

RESEARCH ARTICLE

Open Access

# Mathematical model for the distribution of major depressive episode durations

Shinichiro Tomitaka<sup>1\*</sup> and Toshiaki A Furukawa<sup>2</sup>

## Abstract

**Background:** The duration of major depressive episodes varies widely, ranging from one month to more than several years. Despite the accumulation of knowledge regarding the course of major depressive episodes, no mathematical model has been developed to describe the durations of major depressive episodes. We evaluated which mathematical model is fitted to describe the distribution of the durations of major depressive episodes using data from the Group for Longitudinal Affective Disorder Study (GLADS), a prospective study conducted in Japan.

**Results:** The distribution of the cumulative probability of major depressive disorder duration plotted on a double-logarithmic scale exhibited an approximately linear form. A log-normal distribution fit the distribution of major depressive episodes better than an exponential distribution or a Weibull distribution.

**Conclusions:** In this study, we evaluated which mathematical model fit the distribution of major depressive episode durations using data from GLADS. The results showed that a log-normal model and a power law model may fit the distribution of major depressive episode durations.

**Keywords:** Depression, Log-normal distribution, Power law distribution, Mathematical model

## Background

How long does an episode of unipolar major depression last? Studies available to date suggest that the median time to recovery of a major depressive episode is 3–12 months in cohort studies [1,2], with rates of chronicity (a duration of 24 months or more) of between 10% and 30% [3,4]. These considerable rates of chronicity suggest a long-tailed distribution, which can be partly explained by the fact that longer durations of depression are associated with lower rates of recovery [5].

Despite the accumulation of knowledge regarding major depressive episodes, there seems to be no overarching theory to explain the distribution of major depressive episode durations. The difficulty of developing a model describing the distribution of major depressive episode durations lies in the diversity of the results of each study. The indexes of distribution, such as the median time to recovery and the rates of chronicity, differ widely among previous studies. Furthermore, since a

portion of depressive episodes exceeds the given period of the cohort studies, the average or variance of the distribution of major depressive episodes is difficult to estimate.

A survival analysis is used to estimate the distributions of major depressive episode durations. In general, survival curves are governed by non-normal distribution models, such as an exponential distribution, Weibull distribution, or log-normal distribution [6]. However, which model best fits the distribution of major depressive episode durations has received little attention.

Whereas the indexes of the distributions of major depressive episodes differ widely among previous studies, the survival curves of major depressive episode durations appear to exhibit a similar shape characterized by a positively skewed distribution with a long tail [1-5]. The similarity of the survival curve shapes suggests that major depressive episode durations may follow a specific mathematical distribution.

We evaluated whether a mathematical model can be used to describe the distribution of major depressive episode durations using data from the Group for Longitudinal Affective Disorder Study (GLADS), a prospective study conducted in Japan [3].

\* Correspondence: tomitaka.shinichiro@jp.panasonic.com

<sup>1</sup>Department of Mental Health, Panasonic Health Center, Landic building 3F, Nishishinbashi 3-8-3, Minato-ku, Tokyo 105-0003, Japan

Full list of author information is available at the end of the article

## Methods

### Analysis data

We used data from GLADS, which was a 2-year multi-center prospective follow-up study of patients with a mood disorder: the patients were selected so as to be representative of untreated, first-visit patients at 23 psychiatric settings from all over Japan. The details of GLADS have been published elsewhere [3]. The GLADS study was approved by the ethics committees of Nagoya City University School of Medicine and Tokyo Women's Medical University Hospital and present study has been permitted to use the data. Written informed consent was obtained from all the subjects. The cohort consisted of 90 subjects who were diagnosed as suffering from unipolar major depressive disorder, not superimposed on dysthymic disorder, according to the DSM-IV. Each patient in the cohort was treated with, on average, 60 mg of imipramine or equivalent per day upon entry and 85 mg per day at 1 month. Recovery from a major depressive episode was defined in accordance with the definition of the US National Institute for Mental Health (NIMH) as 2 consecutive months with no more than one or two symptoms of mild depression [5]. Patients who recovered within a few days after the commencement of treatment were regarded as having recovered at 0.5 months.

### Statistics

The cumulative probability of the durations of major depressive episodes was estimated using the Kaplan Meier method. A set of the estimated cumulative probabilities of the major depressive episode durations was plotted on probability paper for an exponential distribution, a Weibull distribution, and a log-normal distribution. A

fitting curve was calculated using the maximum likelihood estimation method.

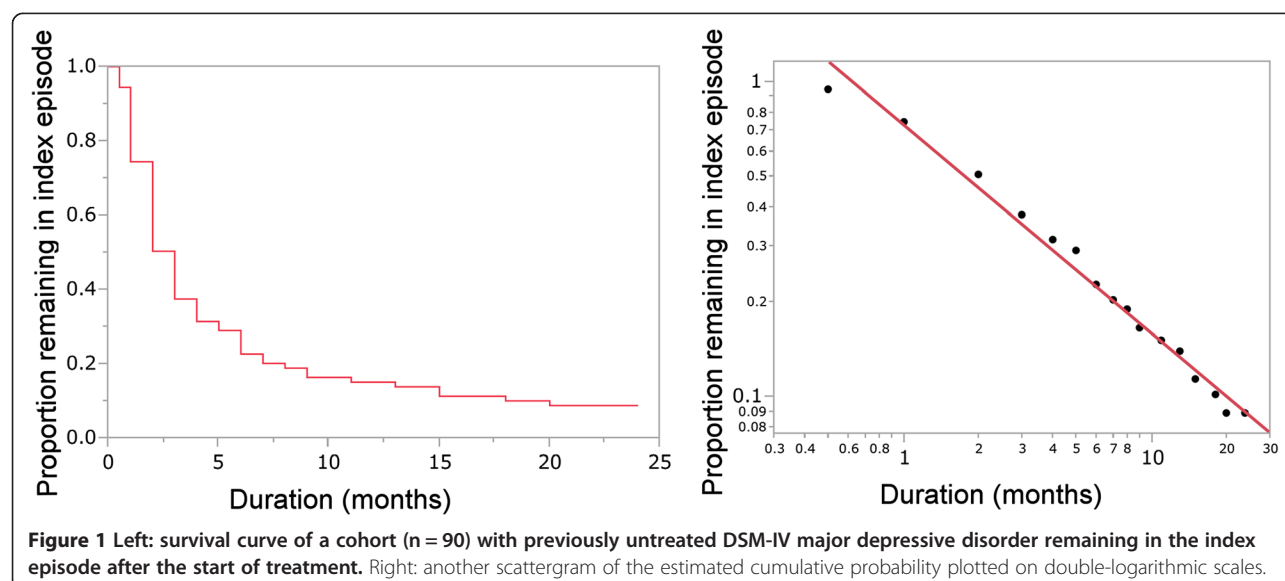
Since the survival curve showed an approximately linear pattern when plotted using a double-logarithmic scale, the power law trend line was calculated using the least squares method. JMP version 11 for Windows was used for all the statistical analyses.

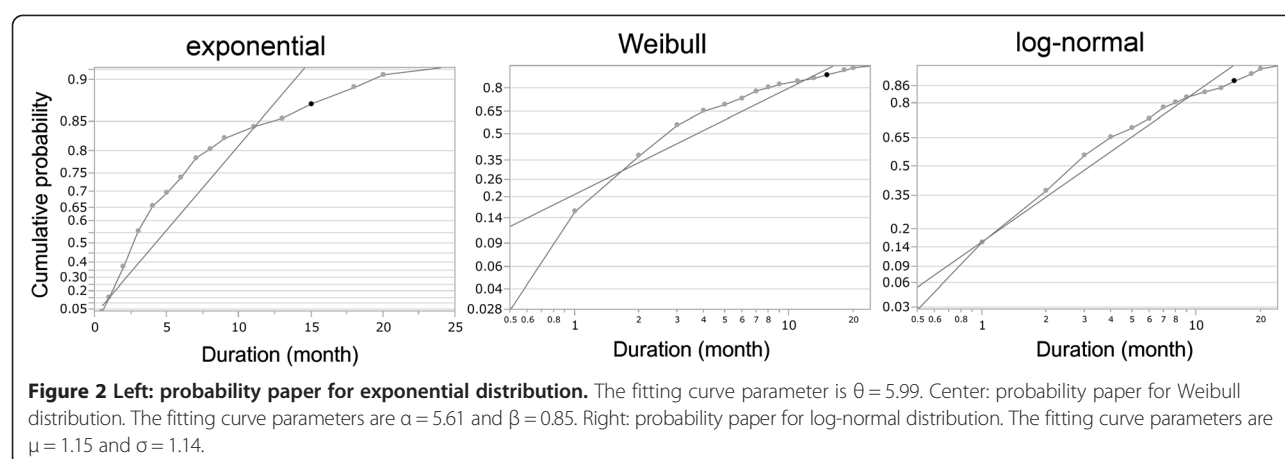
## Results

Figure 1 shows the cumulative probability of the previously untreated subjects with DSM-IV major depressive disorder ( $n = 90$ ) remaining in the index episode after treatment commencement. The median time to recovery from the index episode after treatment was 3 months: 26% of the cohort reached an asymptomatic or minimally symptomatic status by 1 month, 63% by 3 months, 85% by 12 months, and 91% by 24 months.

A set of estimated cumulative probabilities of major depressive episode durations plotted on double-logarithmic scales (Figure 1: Left) showed an approximately linear relation. The trend line was calculated using a power law model and the observed data ( $\log y = -0.209731 - 0.7090606 \cdot \log x$ ).

To evaluate the model of the distribution, a set of estimated cumulative probabilities of major depressive episode durations was plotted on probability paper for an exponential distribution, a Weibull distribution, and a log-normal distribution (Figure 2). The markers on the probability paper tend to follow a straight line when the distributional fit is suitable for the data. Figure 2 shows that the log-normal plot was the closest to a straight line among the three distributions, indicating that a log-normal distribution had the best fit with the observed data.





## Discussion

The goal of this study was to assess which model best fit the survival curve for major depressive episode durations. Our findings showed that a log-normal distribution fit the distribution of major depressive episode durations better than that of an exponential distribution or a Weibull distribution.

A log-normal distribution is a probability distribution for a random variable for which the logarithmic values are normally distributed. Empirically, a log-normal model is accepted as a survival curve model [7]. In fact, a variety of medical examples fit a log-normal distribution. For example, the latent periods (time from infection to first symptoms) of infectious diseases [8,9], the survival time after a diagnosis of cancer [10,11], and the age of onset of diseases and the duration of elderly disabilities [12] have been shown to fit a log-normal distribution.

The underlying process of a log-normal distribution is regarded as a “multiplicative process,” whereas a normal distribution is generated by an “additive process” [13]. When a process requires the completion of sequential subprocesses, the probability of the success or failure of the primary process is considered to show a log-normal distribution because of the central limit theorem [14]. Inversely, if the survival curves of a disease follow a log-normal distribution, disease recovery or progression is thought to occur according to a multiplicative process.

The finding that the distribution of major depressive episode durations fitted a log-normal distribution best suggests that recovery from a major depressive disorder may occur according to a multiplicative process. Environments, genotypes, treatments, and biological systems may be involved in the recovery of major depressive episodes on a multiplicative, not additive, basis.

Interestingly, the distribution of the cumulative probabilities of major depressive disorder durations plotted on a double-logarithmic scale exhibited an approximately linear form, indicating that the duration of depressive

episodes seems to follow a power law distribution. However, as a survival curve model, the use of a power law distribution is obscure.

It is not surprising that both a log-normal distribution and a power law distribution arose together as possible distributions. Log-normal distributions and power law distributions are connected both theoretically and empirically [15]. Whereas the multiplicative model is used to generate a log-normal distribution, only a small change from a log-normal generative process yields a generative process with a power law distribution. As long as there is a bounded minimum that acts as a lower reflective barrier to the multiplicative model, a power law distribution, rather than a log-normal distribution, will be generated [16]. As a matter of fact, arguments over whether a log-normal or a power law distribution is a better fit for some empirically observed distributions have been repeated across many fields, including financial models [17], biology [18], chemistry [19], ecology [20], and information theory [21].

A wide range of data is necessary to determine whether a log-normal or a power law distribution is a better fit for empirically observed distributions. Whereas a log normal distribution will appear as a nearly straight line, similar to a power law distribution, when plotted using a double-logarithmic scale for a large portion of the body of the distribution, the log-normal distribution will appear to have a curved nature in the tail region [16]. However, data for major depressive episode durations of more than two years were not available in our study. Because of this limitation, we were unable to determine whether a log-normal or a power law distribution was a better fit to the observed distribution.

This study has some limitations. First, the sample size ( $N = 90$ ) of the data from the GLADS is relatively small for analysis. Further studies with large sample size are needed to establish our findings. Second, this analysis did not examine any potential factors, such as type or dose of antidepressant, comorbidity and personality, which

influence the major depressive episode durations. Finally, although a log-normal distribution had the best fit with the observed data, we did not perform authentic curve fitting.

Mathematically, there must be a necessary condition for a power law to arise during recovery from major depression. The instantaneous change in the cumulative probability of a cohort remaining in the index episode was defined as  $\frac{1}{N} \frac{dN}{dt}$ , where  $N$  is the number of patients remaining in the index episode and  $\frac{dN}{dt}$  is the derivative of the number of patients remaining in the index episode. When  $\frac{1}{N} \frac{dN}{dt}$  is inversely proportional to the duration of the depressive episodes, the following formula can be produced:

$$\frac{1}{N} \frac{dN}{dt} = -k \frac{1}{t}, \quad (1)$$

where  $k$  is a constant and  $N, t, k > 0$ .

The following power law equation can then be obtained by solving the differential equation (1):

$$N = At^{-k}, \quad (2)$$

where  $A$  is a constant  $> 0$ .

Formula (1) corresponds to the clinical finding that the instantaneous probability of recovery decreases as the duration of the major depressive episode lengthens [5].

From a pragmatic point of view, the finding that a cumulative probability distribution for durations within two years appears as an approximately straight line on double-logarithmic scales is important regardless of the exact model describing the distribution. We conducted an analysis of several cohort studies [1-5] and found that the survival curves of other cohort studies also appeared as an approximately linear line (data not shown). However, the lack of individual patient data prevented us from examining their fit to the models used in this review. To confirm the finding that the cumulative probability distribution appears as an approximately straight line on double-logarithmic scales, further reviews are necessary.

Mathematical models describing diseases are necessary not only for making predictions regarding the course of disease, but also for shedding light on the mechanisms of disease. Our finding will be helpful to developing a model describing the duration of major depressive disorder.

## Conclusion

Our findings showed that a log-normal distribution fit the distribution of major depressive episode durations better than that of an exponential distribution or a Weibull distribution. The distribution of the cumulative probabilities of major depressive disorder durations plotted on a double-logarithmic scale exhibited an approximately linear form. A log-normal model and a power law model may fit the distribution of major depressive episode durations.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

ST carried out the design of the study and the statistical analysis. TF participated in the collection of data and general supervision of the study. All authors read and approve the final manuscript.

## Acknowledgements

We wish to acknowledge the GLADS cohort team. We also thank Dr Tanaka and Dr Kawasaki (Biostatistics Section, Department of Clinical Research and Informatics, Clinical Science Center, National Center for Global Health and Medicine) for their helpful advices.

## Author details

<sup>1</sup>Department of Mental Health, Panasonic Health Center, Landic building 3F, Nishishinbashi 3-8-3, Minato-ku, Tokyo 105-0003, Japan. <sup>2</sup>Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan.

Received: 25 November 2013 Accepted: 9 September 2014

Published: 12 September 2014

## References

1. Angst J: **Prediction of course and outcome.** In *Depressive illness*. Edited by Helgason T, Daly RJ. Berlin: Springer-Verlag; 1988:10–11.
2. Coryell W, Akiskal HS, Leon AC, Winokur G, Maser JD, Mueller TI, Keller TI: **The time course of nonchronic major depressive disorder. Uniformity across episodes and samples.** *Arch Gen Psychiatry* 1994, **51**:405–410.
3. Furukawa TA, Kitamura T, Takahashi K: **Time to recovery of an inception cohort with hitherto untreated unipolar major depressive episodes.** *Br J Psychiatry* 2000, **177**:331–335.
4. Spijker J, de Graaf R, Bijl RV, Beekman AT, Ormel J, Nolen WA: **Duration of major depressive episodes in the general population: results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS).** *Br J Psychiatry* 2002, **181**:208–213.
5. Keller MB, Lavori PW, Mueller TI, Endicott J, Coryell W, Hirschfeld RMA, Shea T: **Time to recovery, chronicity, and levels of psychopathology in major depression: a 5-year prospective follow-up of 431 subjects.** *Arch Gen Psychiatry* 1992, **49**:809–816.
6. Klein JP, Zhang MJ: *Survival Analysis, Software. Encyclopedia of Biostatistics.* 111 River Street Hoboken, NJ USA: Wiley Online Library; 2005.
7. Lee ET, Wang J: *Statistical methods for survival data analysis.* 3rd edition. Hoboken: Wiley; 2003.
8. Sartwell PE: **The distribution of incubation periods of infectious disease.** *Am J Hygiene* 1950, **51**:310–318.
9. Sartwell PE: **The incubation period of poliomyelitis.** *Am J Public Health Nation's Health* 1952, **42**:1403–1408.
10. Royston P: **The Lognormal Distribution as a Model for Survival Time in Cancer, With an Emphasis on Prognostic Factors.** *Statistica Neerlandica* 2001, **55**:89–104.
11. Tai P, Tonita J, Yu E, Skarsgard D: **Twenty-year follow-up study of long-term survival of limited-stage small-cell lung cancer and overview of prognostic and treatment factors.** *Int J Radiat Oncol Biol Phys* 2003, **56**:626–633.
12. Moriyama O, Itoh H, Matsushita S, Matsushita M: **Long-tailed duration distributions for disability in aged people.** *J Phys Soc Jpn* 2003, **72**:2409–2412.
13. Gibrat R: **Une loi des r'eparations 'economiques: l'effet proportionnel.** *Bull Statist g'en Fr* 1930, **19**:469.
14. Montroll EW, Shlesinger MF: **Maximum entropy formalism, fractals, scaling phenomena, and 1/f noise: a tale of tails.** *J Stat Phys* 1983, **32**:209–230.
15. Mitzenmacher M: **A Brief History of Generative Models for Power Law and Lognormal Distributions.** *Internet Math* 2004, **1**:226–251.
16. Champenowne D: **A Model of Income Distribution.** *Economic J* 1953, **63**:318–351.
17. Mandelbrot B: *Fractals and Scaling in Finance.* New York: Springer-Verlag; 1997.
18. Jain R, Ramakumar S: **Stochastic dynamics modeling of the protein sequence length distribution in genomes: Implications for microbial evolution.** *Physica A* 1999, **273**:476–485.

19. Nakajima T, Higurashi A: A Use of two-channel radiances for an aerosol characterization from space. *Geophysical Res Lett* 1998, **25**:3815–3818.
20. Allen AP, Li B, Charnov EL: Population: Fluctuations, power Laws and mixtures of lognormal distributions. *Ecol Lett* 2001, **4**:1–3.
21. Li W: Random texts exhibit Zipf's-law-like word frequency distribution. *IEEE Trans Inf Theory* 1992, **38**:1842–1845.

doi:10.1186/1756-0500-7-636

**Cite this article as:** Tomitaka and Furukawa: Mathematical model for the distribution of major depressive episode durations. *BMC Research Notes* 2014 **7**:636.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

