchers to use more appropriate recurrent event models such as AG, PWP and shared frailty models based on their objectives and research questions.

Furthermore, we recommend not only treating bipolar patients within family members and within the community but also educating them and the community about the features of disorder and creating awareness on each risk factor so as to increase the gap time between the ends of prior relapses to the next relapse.

References

- Amini P, Najafi-Vosough R, Ghaleiha A, Mahjub H (2017). Evaluation of bipolar disorder in several relapses over time using generalized estimating equations. *Iranian Journal of Psychiatry*, 12(3): 182–187.
- Amorim LD, Cai J (2015). Modelling recurrent events: A tutorial for analysis in epidemiology. *International Journal of Epidemiology*, 44(1): 324–333.
- Balan TA, Putter H (2019). FrailtyEM: An R package for estimating semiparametric shared frailty models. *Journal of Statistical Software*, 90(7): 1–29.
- Bebbington P, Ramana R (1995). The epidemiology of bipolar affective disorder. *Social Psychiatry and Psychiatric Epidemiology*, 30(6): 279–292.
- Biglu MH, Biglu S (2013). Scientific attitudes towards bipolar disorders. *Journal of Research in Clinical Medicine*, 2(1): 1–10.
- Bobo WV (2017). The diagnosis and management of bipolar I and II disorders: Clinical practice update. In: *Mayo Clinic Proceedings*, volume 92, 1532–1551. Elsevier.
- Chen LM, Ibrahim JG, Chu H (2014). Sample size determination in shared frailty models for multivariate time-to-event data. *Journal of Biopharmaceutical Statistics*, 24(4): 908–923.
- Claeskens G, Nguti R, Janssen P (2008). One-sided tests in shared frailty models. *Test*, 17(1): 69–82.
- Collett D (2015). *Modelling Survival Data in Medical Research*. CRC press.
- Dempster AP, Laird NM, Rubin DB (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1): 1–22.
- Duchateau L, Janssen P (2007). *The Frailty Model*. Springer Science & Business Media.
- Duchateau L, Janssen P, Kezic I, Fortpied C (2003). Evolution of recurrent asthma event rate over time in frailty models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(3): 355–363.
- Grant BF, Stinson FS, Hasin DS, Dawson DA, Chou SP, Ruan W, et al. (2005). Prevalence, correlates, and comorbidity of bipolar I disorder and axis I and II disorders: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *The Journal of Clinical Psychiatry*, 66(10): 1205–1215.
- Hanagal DD (2011). *Modeling Survival Data Using Frailty Models*. Springer.
- Hougaard P (2012). *Analysis of Multivariate Survival Data*. Springer Science & Business Media.
- Jones S, Lobban F, Cooke A, Hemmingfield J, Kinderman P, Mansell W, et al. (2010). Under-

- standing bipolar disorder: Why some people experience extreme mood states and what can help. British Psychological Society.
- Kessing LV, Andersen PK, Mortensen PB (1998). Predictors of recurrence in affective disorder: A case register study. *Journal of Affective Disorders*, 49(2): 101–108.
- Kessing LV, Olsen EW, Andersen PK (1999). Recurrence in affective disorder: Analyses with frailty models. *American Journal of Epidemiology*, 149(5): 404–411.
- Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, et al. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the National Comorbidity Survey. *Archives of General Psychiatry*, 51(1): 8–19.
- Merikangas KR, Akiskal HS, Angst J, Greenberg PE, Hirschfeld RM, Petukhova M, et al. (2007). Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. *Archives of General Psychiatry*, 64(5): 543–552.
- Ozga AK, Kieser M, Rauch G (2018). A systematic comparison of recurrent event models for application to composite endpoints. *BMC Medical Research Methodology*, 18(1): 1–12.
- Parikh S (2013). Bipolar disorder an information guide. Revised edition. Canada: Center for addiction and mental health. <https://www.camh.ca/-/media/files/guides-and-publications/bipolar-guide-en.pdf>.
- Prentice RL, Williams BJ, Peterson AV (1981). On the regression analysis of multivariate failure time data. *Biometrika*, 68(2): 373–379.
- Rondeau V (2010). Statistical models for recurrent events and death: Application to cancer events. *Mathematical and Computer Modelling*, 52(7-8): 949–955.
- Rowland TA, Marwaha S (2018). Epidemiology and risk factors for bipolar disorder. *Therapeutic Advances in Psychopharmacology*, 8(9): 251–269.
- Shim IH, Woo YS, Wang HR, Bahk WM (2017). Predictors of a shorter time to hospitalization in patients with bipolar disorder: Medication during the acute and maintenance phases and other clinical factors. *Clinical Psychopharmacology and Neuroscience*, 15(3): 248–255.
- Smith K (2018). Bipolar disorder causes. <https://www.psychom.net/bipolar-disorder-causes>.
- Taheri SS, Ardakani MRK, Karimlou M, Rahgozar M (2016). Identifying some risk factors of time to recurrent relapses in bipolar I disorder patients using frailty model of survival analysis. *Journal of Psychiatry*, 19(4): 1–5.
- Therneau TM, Grambsch PM (2000). *Modeling Survival Data: Extending the Cox Model*. Statistics for Biology and Health. Springer.
- Ullah S, Gabbett TJ, Finch CF (2014). Statistical modelling for recurrent events: An application to sports injuries. *British Journal of Sports Medicine*, 48(17): 1287–1293.
- Vaupel JW, Manton KG, Stallard E (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3): 439–454.
- Vejdani-aram S, Fardmal J, Roshanaei G, Najafi-Vosough R, Mahjub H, Ghaleiha A (2017). Identifying the factors influencing the recurrence time of bipolar I disorder using frailty model. *Journal of Mazandaran University of Medical Sciences*, 27(148): 56–67.
- Wienke A (2010). *Frailty Models in Survival Analysis*. CRC press.
- Yadav CP, Sreenivas V, Khan M, RM P (2018). An overview of statistical models for recurrent events analysis: A review. *Epidemiology (Sunnyvale)*, 8(4): 354.
- Yang W, Jepson C, Xie D, Roy JA, Shou H, Hsu JY, et al. (2017). Statistical methods for recurrent event analysis in cohort studies of CKD. *Clinical Journal of the American Society of Nephrology*, 12(12): 2066–2073.

Zhi X, Grambsch P, Eberly L (2005). Likelihood ratio test for the variance component in a semi-parametric shared gamma frailty model. *Technical report*, Research Report 2005-5.

Appendix: Model fit results (all proposed models)

Variables	Models	coef	exp(coef)	se(coef)	robust se	Z	p-value	
Age	AG	-0.002	0.998	0.006	0.006	-0.321	0.748	
	PWP-TT	-0.004	0.996	0.007	0.006	-0.679	0.497	
	PWP-GT	0.003	1.003	0.007	0.006	0.524	0.600	
	Frailty	0.005	1.005	0.007	0.007	0.757	0.450	
Sex (Female)	AG	0.042	1.043	0.113	0.105	0.402	0.688	
	PWP-TT	0.002	1.002	0.117	0.101	0.021	0.983	
	PWP-GT	-0.025	0.975	0.116	0.095	-0.262	0.794	
	Frailty	-0.053	0.948	0.123	0.124	-0.428	0.670	
Marital (unmarried)	AG	0.216	1.241	0.142	0.144	1.502	0.133	
	PWP-TT	0.211	1.235	0.145	0.148	1.429	0.153	
	PWP-GT	0.278	1.320	0.143	0.140	1.979	0.048	
	Frailty	0.381	1.463	0.155	0.156	2.435	0.010	
Region	AG	ORO	-0.052	0.949	0.200	0.160	-0.327	0.744
		BG	-1.467	0.231	0.741	0.187	-7.833	0.000
		SNPR	0.359	1.432	0.224	0.183	1.959	0.050
		AA	0.029	1.029	0.200	0.159	0.180	0.857
	PWP-TT	Or..	-0.053	0.948	0.202	0.159	-0.333	0.739
		BG	-1.536	0.215	0.750	0.219	-7.002	0.000
		SNNPR	0.288	1.334	0.227	0.185	1.558	0.119
		AA	0.014	1.014	0.203	0.155	0.091	0.928
	PWP-GT	Or..	-0.150	0.861	0.205	0.149	-1.000	0.317
		BG	-1.534	0.216	0.750	0.211	-7.275	0.000
		SNNPR	0.232	1.261	0.226	0.197	1.176	0.240
		AA	-0.050	0.951	0.205	0.151	-0.331	0.741
	Frailty	Oro..	-0.164	0.849	0.219	0.219	-0.746	0.460
		BG	-1.862	0.155	0.780	0.799	-2.331	0.020
		SNNPR	0.276	1.318	0.244	0.248	1.113	0.270
		AA	-0.005	0.995	0.217	0.218	-0.023	0.980
Residence (rural)	AG	-0.383	0.682	0.151	0.148	-2.595	0.009	
	PWP-TT	-0.372	0.689	0.156	0.147	-2.524	0.012	
	PWP-GT	-0.373	0.689	0.156	0.146	-2.546	0.011	
	Frailty	-0.400	0.670	0.167	0.167	-2.399	0.020	
Religion	AG	Protestant	-0.181	0.834	0.147	0.149	-1.221	0.222
		Muslim	-0.193	0.824	0.133	0.137	-1.406	0.160
	PWP-TT	Protestant	-0.115	0.891	0.151	0.153	-0.750	0.453
		Muslim	-0.210	0.810	0.136	0.138	-1.526	0.127
	PWP-GT	Protestant	-0.161	0.851	0.150	0.143	-1.122	0.262
		Muslim	-0.061	0.941	0.136	0.127	-0.479	0.632

		Continued						
Variables	Models	coef	exp(coef)	se(coef)	robust se	Z	p-value	
	Frailty	Protestant	-0.261	0.771	0.161	0.163	-1.600	0.110
		Muslim	-0.128	0.880	0.144	0.145	-0.886	0.380
Employment	AG		0.352	1.422	0.169	0.159	2.210	0.190
(Unemployed)	PWP-TT		0.283	1.327	0.175	0.166	1.698	0.089
	PWP-GT		0.269	1.308	0.175	0.155	1.736	0.083
	Frailty		0.423	1.527	0.185	0.188	2.251	0.020
Having	AG		0.045	1.046	0.038	0.034	1.310	0.190
previous	PWP-TT		0.079	1.082	0.040	0.035	2.259	0.024
relapse	PWP-GT		0.061	1.063	0.040	0.033	1.872	0.061
	Frailty		0.066	1.068	0.043	0.043	1.548	0.120
Family	AG		0.096	1.101	0.192	0.187	0.513	0.608
history of	PWP-TT		-0.033	0.968	0.197	0.204	-0.166	0.868
disorder	PWP-GT		0.024	1.024	0.198	0.170	0.139	0.889
(No)	Frailty		-0.046	0.955	0.210	0.212	-0.217	0.830
Childhoods	AG		0.235	1.264	0.297	0.240	0.976	0.329
stress	PWP-TT		0.214	1.239	0.304	0.229	0.938	0.348
factors	PWP-GT		0.131	1.140	0.299	0.210	0.624	0.532
(No)	Frailty		0.129	1.137	0.323	0.324	0.398	0.690
Stressful	AG		-0.139	0.870	0.145	0.127	-1.093	0.274
life events	PWP-TT		-0.229	0.795	0.151	0.134	-1.716	0.086
	PWP-GT		0.072	1.075	0.150	0.130	0.553	0.580
	Frailty		0.105	1.111	0.150	0.170	0.619	0.540
Substance	AG		-0.337	0.714	0.128	0.125	-2.698	0.007
abuse	PWP-TT		-0.342	0.710	0.133	0.129	-2.657	0.008
(No)	PWP-GT		0.222	0.801	0.131	0.121	-1.839	0.066
	Frailty		-0.267	0.766	0.138	0.138	-1.931	0.050
Stopping	AG		-0.128	0.880	0.149	0.128	-1.002	0.316
medication	PWP-TT		-0.182	0.833	0.154	0.134	-1.362	0.173
or using it	PWP-GT		-0.058	0.944	0.154	0.140	-0.411	0.681
erratically	Frailty		-0.170	0.844	0.157	0.160	-1.066	0.290
(No)								
Having	AG		-0.055	0.947	0.239	0.269	-0.204	0.839
chronic	PWP-TT		-0.102	0.903	0.252	0.270	-0.377	0.706
disease	PWP-GT		-0.126	0.882	0.252	0.246	-0.510	0.610
(No)	Frailty		0.268	0.765	0.255	0.255	-1.051	0.290
Seasonal	AG		0.069	1.072	0.616	0.390	0.177	0.859
case	PWP-TT		-0.078	0.925	0.628	0.365	-0.214	0.831
(No)	PWP-GT		0.350	1.420	0.625	0.443	0.790	0.429
	Frailty		0.231	1.260	0.658	0.661	0.349	0.730
Treatments	AG	Therapy	-0.598	0.550	0.406	0.327	-1.830	0.067
		Drug and	-0.231	0.793	0.118	0.118	-1.958	0.050
		therapy						
	PWP-TT	Therapy	-0.578	0.561	0.406	0.318	-1.815	0.070

Variables	Models	Continued					
		coef	exp(coef)	se(coef)	robust se	Z	p-value
	Drug and therapy	-0.277	0.758	0.123	0.124	-2.233	0.026
PWP-GT	Therapy	-0.442	0.643	0.408	0.365	-1.213	0.225
	Drug and therapy	-0.124	0.883	0.121	0.125	-0.996	0.319
Frailty	Therapy	-0.501	0.606	0.434	0.436	-1.147	0.250
	Drug and therapy	-0.097	0.908	0.126	0.126	-0.769	0.440