## ORIGINAL PAPER

# Comparison of Efficiency GEE and QIF Methods for Predicting Factors Affecting on Bipolar I Disorder Under Complete-case in a Longitudinal Studies

Zahra Geraili-Afra[1], Alireza Abadi[2] , Jamshid Yazdani-Charati[3], Somayeh Ahmadi Gooraji[2] , Mehran Zarghami[4], Samaneh Saadat[5]

Department of Biostatistics & Epidemiology, Babol University of Medical Sciences, Babol, Iran

Department of Social Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Department of Biostatistics, Psychiatric Research Center, School of Health, Mazandaran University of Medical Sciences, Sari, Iran

Department of Psychiatry & Behavioral Sciences Research Center, School of Medicine, Mazandaran University of Medical Sciences, Sari, Iran

Research Committee, School of Health, Mazandaran University of Medical Sciences, Sari, Iran

Corresponding author: Somayeh Ahmadi Gooraji, MSc , Shahid Beheshti University of Medical Sciences, Tehran, Iran. ORCID ID: https://orcid.org/0000-0003-3768-2650. Email: ahmadisomayeh90@gmail.com

**ABSTRACT**

**Background:** Mood variation in manic and depression phases during time is common in type I of Bipolar disorder. Analyzing recurrence require to the related statistical methods. In this paper, we compare the two methods of estimating the GEE and the QIF in recurrence data. **Methods:** In this study, data of 255 patients with Bipolar I disorder hospitalized during years of 2007-2011. Recurrence in Bipolar I disorder was as outcome. Patients' characteristics were gender, age of onset, recurrence history in first degree family, and economic status. Under simulation, percentage of missing were generated to vary and handled by complete-case(cc) strategy. Data were analyzed using GEE and QIF methods. Performance of the methods was assessed using Relative Efficiency. **Results:** QIF method had more efficiency than GEE method in the data with missing /without missing. Odds of recurrence in a first-degree family history was 30% more than those without a family history (p=0.009). Also, odds of recurrence in high/moderate level of economic status was 23% more than low level status (p=0.014). **Conclusion:** QIF method was more appropriated for modeling recurrence during time with the structure of more correlation and low dropout rate in data. Family history and economic status were more affected recurrence in type I of Bipolar disorder.

Keywords: Bipolar I disorder, Recurrence, Generalized Estimation Equations, Quadratic Inference Function, missing, simulation.

## 1. INTRODUCTION

Bipolar disorder caused to frequent mood variation in manic and depression or both during time (1). The recurrence in Bipolar I disorder is prevalent in different types of mood disorders (1, 2). About 90% of these patients experience recurrence (3-5). The Bipolar I disorder affected the ability and quality of life, social and occupational status of patients with frequent recurrence (6). Its complications including aggression and anxiety, sleep disorder, sexual problems and suicide (1). Although, Bipolar I disorder is treated through drug therapy, psycho-education of patients and their family; most patients experience recurrence yet (7).

Based on findings of previous studies related to recurrence in Bipolar disorder, factors such as reducing drug dosage, drug dependency and substance abuse, sleep pattern, life events and social-mental factors (low social protection, Social inconsistency and entertainments) can affect recurrence significantly (8-10).

On the other hand, the severity of emotional symptoms in patients with Bipolar disorder is varying frequently. So, that is required to assess emotional status repeatedly by the longitudinal study (repeated measurements of recurrence in each patient during time) in order to reach appropriate estimations (1, 11). Some studies had evaluated recurrence in Bipolar disorder in Iran. In a study in 2011, a multivariate regression was used to predict a recurrence rate for six months (12); in another studies in 2007 and 2002, patients with Bipolar disorder were followed up one year (13, 14); in a study in 2006, patients with first episode mania of Bipolar disorder were followed up for 17 months (15).

On the other hand, observations obtained from repeated measurements are correlated in a longitudinal study (11). There are methods that include the within subject correlation structure in data such as Marginal model. The

Generalized Estimation Equations (GEE) and Quadratic Inference Function (QIF) as a consistent and unbiased estimation methods in a Marginal model are used for modeling the within subject correlation. Also, they are used for any outcome such as count, categorical and normal (16, 17).

On the other hand, missing data are more common in longitudinal studies especially in clinical trials which may be due to lack of accurately record or lost to follow up of patients during time. Large proportion of missing data can affect results, and probably cause to a great bias in analysis. One of the methods for handling incomplete datasets or missing is "complete-case" (18). These estimation methods were applied in a few studies on Bipolar disorder, in which the efficiency of two methods were compared without considering missing in data (19-23). In the present study, we also aimed to comparing effect of missing on two estimation methods by simulation. Besides, GEE and QIF methods are compared under missing data which are handled by the "complete-case strategy.

## 2. 2. METHODS

### Patients and design

This longitudinal study was done by census at Zare hospital in the province of Mazandaran in the north of Iran; during five years follow up from 2007 to 2011. Total of 255 hospitalized patients with Bipolar I disorder aged from13 to 55 years were included in study. Diagnosis of Bipolar disorder and biography of patients was done by psychiatrist. Patients had experienced at least one recurrence or anxiety or depression or both based on DSM-IV-TR criteria of American psychiatry association in year of 2000. Data collecting was done through patients' medical records and interview. The outcome variable is recurrence in Bipolar I disorder (event or not (0, 1)) and the predictor variables are as follow: age of onset, gender, recurrence history in first degree family, economic status.

### Statistical method

For descriptive analysis, mean ±standard deviation (SD) for quantitative variables, and frequency with percent for qualitative variables was used. All other analyses were done with R software by simulation method. Significance level was defined as less than 0.05.

*This section is organized as follows:*

First, the briefly introduction of GEE and QIF used to analysis Bipolar I disorder data. Second, the complete-case strategies used in this study for handling missing. Third, relative efficiency for compare between two methods are introduced.

### Generalized Estimation Equations

This method, first introduced by Liang and Zeger in 1986, is the most widely used statistical inference in marginal models (24). The generalized estimation equations will fit by the Quasi-likelihood method and a working correlation construct that require less correlation (disturbance) parameters. In this equation, is considered as vector of correlated observations, and the covariance of quasi-likelihood function is replaced by a working covariance matrix, in which, the diagonal matrix of the variance of the marginal variables of the response, the working correlation matrix including the correlation parameters and the dispersion parameter α and Ø. The estimation of regression coefficients β is obtained from the solution of the following equation.

In this equation, it is assumed that the marginal mean is related to covariate by link function.

### Method of Quadratic Inference Function

In 2000, Qu et al. used a generalized moment method for marginal regression analysis to improve the method of generalized estimation equations (25). In other words, this method is quasi-likelihood i.e. in generalized linear models, in which the inverse of the correlation matrix is approximated by the linear combination of unknown constant and known matrices. The matrix is identity matrix, are known basis matrices and, are constant coefficients and unknown.

We use the generalized moment method to construct the optimal inference function.

The matrix is a consistent estimate of the sample variance.

The regression parameter is estimated without the need to estimate the nuisance parameter α:

Under certain circumstances, this is unique. The inference function in is called the Quadratic Inference Function. Because the estimator of the Quadratic Inference Function does not estimate the correlation nuisance parameter, therefore there is no emphasis on the correct estimation of the correlation parameter (24-26).

### Mechanism of complete - case (cc)

In this method, units with missing values are deleted and statistical analysis is done based on units with complete data. The advantage of this method is that it can be used for any kind of statistical analysis and does not require a particular computational method. In this method, the estimator is unbiased when missing data being as completely at random. Otherwise, estimator is biased and the accuracy and power of study reduce by losing a lot of information. This method is effective only if there is a few missing in data (18).

### Relative efficiency (R.E):

Relative efficiency (R.E) rate is estimated as proportion of mean square error (MSE) of GEE estimation than MSE of QIF. If proportion rate is estimated more than one, that implied on the better relative efficiency of QIF than GEE.

## 3. RESULTS

Data of 255 patients with bipolar I disorder was analyzed. Based on Table 1, most patients were males (%55.3). Mean age range was between 13- 55 years, and mean age onset of disease was 24.11±8.61 year. Of total, only %43 of patients had experienced recurrence in the first-degree family. Most patients had high/moderate level of economical status (168 (65.9%)).

| N (%) | Category | Variables |
|---|---|---|
| 141(55.3) | male | Gender |
| 114(44.7) | female | |
| 109 (43) | Yes | Hifistory in first-degree family |
| 146(57) | No | |
| 87(34.1) | low decile | Economical status |
| 168(65.9) | (middle /upper) decile | |

**Table 1. Patients' characteristics (n=255)**

The results of demographic variables by QIF method is presented in Table 2. Results showed that the recurrence history in a first-degree family, and economical status were statistically significant (p<0.05). Chance of recurrence in patients with recurrence history in a first-degree family was

| Parameter | % missing | Estimate | S.E | OR* | P-value |
|---|---|---|---|---|---|
| Age of onset | 0 | -0.007 | 0.0044 | 0.993 | 0.114 |
| | 10 | -0.006 | 0.0048 | 0.994 | 0.193 |
| | 20 | -0.001 | 0.0048 | 0.999 | 0.891 |
| Gender (Female) | 0 | 0.021 | 0.0797 | 1.021 | 0.793 |
| | 10 | 0.014 | 0.0909 | 1.014 | 0.876 |
| | 20 | 0.062 | 0.0923 | 1.064 | 0.504 |
| History in first degree family (Yes) | 0 | 0.266 | 0.1025 | 1.305 | 0.009 |
| | 10 | 0.309 | 0.1126 | 1.362 | 0.602 |
| | 20 | 0.293 | 0.1263 | 1.340 | 0.205 |
| Economic status ( high/moderate level) | 0 | 0.209 | 0.0849 | 1.233 | 0.014 |
| | 10 | 0.237 | 0.0941 | 1.268 | 0.117 |
| | 20 | 0.161 | 0.0931 | 1.175 | 0.838 |
| Time | 0 | 0.021 | 0.0117 | 1.021 | 0.072 |
| | 10 | 0.012 | 0.0125 | 1.012 | 0.337 |
| | 20 | 0.019 | 0.0138 | 1.020 | 0.157 |

**Table 2. Regression results of QIF method in marginal model for Bipolar I disorder data (missing / without missing data) * OR: Odds ratio; **It is significant at level of 0.05.**

30% more than those without family history. Besides, the chance of recurrence in high/moderate level of economical status was 27% more than low status (p<0.001).

The results of demographic variables by GEE estimation are presented in Table 3. Results showed that the recurrence history in a first-degree family and economical status were statistically significant (p<0.05). With age of onset increasing, chance of recurrence was reduced about 0.01. Besides, chance of recurrence in patients with recurrence history in a first-degree family was 28% more than those without family history.

| Parameter | % missing | Estimate | S.E | OR* | P-value |
|---|---|---|---|---|---|
| Age of onset | 0 | -0.007 | 0.0044 | 0.993 | 0.096 |
| | 10 | -0.006 | 0.0049 | 0.994 | 0.194 |
| | 20 | -0.004 | 0.0049 | 0.996 | 0.438 |
| Gender (Female) | 0 | 0.029 | 0.0797 | 1.03 | 0.708 |
| | 10 | 0.029 | 0.0909 | 1.030 | 0.742 |
| | 20 | 0.077 | 0.0928 | 1.081 | 0.403 |
| History in first degree family (Yes) | 0 | 0.252 | 0.1029 | 1.286 | 0.014 |
| | 10 | 0.280 | 0.1134 | 1.324 | 0.013 |
| | 20 | 0.304 | 0.1268 | 1.355 | 0.016 |
| Economic status ( high/moderate level) | 0 | 0.184 | 0.0850 | 1.202 | 0.030 |
| | 10 | 0.210 | 0.0949 | 1.234 | 0.027 |
| | 20 | 0.140 | 0.0935 | 1.151 | 0.133 |
| Time | 0 | 0.016 | 0.0128 | 1.016 | 0.216 |
| | 10 | 0.013 | 0.0139 | 1.013 | 0.342 |
| | 20 | 0.023 | 0.0145 | 1.024 | 0.107 |

**Table 3. Regression results of GEE method in marginal model for Bipolar I disorder data. * OR: Odds ratio; **It is significant at level of 0.05.**

| RE Bipolar data | REa (ρ=0.6) | REa (ρ=0.1) | % of missing |
|---|---|---|---|
| 1.142 | 1.065 | 0.984 | 0 |
| 1.135 | 1.089 | 0.989 | 10 |
| 1.130 | 1.035 | 0.993 | 20 |

**Table 4. Brief results of simulation for relative efficiency (R.E) under different rates of correlation (ρ) and missing. a: Is created by simulation**

In continue, the relative efficiency of these two estimation methods were compared for choosing the best method. To do comparison, a simulation method was used considering different coefficient of correlation and dropout rate in data. According to relative efficiency (R.E) rates obtained by simulation in table 4, QIF had the more efficiency than GEE in total

and also when data correlation was high (ρ=0.6) and rate of dropout percent was low.

## 4. DISCUSSION

Regarding the periodic changes of bipolar disorder between Manic and depression, assessing recurrence and emotional status of these patients during time is more important; and the cross-sectional study of recurrence meets to limitations (27). Therefore, the longitudinal study is more applicable, in which that is required to especial statistical methods for modeling correlation structure of data during time. In the present study, we used two estimation methods (GEE and QIF) for correlated data, in order to reach a best method for predicting recurrence and affected factors. In our study, simulation results showed that QIF had more efficiency to predict factors related to recurrence whether with/without missing data. This recent result is accordance with results of previous longitudinal studies conducted by simulation. In a simulation study with count data (24) without missing, QIF and GEE methods were compared considering two cut points of correlation as 0.3 and 0.7. Results of their study showed that when correlation structure in data is given appropriately, QIF and GEE have equally efficiency in range of 1.2-2.07. This finding was accordance with the results of studies by Song et al; Odueyungbo et al; and Kazemi et al (21-23). In our study, QIF had the more efficiency when data correlation was high (ρ=0.6) and rate of dropout percent was low.

Based on results of our study which obtained by QIF method, factors such as recurrence history in a first-degree family and economical status had significantly affected recurrence. Recurrence in patients with family history was more than those without family history. Findings of previous studies had confirmed that genetic has an important effect on cause of disorder, and hypertension can also effective on recurrence in younger patients (1, 3, 27-29).

Results of some studies had shown that the rate of disorder was more in males due to economical problems, and also patients with the lower economical level were more at risk of Bipolar disorder and depression (12, 14, 17). These findings are accordance with the results of our study. Based on reports of other studies, bipolar disorder is more observed in males due to their job stress, but gender had not significantly affected recurrence (1, 12, 14, 30). These finding was similar to results of present study.

According to some studies which had used QIF method, chance of recurrence was more in a lower age of disease onset, which may be due to the role of genetic in made of disorder in childhood. This finding no confirmed by our study (31).

So far, there has not been done a proper longitudinal study in related with the effective factors on type 1 bipolar disorder in our country. So, follow up of patients during time, an estimation of appropriate sample size and trying to achieve a valid statistical method for analyzing longitudinal and correlated data is the most important strong point of our study. Regarding the findings of our study were small step for identifying the predictive factors of this mental disorder, suggestions are proposed to enhance such studies such as controlling effect of drug use as confounding variable which affected recurrence over time. Besides, that is suggested conducting more study on comparing statistical methods for analyzing

longitudinal data to achieve valid results.

## 5. CONCLUSION

QIF method was more appropriated for modeling recurrence during time with the structure of more correlation and low dropout rate in data. Factors such as family history and economic status were more affected recurrence in Bipolar 1 disorder

## REFERENCES

1. Kaplan H, Sadock B. The handbook of psychiatry, behavioral sciences/psychiatry. Poorafkari N (translator); 1st ed. Tehran: Shahrab 2004: 105-14.
2. Pfeffer CR, Klerman GL, Hurt SW, et al. Suicidal children grow up: demographic and clinical risk factors for adolescent suicide attempts. J Am Acad Child Adolesc Psychiatry 1991; 30: 609–16.
3. Sadock BJ, Sadock VA. Comprehensive text book of psychiatry; 8th ed. Philadelphia, Lippincott Williams and Wilkins 2005: 1559-1717.
4. Hirschfeld RM, Calabrese JR, Weissman MM, et al. Screening for bipolar disorder in the community. J Clin Psychiatry 2003; 64(1): 53.
5. Morris C, Miklowitz D, Wisniewski S, et al. Care satisfaction, hope, and life functioning among adults with bipolar disorder: data from the first 1000 participants in the Systematic Treatment Enhancement Program. Compr Psychiatry 2005; 46(2): 98-104.
6. Douglas E. does functioning differ before and after daylight savings time changes among patients with bipolar disorder? Thesis, Graduate School of Public Health of Pittsburgh University 2007: 1-2.
7. Rashedi E, Sohrabi F, Shams J. A study of the efficacy of cognitive behavior therapy in replace prevention of bipolar disorder. Clin Psychol Personal 2011; 2(5): 49-64.
8. Yazdani-Charati J, Zarghami M, Saadat. Determining the factors affecting recurrence in bipolar I disorder using Andersen-Gill counting process model.J Mazand Univ Med Sci 2013; 23(108): 83-9
9. Stefos G, Bauwens F, Staner L, et al. Psychosocial predictors of patients on prophylactic treatment.Acta Psychiatrica Scandinavica 1996; 93(6): 420-426.
10. Ghorayshizadeh M,Ranjbar F,Zakaria Pezeshki M. Evaluation of Risk factors of recurrence in type 1 bipolar disorder and It's relation to demographic feature .Medical Journal of Tabriz university of medical sciences 2009; 31(2): 77-81.
11. Diggle PJ, Heagerty p, Liang KY, et al. Analysis of longitudinal data. London: Oxford University Press 2002: 1-2.
12. Molavi P, Shahrivar Z, Mahmodi Garaee J, et al. Six- month outcome predictor factors of manic and mixed bipolar disorders of children and adolescents admitted in Tehran Rozbeh hospi-tal. J Ardabil Univ Med Sci 2011; 11(3): 269-281.
13. Mousavi SGH, Moalemi S, Sadeghi S. Recurrence and Recurrence in bipolar Mood Disorder. J Res Med Sci 2004; 3: 120-122.
14. Amini H , Sharifi V, Nejatisafa A, et al. One year follow-up of patients with bipolar disorder admitted to roozbeh hospital. Iran J psychiatry Clin Psychol 2009; 15(2): 168-174.
15. Shabani A, Eftekhar M, Daneshamuz B, et al. Degree of recurrence of type 1 bipolar disorder: A 17 month follow up of patients with first episode mania. Adv Cogn Sci 2006; 8(3): 33-42.
16. Liang K-Y, Zeger S. Longitudinal data analysis using Generalized Linear Models. Biometrika 1986; 73: 13-22.
17. Hedeker D, Gibbons DR.Longitudinal data analysis. Wiley 2006: 146-147.
18. Nakai M, and Ke W. Review of the methods for handeling missing data in longitudinal data analysis. Int J Math Analysis 2011: 5(1): 1-13.
19. Hennen J. Statistical methods for longitudinal research on bipolar Disorders. Bipolar Disord 2003; 5: 156–168.
20. Baur R, Glenn A , Alda M, et al. Anti-depressant dosage taken by patients with bipolar disorder: factors associated with irregularity. Int J Bipolar Disord 2013: 1-26.
21. Odueyungbo A, Browne D, Akhtar-Danesh N, et al Comparison of generalized estimating equations and quadratic inference functions using data from the National Longitudinal Survey of Children and Youth (NLSCY). BMC Med Res Methodol 2008: 8:28.
22. Song P, Jiang Z, Park E, et al. Quadratic inference functions in marginal models for longitudinal data. Stat Med 2009; 28: 2683-3692.
23. Khajeh-Kazemi R, Golestan B, Mohammad K, et al. Comparisonof Generalized Estimating Equations and Quadratic Inference Functions in superior versus inferior. J Res Med Sci 2010; 16(3): 235-244.
24. Qu A, Lindsay B, Li B. Improving generalized estimating equations usingquadratic inference functions. Biometrika 2000; 87: 823-826.
25. Horton NJ, Bebchuk JD, Jones CL, et al. Goodness-of-fit for GEE: An example with mental health service utilization. Stat Med 1999; 18 (2): 213-222.
26. Barnhart HX, Williamson JM. Goodness-of-fit tests for GEE modeling with binary responses. Biometrics 1998; 54 (2): 720-729.
27. Ahmadvand A. Prevalence of Mental Disorders in General Population of Kashan City. Iran J Epidemiology 2010; 6(2): 16-24.
28. Duffy A, Alda M, Hajek T, et al. Early course of bipolar disorder in high-risk offspring: prospective study. Br J Psychiatry 2009; 195(5): 457-458.
29. Chiaroni P, Hantouche EG, Gouvernet J, et al. The cyclothymic temperament in healthy controls and familially at risk individuals for mood disorder: Endopheno type for genetic studies? J Affect Disord 2005; 85: 135-145.
30. Das AK, Olfson M, Gameroff MJ, et al. Screening for bipolar disorder in a primary care practice. JAMA 2005; 293(8): 956-63.
31. Altman S, Haeri S, Cohen L, et al. Predictors of Relapse in Bipolar disorder:A Review. J Psychiatr Pract 2006; 12(5): 269-282