**International Journal of Theoretical and Applied Mathematics**
2016; 2(2): 74-78
http://www.sciencepublishinggroup.com/j/ijtam
doi: 10.11648/j.ijtam.20160202.17

# The Type I Generalized Half Logistic Survival Model

**Phillip Oluwatobi Awodutire[1, *], Akintayo Kehinde Olapade[2], Oladapo Adedayo Kolawole[3]**

[1]Department of Statistics, Federal Polytechnic of Oil and Gas, Bonny, Nigeria

[2]Department of Mathematics, Obafemi Awolowo University, Ile-Ife, Nigeria

[3]Department of Surgery, Ladoke Akintola University of Technology Teaching Hospital, Osogbo, Nigeria

## Email address:

phillip.awodutire@gmail.com (P. O. Awodutire), akolapad@oauife.edu.ng (A. K. Olapade), cullerwarley@yahoo.co.uk (O. A. Kolawole)
[*]Corresponding author

## To cite this article:

Phillip Oluwatobi Awodutire, Akintayo Kehinde Olapade, Oladapo Adedayo Kolawole. The Type I Generalized Half Logistic Survival Model. *International Journal of Theoretical and Applied Mathematics.* Vol. 2, No. 2, 2016, pp. 74-78. doi: 10.11648/j.ijtam.20160202.17

**Received**: October 26, 2016; **Accepted**: November 10, 2016; **Published**: December 9, 2016

**Abstract:** In this paper, a three parameter probability distribution function called type I generalized half-logistic distribution is introduced to model survival or time to event data. The survival function, hazard function and median survival time of the survival model were established. Estimation of the parameters of the model was done using the maximum likelihood method. We then applied the type I generalized half-logistic survival model to a breast cancer survival data. The derived result from type I generalized half logistic survival model was compared with the results of some common existing parametric survival models, and this revealed that the type I generalized half-logistic survival model clearly demonstrates superiority over these other models.

**Keywords:** Parameter, Survival Function, Hazard Function, Model, Half Logistic Distribution, Survival Time

## 1. Introduction

Survival analysis is a statistical tool for studying the occurrence and timing of events. Initially, survival analysis was designed for study of death, but it has now been extended to other aspects both in social and natural sciences. It measures time to event e.g death, contacting a disease, equipment failures, earthquakes, automobile accidents, stock market crashes, revolutions, job terminations, births, marriages, divorces, promotions and arrests e.t.c. We may be interested in characterizing the distribution of time to event for a given population as well as comparing this time to event among different groups (e.g., treatment vs. control in a clinical trial or an observational study), or modeling the relationship of time to event to other covariates (sometimes called prognostic factors or predictors). The Hosmer and Lemeshow [1], Lee and Wang [2], Kleinbaum and Klein [3], and Collet [4] books give a detailed overview of survival data modeling techniques. Non-parametric and semi-parametric survival models such as the Cox regression analysis have been the most widely used models in the analysis of time to event survival data. On the other hand, if the assumption for parametric probability distribution is met for the data set under consideration, it will result in more efficient and easier to interpret estimates than non-parametric or semi parametric models. A comprehensive review was given by Efron [5] and Lee and Go [6]. Popular parametric models include the exponential, Weibull, log- logistics and lognormal distributions. The description of the distribution of the survival times and the change in their distribution as a function of predictors is of interest. Model parameters in these settings are usually estimated using maximum likelihood method. Survival estimates obtained from parametric survival models typically yield plots more consistent with a theoretical survival curve. If the investigator is comfortable with the underlying distributional assumption, then parameters can be estimated that completely specify the survival and hazard functions. This efficiency and completeness are the main appeals of using a parametric approach. Foulkes et al. [7] used parametric modeling to assess the prognostic factors in the recurrence of ischemic strokes. Sama et al. [8] used five parametric models to analyze the survival time data of infections and they found that the best fit could be obtained using parametric models. They also indicated that parametric models can be used to model the duration of malaria infections. Kannan et al. [9] used log-logistic probability distribution to model altitude decompression sickness (DCS) risk and symptom onset time.

They concluded that the log-logistic model could provide good estimates of the probability of DCS over time. The AFT model is often viewed as a competitor to the proportional hazards (PH) model when the PH model fails to fit. Although the AFT model with an unspecified error distribution is a natural analog to the PH model, there are no widely accepted methods for implementing this approach. Miller [10], Buckley and James [11], Koul et al. [12] and Christensen and Johnson [13] developed methods for a semi-parametric AFT model, but difficulties inherent in the model made their approaches somewhat un- palatable. Kuo and Mallick [14] and Walker and Mallick [15] present novel Bayesian approaches but it will take time to determine their long-term viability. Aalen [16] noted that parametric survival model is underused in medical research and deserving of more attention. Mohammad A Tabatabai et al. [17] derived a new two-parameter probability distribution called hypertabastic and introduced it to model the survival or time-to-event data. Torabi and Bagheri [18] presented an extended generalized half logistic distribution and studied different methods for estimating its parameter based on complete and censored data. They derived maximum likelihood equations for estimating the parameters based on censored data. Also, the asymptotic confidence intervals of the estimators are presented in which they applied using simulation studies and properties of maximum likelihood of the estimators are given. Alireza Abadi [19] applied the generalized gamma distribution to a set of data of breast cancer patients. He fitted the saturated generalized gamma (GG) distribution, and compared this with the conventional Acceleration Failure Time (AFT) model. Using a likelihood ratio statistic, both models were compared to the simpler forms including the Weibull and lognormal. For semiparametric models, either Cox PH model or stratified Cox model was fitted according to the PH assumption and tested using Schoenfeld residuals. The GG family was compared to the log-logistic model using Akaike information criterion (AIC) and Bayesian information criterion (BIC).

# 2. Type I Generalized Half Logistic Distribution

In this section we introduce the four parameter type I generalized half logistic distribution. One of the probability distributions, which is a member of the family of logistic distribution, is the half logistic distribution with probability density function (pdf).

$$f(t) = \frac{2e^{t}}{(1+e^{t})^{2}} \quad t \geq 0 \tag{1}$$

and cumulative distribution function as

$$F(t) = \frac{e^{t}-1}{1+e^{t}} \ t \geq 0 \tag{2}$$

Balakrishnan [20] studied order statistics from the half logistic distribution, Balakrishnan and Puthenpura [21] obtained best unbiased estimates of the location and scale parameter of the distribution while Olapade [22] presented some theorems that characterized the distribution. Balakrishnan and Wong [23] obtained approximate maximum likelihood estimates for the location and scale parameters of the half logistic distribution. Torabi and Bagheri [18] presented an extended generalized half logistic distribution and studied different methods for estimating its parameters based on complete and censored data. Olapade [24] obtain a generalized form of half logistic distribution through a transformation of an exponential random variable called four parameter type I half logistic distribution as

$$f(t) = \alpha 2^{\alpha}\left[\frac{\exp(\frac{t-\mu}{\sigma})}{\sigma\left(1+\exp(\frac{t-\mu}{\sigma})\right)^{\alpha+1}}\right] \ \alpha, \lambda, \mu, \sigma > 0 \tag{3}$$

with cumulative distribution function as

$$F(t) = 1 - \left[\frac{2}{1+\exp\left(\frac{t-\mu}{\sigma}\right)}\right]^{\alpha} \ \alpha, \lambda, \mu, \sigma > 0 \tag{4}$$

### *2.1. Type I Generalized Half Logistic Survival Function*

Let T be a continuous random variable representing the waiting time until the occurrence of an event. Then the four parameter type I generalized half logistic survival function s(t) is defined as

$$s(t) = 1 - F(t) \tag{5}$$

$$s(t) = 1 - \ 1 + \left[\frac{2}{\exp\left(\frac{t-\mu}{\sigma}\right)+1}\right]^{\alpha} \tag{6}$$

$$s(t) = \left[\frac{2}{\exp(\frac{t-\mu}{\sigma})+1}\right]^{\alpha} \tag{7}$$

Figure 1 shows the graph shows different shapes of the curve with different values of the shape parameter.

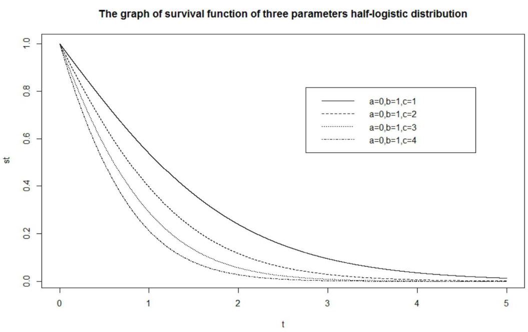


***Fig. 1.*** *Graph of survival function of type I generalized half logistic distribution.*

### *2.2. The Type I Generalized Half Logistic Hazard Function*

The hazard function is a conditional failure rate which gives the instantaneous potential for failing at time t per unit time for an individual surviving to time t.

$$h(t) = \frac{f(t)}{s(t)} \tag{8}$$

Let T be a continuous random variable representing the waiting time until the occurrence of an event. Then the type I generalized half logistic hazard function h(t) which is the instantaneous failure rate at time t given survival up to time t is defined as

$$h(t) = \frac{\alpha 2^{\alpha}\left[\frac{\exp(\frac{t-\mu}{\sigma})}{\sigma\left(1+\exp(\frac{t-\mu}{\sigma})\right)^{\alpha+1}}\right]}{\left[\frac{2}{\exp(\frac{t-\mu}{\sigma})+1}\right]^{\alpha}} \tag{9}$$

which gives

$$h(t) = \frac{\alpha\exp(\frac{t-\mu}{\sigma})}{\sigma(1+\exp\left(\frac{t-\mu}{\sigma}\right))} \tag{10}$$

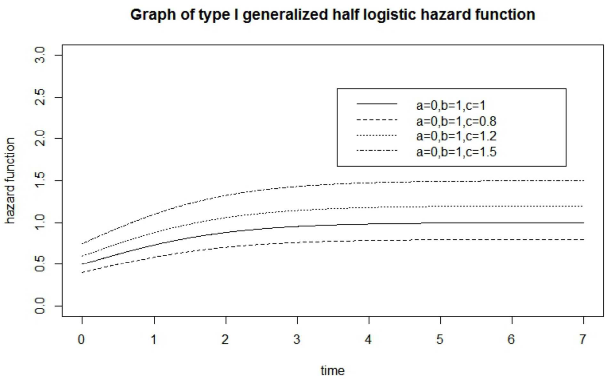


***Fig. 2.*** *Graph of survival function of type I generalized half logistic distribution.*

From Figure 2, one thing to note is that the hazard function has a monotonically increasing hazard rate for all parameter values, a property shared by relatively few distributions which have support on the positive real half-line.

### *2.3. The Median Survival Time of the Type I Generalized Half Logistic Survival Model*

The median survival time is defined as

$$s(t_{0.5}) = \frac{1}{2} \tag{11}$$

$$\left[\frac{2}{\exp(\frac{t_{0.5}-\mu}{\sigma})+1}\right]^{\alpha} = \frac{1}{2} \tag{12}$$

$$2(2)^{\alpha} = \left(\exp(\frac{t_{0.5}-\mu}{\sigma})+1\right)^{\alpha} \tag{13}$$

$$\sqrt[\alpha]{2}\,(2) = \exp\left(\frac{t_{0.5}-\mu}{\sigma}\right)+1 \tag{14}$$

$$\exp\left(\frac{t_{0.5}-\mu}{\sigma}\right) = \sqrt[\alpha]{2}\,(2)-1 \tag{15}$$

$$\frac{t_{0.5}-\mu}{\sigma} = \log\left(\sqrt[\alpha]{2}\,(2)-1\right) \tag{16}$$

$$t_{0.5} = \mu+\sigma\left(\log\left(\sqrt[\alpha]{2}\,(2)-1\right)\right) \tag{17}$$

### *2.4. Accelerated Failure Time Model (AFT) of the Type I Generalized Half Logistic Survival Model*

Generally, the AFT model is of the form

$$lnT = \delta + x'\beta + \varepsilon \tag{18}$$

where

1. ε is said to follow a distribution
2. $x$ is sets of covariates with parameters β
3. δ is the intercept of the model

Given that survival times $T_1, T_2, \ldots, T_n$ of size n is assumed to follow the type I generalized half-logistic distribution, the ε is assumed to be distributed to exponentiated Type I half logistic distribution. We estimate the parameters of the AFT model using the maximum likelihood estimation (MLE) method, involving optimization technique.

The method of the MLE is to find an estimator that maximizes the likelihood function, or in other words, which is "most likely" to have produced the observed data. The maximum likelihood function for the AFT model is given as

$$L = \prod_{i=1}^{r} f(t_i/x,\beta) + \prod_{r+1}^{n} s(t_i^{+}/x,\beta) \tag{19}$$

which equivalently is

$$L = \prod_{i=1}^{r} g(x/\beta) f_o\left(Z(t_i)\right) + \prod_{r+1}^{n} s_o\left(Z(t_i)\right) \tag{20}$$

where $g(x/\beta)$=exp$(-x\beta)$ and $Z(t) = t\exp(-x\beta)$.

Also $f_o(Z(t))$ and $s_o(Z(t))$ are the baseline probability density function and baseline survival function respectively. This now gives

$$\mathrm{L} = g(x/\beta)^{r}\,\frac{(2)^{r\alpha}\alpha^{r}\exp\sum_{i=1}^{r}\left(\frac{Z(t_i)-\mu}{\sigma}\right)}{\sigma^{r}\prod_{i=1}^{r}\left(1+\exp\left(\frac{Z(t_i)-\mu}{\sigma}\right)\right)^{\alpha+1}} + \frac{(2)^{\alpha(n-r)}}{\prod_{i=r+1}^{n}\left(1+\exp\left(\frac{Z(t_i^{+})-\mu}{\sigma}\right)\right)^{\alpha}} \tag{21}$$

Taking logarithm of both sides,

$$lnL = l = r(-x'\beta) + rln\alpha + r\alpha ln2 + \sum_{i=1}^{r}\left(1+\exp\left(\frac{Z(t_i)-\mu}{\sigma}\right)\right) - rln\sigma - (\alpha+1)\sum_{i=1}^{r} ln\left(1+\exp\left(\frac{Z(t_i)-\mu}{\sigma}\right)\right) + \alpha(n-r)\,ln(2) - \alpha\sum_{i=r+1}^{n} ln\left(1+\exp\left(\frac{Z(t_i{}^{+})-\mu}{\sigma}\right)\right) \tag{22}$$

we obtain the maximum likelihood estimate of the shape parameter α as

$$\frac{dl}{d\propto} = \frac{r}{\propto} + rln(2) - \sum_{i=1}^{r}\left(1+\exp\left(\frac{Z(t_i)-\mu}{\sigma}\right)\right) + (n-r)\ln 2 - \sum_{i=r+1}^{n} ln\left(1+\exp\left(\frac{Z(t_i{}^{+})-\mu}{\sigma}\right)\right) \tag{23}$$

Equating this to zero to derive α,

$$= \frac{r}{\sum_{i=1}^{r}\left(1+\exp\left(\frac{Z(t_i)-\mu}{\sigma}\right)\right)-(n-r)\ln(2)-rln(2)+\sum_{i=r+1}^{n} ln\left(1+\exp\left(\frac{Z(t_i^{+})-\mu}{\sigma}\right)\right)} \tag{24}$$

To estimate the location parameter μ, and scale parameter σ,

$$\frac{dL}{d\mu} = -\frac{r}{\sigma} + \frac{\alpha+1}{\sigma}\sum_{i=1}^{r}\frac{\exp\left(\frac{Z(t_i)-\mu}{\sigma}\right)}{\left(1+\exp\left(\frac{Z(t_i)-\mu}{\sigma}\right)\right)} + \frac{\alpha}{\sigma}\sum_{i=r+1}^{n}\frac{\exp\left(\frac{Z(t_i^{+})-\mu}{\sigma}\right)}{\left(1+\exp\left(\frac{Z(t_i^{+})-\mu}{\sigma}\right)\right)} \tag{25}$$

$$\frac{dL}{d\sigma} = -\frac{r}{\sigma} - \frac{\sum_{i=1}^{r} Z(t_i)-\mu}{\sigma^2} + \frac{\alpha+1}{\sigma^2}\sum_{i=r+1}^{n}(Z(t_i^{+})-\mu)\frac{\exp\left(\frac{Z(t_i^{+})-\mu}{\sigma}\right)}{\left(1+\exp\left(\frac{Z(t_i^{+})-\mu}{\sigma}\right)\right)} \tag{26}$$

Since the equations (25) and (26) above are nonlinear in the parameters, we can use numerical iterative method with the aid of computer program to estimate the parameters from a relevant data sample.

## 3. Application of the Type I Generalized Half Logistic Survival Model

To investigate the performance of the type I generalized half logistic survival model and to compare it with common existing models, we retrospectively analysed a set of data for breast cancer patients from Ladoke Akintola University of Technology Teaching Hospital (LTH), Osogbo. The data contain information on 89 patients with breast cancer disease. The data is right censored and the survival time is measured in days, starting from the date of admission for treatment and the date of last contact (death, alive or loss to follow up). The histogram of the data shows that the data can be analyzed using a type I generalized half logistic or the half logistic distribution of Olapade [24]. Besides the type I generalized half logistic survival model, we fit Weibull, log-normal, log-logistic and exponential models and the figure 3 shows that the models fit the data. We present the log-likelihood estimates, as well as AIC, to assess the goodness-of-fit of the models. The data was analyzed using the R-statistical software and the result is derived.

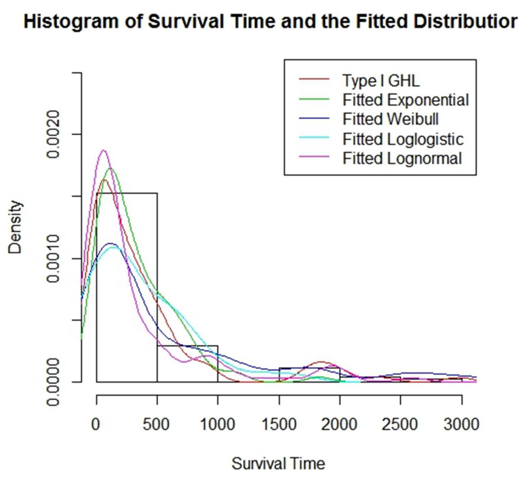


***Fig. 3.*** *Graph displaying the comparison of the fitted survival models.*

The parametric estimation estimates the parameter variables. Using the survival data collected in the hospital, we have the μ= 396.842161, σ= 586.200871 and α= 5.803715. The loglikelihood value is -336.214 and the AIC = 678.428.

Comparing it with common existing survival models, the result is displayed in the table 1

***Table 1.*** *Table displaying comparison of Type I Generalized Half Logistic Survival Model with common existing models.*

| **Model** | **Loglikelihood** | **AIC** |
|---|---|---|
| Type I GHL | -336.2 | 678.4 |
| Exponential | -374.3 | 773.2 |
| Log-normal | -337.9 | 693.2 |
| Log-logistic | -338.7 | 695.8 |
| Weibull | -348.2 | 702.4 |

From table 1, the Type I GHL Survival model (with AIC value at 674.4) gives the smallest value and thus is the best model when compared with the common existing ones.

## 4. Conclusion

In this paper, we have introduced a new survival model called the four parameter type I generalized survival model and presented some theoretical work on it.

## References

[1] Hosmer, D. W. and Lemeshow S. (1999): Applied Survival Analysis; Regression Modeling of Time to Event Data. New York: John Wiley Sons.

[2] Lee, E. T. and Wang, J. W. Statistical Methods for Survival Data Analysis 3rd Edition. New York: Wiley; 2003.

[3] Kleinbaum, D. G. and Klein M. (2004): Survival Analysis: A self-Learning Text. New York: Springer - Verlag; 2004.

[4] Collet, D. (1994): Modeling Survival Data in Medical Research New York: Chapman and Hall/CRC

[5] Efron, B. (1977): The Efficiency of Cox's Likelihood Function for Censored Data. Journal of the American Statistical Association.

[6] Lee, E. T., Go, O. T. (1997): Survival Analysis in Public Health Research. Annual Reviews in Public Health, 18:105-134.

[7] Foulkes M. A., Sacco R. L., Mohr J. P., Hier, D. B., Price, T. R., Wolf, P. A. (1994): Parametric Modeling of Stroke Recurrence. Neuroepidemiology 13:19-27.

[8] Sama, W., Dietz, K., Smith. T. (2006): Distribution of Survival Times of Deliberate Plasmodium Falciparum Infections in Tertiary Syphilis Patients. Transaction of the Royal society of Tropical Medicine and Hygiene, 100: 811-816.

[9] Kannan, N., Raychaudhuri, A., Pilamanis, A. A. (1998): A Log Logistic Model for Altitude Decomposition Sickness. Aviation, Space and Environmental Medicine 69:965-970.

[10] Miller, R. G. (1976): Least squares regression with censored data. Biometrika 63, 449-464.

[11] Buckley, J. and James, I. (1979): Linear regression with censored data. Biometrika 66, 429-436.

[12] Koul,H., Susarla, V. and Van Ryzin, J. (1981): Regression Analysis with Randomly Right Censored Data. Annals of Statistics 8, 1276-1288.

[13] Christensen, R. and Johnson, W. (1988): Modeling Accelerated Failure Time with a Dirichlet Process. Biometrika 75, 793-704.

[14] Kuo, L. and Mallick, B. (1997): Bayesian Semiparametric Inference for Accelerated Failure Time Model. Canadian Journal of Statistics 25, 457-472.

[15] Walker, S. and Mallick, B. K. (1999): A Bayesian Semiparametric Accelerated Failure Time Model. Biometrics 55, 477-483.

[16] Aalen, O. O. (2000): Medical statistics-no time for complacency. Statistical Methods in Medical Research 9, 31-40.

[17] Mohammad, A. T., Zoran, B., David, K. W. and Karan, P. S. (2007): Hypertabastic Survival Model. [http://www.biomedcentral.com/content/supplementary/1742-4682-4-40-S1.doc]

[18] Torabi, H. and Bagheri, F. L. (2010): Estimation of Parameters for an Extended Generalized Half Logistic Distribution Based on Complete Censored data. JIRSS Vol. 9, No. 2, pp 171-19.

[19] Alireza, A. (2012): Breast Cancer Survival Analysis : Applying the Generalized Gamma Distribution under Different Conditions of the Propotional Hazads and Accelerated Failure Time Assumptions. International Journal of Preventive Medicine, Vol 3, No 9.

[20] Balakrishnan, N. (1985): Oder statistics from the Half Logistic Distrinution. Journal of Statistics and Computer Smulation, 20, 27-309.

[21] Balakrishnan, N. and Puthenpura, S. (1986): Best Linear Unbiased Estimators of Location and Scale Parameters of the Half Logistic Distribution. Journal of Statistical Computation and Simulation, 25, 193-204.

[22] Olapade, A. K. (2003): On Characterizations of the Half Logistic Distribution. InterStat, February Issue, 2, http://interstat.stat.vt.edu/ InterStat/ARTICLES/2003articles/F06002.pdf

[23] Balakrishnan, N. and Wong, K. H. T. (1991): Approximate MLE's for the Location and Scale Parameters of the Half-Logistic Disribution with Type II Right Censoring. IEEE Transactions on Reliability, 40(2), 140-145.

[24] Olapade A. K. (2014): The type I generalized half logistic distribution. JIRSS Vol. 13, No. 1, pp 69-82.