Check for
updates

Parametric Regression Model Based on Reversed Hazard Rate: An Application to left censored heavy tailed HIV Viral Load Data

Akhtar Hossain[1] · **Farahnaz Islam**[1] · **Hrishikesh Chakraborty**[2]

Abstract

The parametric survival model with Weibull distribution can be used to model a wide range of practical lifetime data. While there have been several studies comparing the fit of various distributions to right-censored and interval censored data, there are no recommendations in the literature on optimal distributions to use for left-censored heavy-tailed data. Parametric Reverse Hazards (PRH) has gained considerable attention from time-to-event data researchers for its excellent properties and appropriateness to analyzing left-censored survival data. To analyze left-censored with heavy-tailed data, we derived the PRH model for a variety of distributions including the Exponential, Log-normal, Inverse Gaussian, Log-logistic, Gompertz–Makeham, Gamma, Generalized Gamma, Inverse Gamma, Generalized Inverse Gamma, Weibull, Inverse Weibull, Generalized Inverse Weibull, Modified Weibull, Flexible Weibull, Power Generalized Weibull, and Marshal–Olkin distributions. Extensive statistical simulations were used to assess the performance of the derived PRH models and compare these to establish a guideline for which distribution/s would "best" fit for left-censored heavy-tailed data. We then applied the best performing model to the South Carolina Enhanced HIV/AIDS Reporting Surveillance System data to explain the effects of

✉ Hrishikesh Chakraborty
  hrishikesh.chakraborty@duke.edu

  Akhtar Hossain
  akhtar.bios@gmail.com

  Farahnaz Islam
  fislam@email.sc.edu

[1] Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC, USA

[2] Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA

different demographic, social, and treatment factors on patients' viral load transition from detectable-to-undetectable levels.

1 Introduction

Human immunodeficiency virus (HIV) is a chronic disease which weakens the immune system, leading to increased susceptibility to a wide range of infections and some types of cancer [46]. An important biomarker to measure HIV disease progression is HIV viral load (VL), the number of copies of actively replicating HIV virus in an individual [42]. By the US Health and Human Services guideline, if the number of copies is less than or equal to 200 per milliliter of blood, VL is classified as undetectable; otherwise, it is classified as detectable [14]. To date, there is no cure for HIV but the suppression of VL to undetectable levels improves physical functioning, reduces opportunistic infections, reduces HIV related mortality, and is associated with a substantial decrease in the probability of transmitting HIV to others [6, 10, 15]. Not only is suppressing VL important on an individual level, it also has the potential to decrease HIV incidence rates in a community because of reduced infectivity [10, 13]. Consequently, the focus of care has shifted from survival to improving health outcomes and the success of highly active antiretroviral therapy (ART) to suppress VL to undetectable levels for prolonged periods of time has transformed HIV into a manageable chronic disease [42].

To gain insight into the HIV endemic, survival models of patient VL may be an effective way since traditional regression models are not able to handle censored data directly. Additionally, these models can be used to assess the effect of various factors and treatments on VL suppression. The commonly used form of survival model can be written as

$$\lambda(t) = \lambda_0(t)e^{x_i\beta}$$

where $\lambda_0(t)$ is the baseline hazard function, x_i is the set of covariates, and β are parameters estimating covariate effects on hazard. In semi-parametric survival models, the regression coefficients are estimated leaving the baseline hazard unspecified. For example, the Cox Proportional Hazards (PH) model [11] introduced the use of the partial-likelihood function to estimate the coefficients without needing to characterize the baseline Hazard Rate. To avoid making distributional assumptions about the baseline hazard, several studies used nonparametric methods to correct for censoring [18, 31, 34, 35]. However, this can also be disadvantageous since assuming an underlying distribution naturally smooths the data so that censoring has less impact on parameter estimates.

While the parametric survival models can be advantageous in many respect, a well-suited parametric distribution for baseline hazard generally ensures more precise

estimation of hazard parameters when compared to the semi-parametric counterpart. However, such benefits also come with a very commonly faced challenge for the applied researchers of selecting the appropriate parametric distribution. This gets even more challenging in cases where the data present characteristics, e.g., left censored time-to-events, heavy-tailed event time density, not very frequently studied in related literature. Given the importance of choosing right distribution in a parametric survival model, any guidance on choosing well-fitting parametric distributions can be a useful addition to the related literature helping applied researchers.

There have been several studies (e.g., [23, 24, 39]) that studied the comparative fits of various distributions to the right-censored and interval-censored lifetime data. However, there are no known recommendations or exploration in the literature on guiding the optimal choice of distribution to use while modeling left-censored time-to-event with heavy-tailed event times. These data features not rare in chronic disease biomarker settings, e.g., HIV VL as discussed in above and characterized further in below (see Fig. 1).

Among the very limited studies in literature analyzing time-to-event of HIV VL suppression, it may be notable that [40] applied a lognormal survival model using a fully parametric approach to take into account the left-censored HIV VL counts. The choice of lognormal distribution was guided by a previous work of [17] based on the estimated lognormal survival distribution function was contained within the 95% confidence interval of nonparametric Kaplan–Meier estimate. Despite [40] presented sensitivity analysis by comparing the lognormal survival model to a univariate mixed model and a Cox PH model, it was not known if any other survival distributions could provide a better fit to the data.

While analyzing left-censored event time data, a further challenge can be the use of the appropriate hazard function for estimation of event risk. The common and widely used estimate of time-to-event risk, Hazard Rate (HR), is appropriate for using with right-censored lifetime data and may be very unstable if used for analyzing left-censored event time risk [43]. A more appropriate choice of estimating left-censored time-to-event risk can be the Reversed Hazard Rate (RHR) [43].

Since its introduction in 1963 [2], the RHR has been used in various applications and several articles [5, 12, 16, 20, 28, 29, 36, 37] studying the properties of the RHR function and devising methodologies based on it to analyze left-censored lifetime data are found in the literature. One recent development is the Parametric Reversed Hazards (PRH) model based on the RHR to be applied to left-censored lifetime data [43]. In this formulation [43], the lifetime random variable was assumed to be distributed as inverted Weibull.

This current study derives the PRH model for a variety of distributions which may be appropriate for left-censored heavy-tailed data including the Exponential, Log-normal, Inverse Gaussian, Log-logistic, Gompertz–Makeham, Gamma, Generalized Gamma, Inverse Gamma, Generalized Inverse Gamma, Weibull, Inverse Weibull, Generalized Inverse Weibull, Modified Weibull, Flexible Weibull, Power Generalized Weibull, and Marshal–Olkin distributions. Extensive statistical simulations are used to assess the performance of the derived PRH models and compare these to establish a guideline for which distribution/s would "best" fit left-censored, heavy-tailed HIV VL data. We applied the selected best performing model to the South Carolina
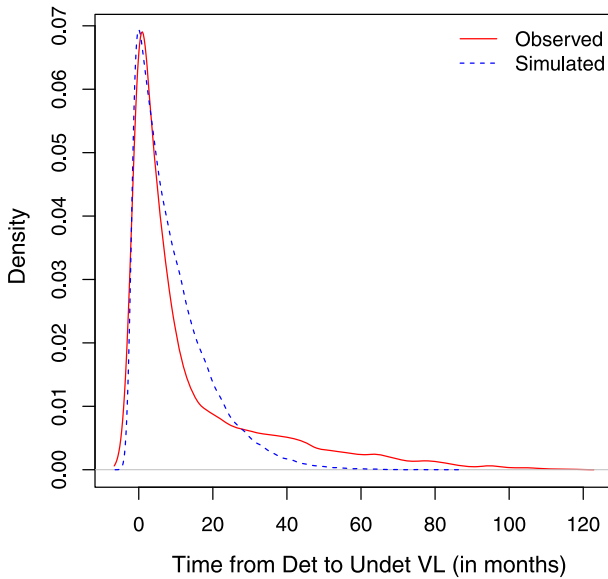
Fig. 1 Density of observed and simulated data

Enhanced HIV/AIDS Reporting Surveillance System (SC eHARS) data to explain effects of different demographic, social, and treatment factors on patients' VL transition from detectable-to-undetectable levels. Recommendations from this study may help researchers apply more accurate models for this type of censoring, specifically in HIV VL-related studies where left censoring may be a common occurrence and the data demonstrate considerably uncommon features, e.g., being heavy-tailed.

2 The Parametric Reverse Hazards model

The Parametric Reversed Hazard (PRH) model [43] is a fully parametric model based on the Reversed Hazard Rate (RHR) for the analysis of left-censored data. The Hazard Rate (HR) used for analyzing more common right-censored time-to-event data is defined as the instantaneous rate of an event in an infinitesimal time width, Δt, following an event free time t and expressed mathematically as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(T \leq t + \Delta t | T \leq t)}{\Delta t}$$

Unlike the above, RHR of T is the instantaneous rate of the event occurring in an infinitesimal time width, Δt, preceding t, given that the event occurred before time t. It is defined as

$$\lambda_r(t) = \lim_{\Delta t \to 0} \frac{P(t - \Delta t \le T | T \le t)}{\Delta t}$$

In terms of the distribution function, $F(t)$, and probability density function, $f(t)$, the RHR function can be written as

$$\lambda_r(t) = \frac{f(t)}{F(t)}$$

By letting X be a $p \times 1$ vector of covariates, we can now define the PRH model as

$$\lambda_r(t|X) = \lambda_{r0}(t)g(\beta; X)$$

where $\lambda_0(t)$ is the baseline RHR, $g(\beta; X)$ is a nonnegative function of X and β (a $p \times 1$ vector of regression parameters), and $\lambda(t|X)$ is the RHR of T given the covariates X.

The PRH model can also be expressed in terms of the distribution function as

$$F(t|X) = F_0(t)^{g(\beta;X)}$$

where $F(t|X)$ is the distribution function of T given X and $F_0(t)$ is the baseline distribution function in the absence of covariates.

Suppose that the lifetime random variable T is randomly left-censored by Z. In practice, we may observe the vectors (Y, δ, X), where $Y = \max(T, Z)$ and $\delta = I(T = Y)$ with $I(.)$ being the indicator function. The likelihood function can then be written as

$$L(\beta, y) = \prod_{i=1}^{n} f(y_i|x_i)^{\delta_i} F(y_i|x_i)^{1-\delta_i}$$

Using this general notation, we show the derivation assuming Generalized Inverse Weibull as the baseline hazard distribution. See supplementary materials for model derivations for the other baseline hazard distributions.

When the lifetime random variable follows a Generalized Inverse Weibull distribution, the baseline distribution function is given by

$$F_0(t) = e^{-\gamma \left(\frac{\lambda}{t}\right)^{\alpha}}, \quad t > 0; \alpha, \gamma, \lambda > 0$$

The baseline RHR of T is then obtained as

$$\lambda_0(t) = \alpha \gamma \lambda^{\alpha} t^{-(\alpha-1)}$$

In the presence of the covariates X, we have

$$\lambda(t|X) = \alpha \gamma \lambda^{\alpha} t^{-(\alpha-1)} e^{x_i \beta}$$

$$F(t|X) = \left[e^{-\gamma \left(\frac{\lambda}{t}\right)^{\alpha}} \right]^{\exp(x_i \beta)}$$

$$f(t|X) = \alpha \gamma \lambda^{\alpha} t^{-(\alpha-1)} e^{x_i \beta} \left[e^{-\gamma \left(\frac{\lambda}{t} \right)^{\alpha}} \right]^{\exp(x_i \beta)}$$

From these, the likelihood and the log-likelihood functions are obtained as

$$L(\alpha, \gamma, \lambda, t) = \prod_{i=1}^{n} \left[\alpha \gamma \lambda^{\alpha} t_i^{-(\alpha-1)} e^{x_i \beta} \right]^{\delta_i} \left[e^{-\gamma \left(\frac{\lambda}{t_i} \right)^{\alpha}} \right]^{\exp(x_i \beta)}$$

$$l(\alpha, \gamma, \lambda, t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i \ln \alpha + \sum_{i=1}^{n} \delta_i \ln \gamma + \sum_{i=1}^{n} \delta_i \alpha \ln \lambda$$

$$- (\alpha - 1) \sum_{i=1}^{n} \delta_i \ln t_i - \sum_{i=1}^{n} \gamma \left(\frac{\lambda}{t_i} \right)^{\alpha} e^{x_i \beta}$$

Similar derivations for several other distributions including the Exponential, Lognormal, Inverse Gaussian, Log-logistic, Gompertz–Makeham, Gamma, Generalized Gamma, Inverse Gamma, Generalized Inverse Gamma, Weibull, Inverse Weibull, Modified Weibull, Flexible Weibull, Power Generalized Weibull, and Marshal–Olkin distributions are provided in supplementary materials.

3 Simulation Study

We used the SC eHARS HIV VL data, further described in the next section, as a real-life example of such data and simulate data with similar distribution for the time to transit from detectable VL to undetectable VL state after HIV diagnosis. Figure 1 presents the density of time to transition from detectable-to-undetectable VL transition for both observed data and a randomly selected set of simulated data showing similarities in densities. The time to detectable-to-undetectable VL transition data were simulated from a Skewed Normal distribution with location, scale, and shape parameters, respectively, as 5, 30, 50. Different parameters were tested under the Skewed Normal distribution using a trial-and-error approach until the simulated data matched as close as possible to the SC VL data.

To assess the model fits best, we used information criteria including

1. Akaike Information Criterion (AIC) rewards goodness of fit but penalizes the model for increasing the number of estimated parameters:

$$AIC = 2k - 2 \ln(L)$$

2. Bayesian Information Criterion (BIC), which uses a larger penalty than AIC:

$$BIC = k \ln(n) - 2 \ln(L)$$

3. Corrected Akaike Information Criterion ($AICC$), which corrects the AIC for overfitting of the data in cases where the sample size is relatively small compared to

the number of parameters in the model:

$$AICC = AIC + (2k(k + 1))/(n - k - 1)$$

4. Hannan–Quinn Information Criterion ($HQIC$), which is often cited in the literature but, unlike AIC, it is not asymptotically efficient:

$$HQIC = 2k \ln(\ln(n)) - 2\ln(L)$$

5. Bozdogan's Consistent Akaike Information Criterion ($CAIC$), is another adjusted form of AIC which is consistent:

$$CAIC = k(\ln(n) + 1) - 2\ln(L)$$

Where k is the number of parameters to be estimated, L is the maximum value of the likelihood function, and n is the number of observations. The model with the smallest average AIC, BIC, $AICC$, $HQIC$, and $CAIC$ value was determined to be the model with the best fit. The simulation studies were conducted using the Statistical Computing Software, R version 3.2.5. The summaries of the simulation results are presented in Tables 1, 2, 3.

Table 1 summarizes the results for the simulated data with a censoring rate of 20%, Table 2 for data with censoring rate of 30%, and Table 3 for data with censoring rate 40%. From these tables, it is clear that the Generalized Inverse Weibull distribution consistently performs the best, having the lowest average AIC, BIC, $AICC$, $HQIC$, and $CAIC$. Following closely behind in performance are the Log-Logistic, Log-Normal, Inverse Gaussian, and Gamma distributions, respectively. This is consistent across all censoring rates and sample sizes. The consistently worst performing distributions are the Modified Weibull, Inverse Weibull, Inverse Gamma, Power Generalized Weibull, and Exponential distributions, respectively.

4 Application to SC eHARS Data

The HIV endemic disproportionately impacts the Southern states in the US in terms of the overall number of people living with HIV/AIDS (PLWHA), and survival rates after diagnosis [33]. SC, like many Southern states, ranks high for poverty, unemployment, and low educational completion which are all characteristics that may promote disease transmission. The number of PLWHA in SC has increased from 12,089 in 2004 to 16,311 in 2014 [38]. Studies on retention in HIV care found that a large proportion of PLWHA in SC failed to remain in care on a regular basis [30, 41]. Given the HIV burden in SC and the need to focus on retention in HIV care within the context of the National HIV/AIDS Strategy goals, it is important to identify factors which suppress VL. Identifying these factors will assist in developing targeted strategies to reduce the HIV burden in SC.

Since January 2004, all health care providers, hospitals, and laboratories in SC are legally mandated to report all CD4 count and VL measurements to the SC Depart-

Table 1 Average summary measures across 5000 simulations from simulation study with censoring rate 20%

Distribution	Sample Size = 1000				
	AIC	*BIC*	*AICC*	*HQIC*	*CAIC*
Exponential	8753.05	8757.96	8753.06	8754.92	8758.96
Log-Normal	6162.11	6167.02	6162.12	6165.84	6173.92
Inverse Gaussian	6163.50	6168.41	6163.51	6167.23	6175.31
Gamma	6199.30	6204.21	6199.31	6203.03	6211.12
Generalized Gamma	6206.54	6209.45	6206.57	6212.14	6224.27
Inverse Gamma	10622.02	10626.93	10622.03	10625.75	10633.83
Log-Logistic	6115.87	6120.78	6115.88	6119.60	6127.68
Weibull	6502.31	6507.22	6502.32	6506.04	6514.13
Inverse Weibull	13992.74	13997.65	13992.75	13996.47	14004.55
Generalized Inverse Weibull	**5959.85**	**5964.76**	**5959.87**	**5965.44**	**5977.57**
Flexible Weibull	7768.04	7772.95	7768.05	7771.77	7779.86
Marshal–Olkin	7532.91	7537.82	7532.94	7538.51	7550.63
Power Generalized Weibull	10492.37	10497.28	10492.38	10496.10	10504.19
Modified Weibull	2.97e63	2.97e63	2.97e63	2.97e63	2.97e63
Gompertz	8177.79	8182.70	8177.80	8181.52	8189.60
Distribution	Sample Size = 2000				
	AIC	*BIC*	*AICC*	*HQIC*	*CAIC*
Exponential	17504.19	17508.97	17504.20	17506.25	17510.79
Log-Normal	12323.86	12329.46	12323.87	12327.97	12337.06
Inverse Gaussian	12326.67	12332.27	12326.68	12330.79	12339.88
Gamma	12398.14	12403.74	12398.14	12402.25	12411.34
Generalized Gamma	12410.79	12416.39	12410.80	12416.96	12430.60
Inverse Gamma	21240.12	21245.72	21240.13	21244.24	21253.32
Log-Logistic	12230.91	12236.51	12230.92	12235.03	12244.12
Weibull	13007.93	13013.53	13007.94	13012.05	13021.14
Inverse Weibull	27981.55	27987.15	27981.56	27985.66	27994.75
Generalized Inverse Weibull	**11916.55**	**11922.15**	**11916.56**	**11922.72**	**11936.35**
Flexible Weibull	15488.38	15493.98	15488.39	15492.50	15501.59
Marshal–Olkin	15059.88	15065.48	15059.90	15066.05	15079.69
Power Generalized Weibull	20980.81	20986.41	20980.82	20984.92	20994.01
Modified Weibull	1.27e68	1.27e68	1.27e68	1.27e68	1.27e68
Gompertz	16177.93	16183.53	16177.94	16182.05	16191.13

Table 1 continued

Distribution	Sample Size = 3000				
	AIC	*BIC*	*AICC*	*HQIC*	*CAIC*
Exponential	26255.36	26261.37	26255.36	26257.52	26262.37
Log-Normal	18485.63	18491.64	18485.64	18489.96	18499.65
Inverse Gaussian	18489.88	18495.89	18489.89	18494.20	18503.90
Gamma	18597.16	18603.17	18597.16	18601.48	18611.17
Generalized Gamma	18615.33	18621.34	18615.34	18621.82	18636.35
Inverse Gamma	31858.19	31864.20	31858.19	31862.51	31872.20
Log-Logistic	18346.09	18352.10	18346.09	18350.41	18360.10
Weibull	19515.02	19521.03	19515.03	19519.34	19529.04
Inverse Weibull	41970.35	41976.36	41970.35	41974.67	41984.36
Generalized Inverse Weibull	**17873.49**	**17879.50**	**17873.50**	**17879.97**	**17894.51**
Flexible Weibull	23311.80	23317.87	23311.80	23316.12	23325.81
Marshal–Olkin	22586.86	22592.87	22586.87	22593.34	22607.88
Power Generalized Weibull	31469.28	31475.29	31469.28	31473.60	31483.29
Modified Weibull	1.00e68	1.00e68	1.00e68	1.00e68	1.00e68
Gompertz	24099.00	24105.01	24099.01	24103.32	24113.01

Table 2 Average summary measures across 5000 simulations from simulation study with censoring rate 30%

Distribution	Sample Size = 1000				
	AIC	*BIC*	*AICC*	*HQIC*	*CAIC*
Exponential	7752.73	7757.64	7752.74	7754.60	7758.64
Log-Normal	5571.09	5576.00	5571.10	5574.82	5582.91
Inverse Gaussian	5572.56	5577.47	5572.57	5576.29	5584.38
Gamma	5603.77	5608.68	5603.78	5607.50	5615.58
Generalized Gamma	5611.47	5616.38	5611.50	5617.07	5629.20
Inverse Gamma	9731.22	9736.13	9731.23	9734.95	9743.04
Log-Logistic	5521.91	5526.82	5521.92	5525.64	5533.73
Weibull	5848.74	5853.65	5848.75	5852.47	5860.55
Inverse Weibull	12430.99	12435.90	12431.01	12434.72	12442.81
Generalized Inverse Weibull	**5401.80**	**5406.71**	**5401.83**	**5407.40**	**5419.53**
Flexible Weibull	6843.80	6848.71	6843.81	6847.53	6855.61
Marshal–Olkin	7118.74	7123.65	7118.77	7124.34	7136.47
Power Generalized Weibull	9251.89	9256.80	9251.90	9255.62	9263.70
Modified Weibull	6.15e63	6.15e63	6.15e63	6.15e63	6.15e63
Gompertz	7117.94	7122.85	7117.95	7121.67	7129.76

Table 2 continued

Distribution	Sample Size = 2000				
	AIC	BIC	AICC	HQIC	CAIC
Exponential	15503.37	15508.97	15503.37	15505.43	15509.97
Log-Normal	11141.41	11147.01	11141.41	11145.52	11154.61
Inverse Gaussian	11144.38	11149.98	11144.39	11148.49	11157.58
Gamma	11206.68	11212.28	11206.68	11210.79	11219.88
Generalized Gamma	11219.82	11225.42	11219.84	11225.99	11239.63
Inverse Gamma	19458.34	19463.94	19458.34	19462.45	19471.54
Log-Logistic	11042.28	11047.88	11042.28	11046.39	11055.48
Weibull	11700.72	11706.32	11700.73	11704.83	11713.92
Inverse Weibull	24857.86	24863.46	24857.87	24861.98	24871.07
Generalized Inverse Weibull	**10799.21**	**10804.81**	**10799.22**	**10805.38**	**10819.01**
Flexible Weibull	13628.37	13633.97	13628.37	13632.48	13641.57
Marshal–Olkin	14231.37	14236.97	14231.38	14237.54	14251.17
Power Generalized Weibull	18499.64	18505.24	18499.64	18503.75	18512.84
Modified Weibull	1.95e63	1.95e63	1.95e63	1.95e63	1.95e63
Gompertz	14424.68	14430.28	14424.68	14428.79	14437.88
Distribution	Sample Size = 3000				
	AIC	BIC	AICC	HQIC	CAIC
Exponential	23254.14	23260.15	23254.14	23256.30	23261.15
Log-Normal	16711.43	16717.44	16711.43	16715.75	16725.44
Inverse Gaussian	16715.90	16721.91	16715.90	16720.22	16729.91
Gamma	16809.04	16815.05	16809.04	16813.36	16823.05
Generalized Gamma	16828.00	16834.01	16828.01	16834.48	16849.02
Inverse Gamma	29185.65	29191.66	29185.66	29189.97	29199.67
Log-Logistic	16562.88	16568.89	16562.88	16567.20	16576.89
Weibull	17551.99	17557.99	17552.00	17556.31	17566.01
Inverse Weibull	37284.93	37290.94	37284.94	37289.25	37298.95
Generalized Inverse Weibull	**16197.20**	**16203.21**	**16197.20**	**16203.68**	**16218.22**
Flexible Weibull	20402.93	20408.94	20402.93	20407.25	20416.94
Marshal–Olkin	21344.14	21350.15	21344.15	21350.62	21365.16
Power Generalized Weibull	27747.57	27753.58	27747.58	27751.89	27761.59
Modified Weibull	5.20e64	5.20e64	5.20e64	5.20e64	5.20e64
Gompertz	21422.25	21428.26	21422.25	21426.57	21436.26

ment of Health and Environmental Control (DHEC) [7]. These data are stored in the SC eHARS database along with the patient's socio-demographic characteristics. The quality rating of the SC eHARS database exceeds the CDC minimum standards of reporting timeliness with 95% of new cases being reported within 6 months of HIV diagnosis and 98% of all HIV cases reported [44]. Our sample consisted of 6,221

Table 3 Average summary measures across 5000 simulations from simulation study with censoring rate 40%

Distribution	Sample Size = 1000				
	AIC	BIC	AICC	HQIC	CAIC
Exponential	6741.30	6746.21	6741.31	6743.17	6747.21
Log-Normal	4953.90	4958.81	4953.91	4957.63	4965.71
Inverse Gaussian	4955.39	4960.30	4955.40	4959.12	4967.20
Gamma	4982.80	4987.71	4982.82	4986.53	4994.62
Generalized Gamma	4986.95	4991.86	4986.98	4992.55	5004.68
Inverse Gamma	8650.96	8655.97	8650.97	8654.69	8662.77
Log-Logistic	4905.78	4910.69	4905.79	4909.51	4917.59
Weibull	5174.16	5179.07	5174.17	5177.89	5185.97
Inverse Weibull	10840.67	10845.58	10840.68	10844.40	10852.48
Generalized Inverse Weibull	**4815.45**	**4820.36**	**4815.48**	**4821.05**	**4833.18**
Flexible Weibull	5926.41	5931.32	5926.42	5930.14	5938.23
Marshal–Olkin	6704.64	6709.55	6704.66	6710.23	6722.36
Power Generalized Weibull	8003.90	8008.81	8003.91	8007.63	8015.72
Modified Weibull	4.60e63	4.60e63	4.60e63	4.60e63	4.60e63
Gompertz	6218.98	6223.89	6218.99	6222.71	6230.79
Distribution	Sample Size = 2000				
	AIC	BIC	AICC	HQIC	CAIC
Exponential	13480.48	13486.08	13480.48	13482.54	13487.08
Log-Normal	9905.84	9911.44	9905.85	9909.95	9919.04
Inverse Gaussian	9908.88	9914.48	9908.88	9912.99	9922.08
Gamma	9964.45	9970.05	9964.46	9968.56	9977.65
Generalized Gamma	9972.46	9978.06	9972.47	9978.63	9992.26
Inverse Gamma	17297.85	17303.45	17297.85	17301.96	17311.05
Log-Logistic	9809.98	9815.58	9809.99	9814.09	9823.18
Weibull	10349.18	10354.78	10349.19	10353.29	10362.38
Inverse Weibull	21677.25	21682.85	21677.26	21681.37	21690.46
Generalized Inverse Weibull	**9626.12**	**9631.72**	**9626.13**	**9632.29**	**9645.92**
Flexible Weibull	11890.95	11896.55	11890.96	11895.07	11904.15
Marshal–Olkin	13403.17	13408.77	13403.18	13409.34	13422.97
Power Generalized Weibull	16003.69	16009.29	16003.69	16007.80	16016.89
Modified Weibull	2.09e62	2.09e62	2.09e62	2.09e62	2.09e62
Gompertz	12435.16	12440.76	12435.17	12439.27	12448.36

Table 3 continued

Distribution	Sample Size = 3000				
	AIC	*BIC*	*AICC*	*HQIC*	*CAIC*
Exponential	20220.29	20226.30	20220.29	20222.45	20227.30
Log-Normal	14861.39	14867.40	14861.39	14865.71	14875.40
Inverse Gaussian	14865.96	14871.97	14865.97	14870.29	14879.98
Gamma	14948.26	14954.27	14948.26	14952.58	14962.27
Generalized Gamma	14961.84	14967.85	14961.85	14968.33	14982.86
Inverse Gamma	25945.08	25951.09	25945.09	25949.41	25959.10
Log-Logistic	14716.61	14722.62	14716.62	14720.93	14730.62
Weibull	15527.55	15533.56	15527.55	15531.87	15541.56
Inverse Weibull	32514.24	32520.25	32514.24	32518.56	32528.25
Generalized Inverse Weibull	**14440.87**	**14446.88**	**14440.88**	**14447.35**	**14461.89**
Flexible Weibull	17877.73	17883.74	17877.73	17882.05	17891.74
Marshal–Olkin	20102.14	20108.15	20102.15	20108.62	20123.16
Power Generalized Weibull	24003.98	24009.99	24003.98	24008.31	24018.00
Modified Weibull	9.34e65	9.34e65	9.34e65	9.34e65	9.34e65
Gompertz	18726.64	18732.65	18726.65	18730.97	18740.66

residents in SC who were aged $\geq$ 13 years or older; diagnosed or living with HIV infection between January 1, 2005, and December 31, 2013; had detectable VL at the start of the study period; had at least two reported VL values during the study period.

This study applies the best model as determined from the simulation study to left-censored heavy-tailed HIV VL data from South Carolina. The aim of applying the PRH model to this dataset is to explain the risk behavior of transitioning from detectable VL to undetectable VL. Patients with undetectable VL at the beginning of the study were defined as being left-censored. Covariates that were assessed include gender (male or female), race (White, Black, or other), HIV risk exposure group (heterosexual, men who have sex with men, or other), place of residence (rural or urban), age at baseline, initial treatment regimen (single tablet regimen, multiple tablet regimen), and baseline CD4 count (200 or less, 201 to 350, 351 to 500, or more than 500). Note that HIV risk exposure group refers to how the patient was first exposed to HIV with options including heterosexual HIV infected partner, men who have sex with other men, injecting drug user, no identifiable risk, and no risk reported. Results from the PRH model are presented and discussed in the next section.

Of the individuals in our sample, 1703 (27%) had an undetectable VL at the beginning of the observation period, so they were considered as being left-censored (Table 4). Mean age of the sample at baseline was 40.0 years (range = 14.8–81.6). The majority of subjects were male ($n = 3657$, 58.8%), Black ($n = 4966$, 79.8%), and lived in an urban county when diagnosed with HIV ($n = 4208$, 67.6%). The CD4 count at the beginning of the study was less than 200 cells/mm^3 for just over one third of the individuals (34.03%). Almost half of the sample had missing treatment regimen ($n = 2928$, 47.1%).

Table 4 Characteristics of persons living with HIV in South Carolina, 2005-2013

Characteristics	Frequency (%)/Summary statistics $n = 6,221$
Gender	
Female	2564 (41.22%)
Male	3657 (58.78%)
Race	
Black	4966 (79.83%)
White	1086 (17.46%)
Others	169 (2.72%)
Risk of Exposure	
Heterosexual	2295 (36.89%)
Men who have Sex with Men (MSM)	1911 (30.72%)
Others	2015 (32.39%)
Place of Residence	
Urban	4208 (67.64%)
Rural	2013 (32.36%)
Starting Treatment Regimen	
Single Tablet Regimen (STR)	1056 (16.97%)
Multiple Tablet Regimen (MTR)	2237 (35.96%)
Missing	2928 (47.07%)
Baseline Age (in years)	Range = 14.84-81.58; Mean = 39.99; SD = 11.46
CD4 Count (cell/mm^3)	
200 or less	2136 (34.34%)
201 to 350	1253 (20.14%)
351 to 500	1117 (17.96%)
More than 500	1715 (27.57%)
Outcome	
Event (Det VL to Undet VL)	4518 (72.62%)
Left-censored	1703 (27.38%)

SD standard deviation, *Det* detectable, *VL* viral load, *Undet* undetectable

The Generalized Inverse Weibull distribution, which was found to be the best performing distribution from the simulation study, is applied to analyze the left-censored SC eHARS data time-to-event data for detectable-to-undetectable VL transition. Table 5 shows the results of the estimated PRH model using a Generalized Inverse Weibull distribution. Information on treatment regimen is a very important variable to use in our model to assess which type of treatment has the most, if any, impact on the transition from detectable-to-undetectable VL. However, this information is missing in almost 50% of the subjects in our sample. Thus, we fit the model without this starting treatment regimen information (Model 1) and then we fit a second model with reduced sample size after including the treatment variable in the model (Model 2). It should be

Table 5 Estimated Reverse Hazard Rates (HR) using Generalized Inverse Weibull Reverse Hazard model (without interactions) of SC adult HIV patients

Characteristics	Model 1 $n = 6{,}221$ Reverse HR (95% CI)	Model 2 $n = 3{,}293$ Reverse HR (95% CI)
Treatment Regimen		
Single Tablet Regimen (STR)	–	Ref
Multiple Tablet Regimen (MTR)	–	1.32 (1.22, 1.42)
Gender		
Female	Ref	Ref
Male	1.27 (1.19, 1.37)	5.23 (4.78, 5.72)
Race		
Black	Ref	Ref
White	1.21 (1.14, 1.29)	1.18 (1.10, 1.26)
Others	2.05 (1.83, 2.30)	1.02 (0.85, 1.24)
Risk of Exposure		
Heterosexual	Ref	Ref
Men who have Sex with Men (MSM)	1.42 (1.30, 1.55)	0.73 (0.67, 0.79)
Others	1.46 (1.36, 1.57)	0.57 (0.53, 0.62)
Place of Residence		
Rural	Ref	Ref
Urban	2.67 (2.48, 2.87)	2.01 (1.88, 2.15)
Baseline Age (in years)	0.98 (0.98, 0.99)	0.95 (0.95, 0.96)
CD4 Count (cell/mm^3)		
200 or less	Ref	Ref
201 to 350	0.97 (0.90, 1.06)	0.75 (0.70, 0.82)
351 to 500	1.22 (1.13, 1.32)	0.87 (0.81, 0.94)
More than 500	0.93 (0.87, 0.99)	0.38 (0.35, 0.41)

noted that if there was not such a large proportion of missing values in the treatment variable, we would fit only one model, Model 2.

While several covariates have been shown to have an effect on the time-to-event of transitioning from detectable-to-undetectable VL level, the significant change in behavior of some of these covariates comparing the model incorporating the treatment variable compared to the model without this important factor suggests that an interaction may be present between treatment regimen and each of the other covariates. Additional models were run testing for these interactions. The only statistically significant interaction found was between treatment regimen and age, the results of which are shown in Table 6.

The final model is shown in Table 6. Males are 1.11 times more likely to reach undetectable levels faster than their female counterparts (95% CI 1.00, 1.24). White individuals are 1.53 times more likely to reach undetectable levels faster than Black individuals (95% CI 1.40, 1.67). Other races are 0.86 times less likely to reach unde-

Table 6 Estimated Reverse Hazard Rates using Generalized Inverse Weibull Reverse Hazard model (with interactions) of SC adult HIV patients

Characteristics	Reverse HR (95% CI)
Treatment Regimen	
Single Tablet Regimen (STR)	Ref
Multiple Tablet Regimen (MTR)	0.86 (0.59, 1.24)
Gender	
Female	Ref
Male	1.11 (1.00, 1.24)
Race	
Black	Ref
White	1.53 (1.40, 1.67)
Others	0.86 (0.63, 1.16)
Risk of Exposure	
Heterosexual	Ref
Men who have Sex with Men (MSM)	0.99 (0.87, 1.12)
Others	1.00 (0.90, 1.11)
Place of Residence	
Rural	Ref
Urban	0.94 (0.86, 1.02)
Baseline Age (in years)	0.94 (0.93, 0.95)
CD4 Count (cell/mm^3)	
200 or less	Ref
201 to 350	0.92 (0.79, 1.06)
351 to 500	0.98 (0.85, 1.12)
More than 500	0.98 (0.86, 1.11)
Interaction of treatment by age	1.01 (1.00, 1.02)

tectable levels faster than Black individuals, though this finding is not significant (95% CI 0.63, 1.16). Risk of exposure, place of residence (rural vs urban), and CD4 count do not seem to have any statistically significant impact on the time taken to transition from detectable-to-undetectable VL levels. The significant interaction between treatment regimen and age highlights that older people living with HIV/AIDS are 0.97 times less likely to reach undetectable levels faster than their younger counterparts (95% CI 0.97, 0.98).

5 Discussion

The current study derived several extensions of the PRH model and conducted extensive simulation studies to evaluate the usefulness of parametric regression models based on the Reversed Hazard Rate for analyzing left-censored heavy-tailed HIV viral load time-to-event data. Simulation studies suggested the best distribution to use under

the PRH model is the Generalized Inverse Weibull distribution followed in order of performance by Log-Logistic, Log-Normal, Inverse Gaussian, and Gamma distributions.

Application of this best performing model on the SC eHARS data revealed important factors on the time to transition from detectable-to-undetectable viral load levels. Males were found to be more likely to reach undetectable levels faster than females. This trend is also evident in several recent studies [4, 7, 25]. A possible reason for this disparity may be attributed to the higher rates of treatment adherence among males compared to females. Though some studies did not find an association between gender and treatment adherence, a meta-analysis [22] of 207 studies concluded that males adhere more to ART than females.

White individuals are more likely to reach undetectable levels faster than Black individuals. This is supported by several studies which show that Black individuals are disproportionately affected by HIV/AIDS as they tend to have poorer access to health care, are less likely to receive treatment, less likely to adhere to treatment, and less likely to survive HIV/AIDS [4, 7–9, 19, 27, 32].

This study did not find any statistically significant association between place of residence and time to transition from detectable-to-undetectable VL levels. This may seem in contrary to the expectation that individuals who live in urban areas would be more likely to reach undetectable levels faster than those who live in rural areas due to the typically increased access to health care and higher range of specialists available to people living with HIV/AIDS in urban areas [44, 45]. However, other studies (e.g.,[9]) have also reported analyses supporting the current study reporting no significant effect of place of residence on detectable-to-undetectable VL transition.

Finally, the interaction between drug regimen and age highlights that older people who are on a multiple treatment regimen are likely to reach undetectable levels slower than their younger counterparts. There are mixed findings on this in the existing literature. Young people with HIV tend to have delayed diagnosis and thus higher VL at baseline. One study [7] suggests that this along with underutilization of health care due to HIV-related stigma explains their finding that younger people with HIV reach undetectable levels slower than their older counterparts. A possible explanation of our result may be that older people are not as adherent to treatment [22] or perhaps they have a co-existing morbidity which effects the rate at which they reach undetectable levels.

There are several limitations of the SC eHARS database. Data on VL and CD4 count measurements were not available for those who dropped out of medical care after initial diagnosis—this includes those who passed away, moved to a different state, etc. Additionally, persons living with HIV/AIDS who have not been diagnosed were not captured in this database. The database also does not include information on morbidities which may be co-existing with HIV/AIDS which can impact the effect of drug regimens, especially in older people. Since the interaction between age and drug regimen is found to be statistically significant to have an impact on the VL transition, co-existing conditions warrant further exploration. These limitations may have resulted in not finding an association with factors we would expect based on prior research.

Regardless of these limitations, the application to the SC eHARS database provides important information on the trajectories of VL in SC over time. The results obtained in this study can be used to direct researchers in applying more accurate models when studying similar databases. We recommend that the Generalized Inverse Weibull PRH model be used for analyses involving skewed, left-censored heavy-tailed HIV VL data.

Appendix

Exponential Distribution

When the lifetime random variable follows an Exponential distribution, the baseline distribution function is given by

$$F_0(t) = 1 - e^{-t/\gamma}, \quad t > 0; \gamma > 0$$

The baseline Reversed Hazard Rate of T is then obtained as

$$\lambda_0(t) = \frac{e^{-t/\gamma}}{\gamma(1 - e^{-t/\gamma})}$$

In the presence of the covariates X, we have the following

$$\lambda(t|X) = \frac{e^{-t/\gamma}}{\gamma(1 - e^{-t/\gamma})} \exp(x_i\beta)$$

$$F(t|X) = (1 - e^{-t/\gamma})^{\exp(x_i\beta)}$$

$$f(t|X) = \frac{e^{-t/\gamma}}{\gamma} \exp(x_i\beta)(1 - e^{-t/\gamma})^{\exp(x_i\beta)-1}$$

From these, the likelihood function for the Exponential distribution is obtained as

$$L(\beta, \gamma, t) = \prod_{i=1}^{n} \left[\frac{e^{-t_i/\gamma}}{\gamma} \exp(x_i\beta)(1 - e^{-t_i/\gamma})^{\exp(x_i\beta)-1} \right]^{\delta_i} \left[(1 - e^{-t_i/\gamma})^{\exp(x_i\beta)} \right]^{1-\delta_i}$$

so that the log likelihood function is

$$l(\beta, \gamma, t) = \sum_{i=1}^{n} \delta_i x_i \beta - \frac{1}{\gamma} \sum_{i=1}^{n} \delta_i t_i - \sum_{i=1}^{n} \delta_i \ln \gamma + \sum_{i=1}^{n} (e^{x_i \beta} - \delta_i) \ln(1 - e^{-t_i/\gamma})$$

Log-normal Distribution

When the lifetime random variable follows a Log-normal distribution, the baseline distribution function is given by

$$F_0(t) = \Phi\left(\frac{\ln(t) - \mu}{\sigma}\right), \quad t > 0; \mu, \sigma > 0$$

The baseline Reversed Hazard Rate of T is then obtained as

$$\lambda_0(t) = \frac{\frac{1}{\sqrt{2\pi}\sigma t} \exp\left(-\frac{[\ln(t) - \mu]^2}{2\sigma^2}\right)}{\Phi\left(\frac{\ln(t) - \mu}{\sigma}\right)}$$

In the presence of the covariates X, we have the following

$$\lambda(t|X) = \frac{\exp\left(-\frac{[\ln(t) - \mu]^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma t \,\Phi\left(\frac{\ln(t) - \mu}{\sigma}\right)} \exp(x_i \beta)$$

$$F(t|X) = \left[\Phi\left(\frac{\ln(t) - \mu}{\sigma}\right)\right]^{\exp(x_i \beta)}$$

$$f(t|X) = \frac{1}{\sqrt{2\pi}\sigma t} \exp\left(-\frac{[\ln(t) - \mu]^2}{2\sigma^2} + x_i \beta\right) \left[\Phi\left(\frac{\ln(t) - \mu}{\sigma}\right)\right]^{\exp(x_i \beta) - 1}$$

From these, the likelihood function for the Log-normal distribution is obtained as

$$L(\mu, \sigma, t) = \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}\sigma t} \exp\left(-\frac{[\ln(t) - \mu]^2}{2\sigma^2} + x_i \beta\right) \left[\Phi\left(\frac{\ln(t) - \mu}{\sigma}\right)\right]^{\exp(x_i \beta) - 1}\right]^{\delta_i}$$

$$\times \left[\Phi\left(\frac{\ln(t) - \mu}{\sigma}\right)\right]^{(1-\delta_i)\exp(x_i \beta)}$$

so that the log likelihood function is

$$l(\mu, \sigma, t) = \sum_{i=1}^{n} \delta_i x_i \beta - \sum_{i=1}^{n} \delta_i \ln(\sqrt{2\pi}\sigma t_i) + \sum_{i=1}^{n} \delta_i \frac{[\ln(t) - \mu]^2}{2\sigma^2}$$

$$+ \sum_{i=1}^{n} (e^{x_i \beta} - \delta_i) \ln\left[\Phi\left(\frac{\ln(t) - \mu}{\sigma}\right)\right]$$

Inverse Gaussian Distribution

When the lifetime random variable follows a Inverse Gaussian distribution, the baseline distribution function is given by

$$F_0(t) = \Phi\left[-\sqrt{\frac{\gamma}{t}}\left(\frac{t}{\alpha}-1\right)\right] - exp\left(\frac{2\gamma}{\alpha}\right)\Phi\left[-\sqrt{\frac{\gamma}{t}}\left(\frac{t}{\alpha}+1\right)\right], \quad t>0; \alpha, \gamma>0$$

The baseline Reversed Hazard Rate of T is then obtained as

$$\lambda_0(t) = \frac{\sqrt{\frac{\gamma}{2\pi t^3}}\exp\left[\frac{-\gamma(t-\alpha)^2}{2\alpha^2 t}\right]}{\Phi\left[-\sqrt{\frac{\gamma}{t}}\left(\frac{t}{\alpha}-1\right)\right] - exp\left(\frac{2\gamma}{\alpha}\right)\Phi\left[-\sqrt{\frac{\gamma}{t}}\left(\frac{t}{\alpha}+1\right)\right]}$$

In the presence of the covariates X, we have the following

$$\lambda(t|X) = \frac{\sqrt{\frac{\gamma}{2\pi t^3}}\exp\left[\frac{-\gamma(t-\alpha)^2}{2\alpha^2 t}+x_i\beta\right]}{\Phi\left[-\sqrt{\frac{\gamma}{t}}\left(\frac{t}{\alpha}-1\right)\right] - exp\left(\frac{2\gamma}{\alpha}\right)\Phi\left[-\sqrt{\frac{\gamma}{t}}\left(\frac{t}{\alpha}+1\right)\right]}$$

$$F(t|X) = \left\{\Phi\left[-\sqrt{\frac{\gamma}{t}}\left(\frac{t}{\alpha}-1\right)\right] - \exp\left(\frac{2\gamma}{\alpha}\right)\Phi\left[-\sqrt{\frac{\gamma}{t}}\left(\frac{t}{\alpha}+1\right)\right]\right\}^{\exp(x_i\beta)}$$

$$f(t|X) = \sqrt{\frac{\gamma}{2\pi t^3}}\exp\left[\frac{-\gamma(t-\alpha)^2}{2\alpha^2 t}+x_i\beta\right]$$
$$\times\left\{\Phi\left[-\sqrt{\frac{\gamma}{t}}\left(\frac{t}{\alpha}-1\right)\right] - \exp\left(\frac{2\gamma}{\alpha}\right)\Phi\left[-\sqrt{\frac{\gamma}{t}}\left(\frac{t}{\alpha}+1\right)\right]\right\}^{\exp(x_i\beta)-1}$$

From these, the likelihood function for the Inverse Gaussian distribution is obtained as

$$L(\alpha,\gamma,t) = \prod_{i=1}^{n}\left\{\sqrt{\frac{\gamma}{2\pi t_i^3}}\exp\left[\frac{-\gamma(t_i-\alpha)^2}{2\alpha^2 t_i}+x_i\beta\right]\right\}^{\delta_i}$$
$$\times\left\{\Phi\left[-\sqrt{\frac{\gamma}{t_i}}\left(\frac{t_i}{\alpha}-1\right)\right] - \exp\left(\frac{2\gamma}{\alpha}\right)\Phi\left[-\sqrt{\frac{\gamma}{t_i}}\left(\frac{t_i}{\alpha}+1\right)\right]\right\}^{\exp(x_i\beta)-\delta_i}$$

so that the log likelihood function is

$$l(\alpha,\gamma,t) = \sum_{i=1}^{n}\delta_i x_i\beta + \frac{1}{2}\sum_{i=1}^{n}\delta_i\ln\left(\frac{\gamma}{2\pi t_i^3}\right) - \sum_{i=1}^{n}\left[\frac{\delta_i\gamma(t_i-\alpha)^2}{2\alpha^2 t_i}\right]$$
$$+ \sum_{i=1}^{n}\left[e^{x_i\beta}-\delta_i\right]\ln\left\{\Phi\left[-\sqrt{\frac{\gamma}{t_i}}\left(\frac{t_i}{\alpha}+1\right)\right]\right\} + \sum_{i=1}^{n}\left[e^{x_i\beta}-\delta_i\right]\ln\left[1-\exp\left(\frac{2\gamma}{\alpha}\right)\right]$$

Log-logistic Distribution

When the lifetime random variable follows a Log-logistic distribution, the baseline distribution function is given by

$$F_0(t) = \frac{1}{1 + \left(\frac{t}{\alpha}\right)^{-\omega}}, \quad t > 0; \alpha, \omega > 0$$

The baseline Reversed Hazard Rate of T is then obtained as

$$\lambda_0(t) = \frac{\frac{\omega}{t}}{1 + \left(\frac{t}{\alpha}\right)^{\omega}}$$

In the presence of the covariates X, we have the following

$$\lambda(t|X) = \frac{\frac{\omega}{t} \exp(x_i \beta)}{1 + \left(\frac{t}{\alpha}\right)^{\omega}}$$

$$F(t|X) = \left[\frac{1}{1 + \left(\frac{t}{\alpha}\right)^{-\omega}}\right]^{\exp(x_i \beta)}$$

$$f(t|X) = \left[\frac{\frac{\omega}{t} \exp(x_i \beta)}{1 + \left(\frac{t}{\alpha}\right)^{\omega}}\right]\left[\frac{1}{1 + \left(\frac{t}{\alpha}\right)^{-\omega}}\right]^{\exp(x_i \beta)}$$

From these, the likelihood function for the Log-logistic distribution is obtained as

$$L(\alpha, \omega, t) = \prod_{i=1}^{n} \left[\frac{\frac{\omega}{t_i} \exp(x_i \beta)}{1 + \left(\frac{t_i}{\alpha}\right)^{\omega}}\right]^{\delta_i}\left[\frac{1}{1 + \left(\frac{t_i}{\alpha}\right)^{-\omega}}\right]^{\exp(x_i \beta)}$$

so that the log likelihood function is

$$l(\alpha, \omega, t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i \ln\left(\frac{\omega}{t_i}\right) - \sum_{i=1}^{n} \delta_i \ln\left[1 + \left(\frac{t_i}{\alpha}\right)^{\omega}\right] + \sum_{i=1}^{n} e^{x_i \beta} \ln\left[1 + \left(\frac{t_i}{\alpha}\right)^{-\omega}\right]$$

Gompertz–Makeham Distribution

When the lifetime random variable follows a Gompertz–Makeham distribution, the baseline distribution function is given by

$$F_0(t) = 1 - \exp\left[-\frac{\alpha}{\gamma}\left(e^{\gamma t} - 1\right)\right], \quad t > 0; \alpha, \gamma > 0$$

The baseline Reversed Hazard Rate of T is then obtained as

$$\lambda_0(t) = \frac{\alpha e^{\gamma t} \exp\left[-\frac{\alpha}{\gamma}\left(e^{\gamma t} - 1\right)\right]}{1 - \exp\left[-\frac{\alpha}{\gamma}\left(e^{\gamma t} - 1\right)\right]}$$

In the presence of the covariates X, we have the following

$$\lambda(t|X) = \frac{\alpha e^{\gamma t} \exp\left[-\frac{\alpha}{\gamma}\left(e^{\gamma t} - 1\right)\right] \exp(x_i \beta)}{1 - \exp\left[-\frac{\alpha}{\gamma}\left(e^{\gamma t} - 1\right)\right]}$$

$$F(t|X) = \left\{1 - \exp\left[-\frac{\alpha}{\gamma}\left(e^{\gamma t} - 1\right)\right]\right\}^{\exp(x_i \beta)}$$

$$f(t|X) = \alpha e^{\gamma t} \exp\left[-\frac{\alpha}{\gamma}\left(e^{\gamma t} - 1\right)\right] e^{x_i \beta} \left\{1 - \exp\left[-\frac{\alpha}{\gamma}\left(e^{\gamma t} - 1\right)\right]\right\}^{\exp(x_i \beta) - 1}$$

From these, the likelihood function for the Gompertz–Makeham distribution is obtained as

$$L(\alpha, \gamma, t) = \prod_{i=1}^{n} \left\{\alpha e^{\gamma t_i} \exp\left[-\frac{\alpha}{\gamma}\left(e^{\gamma t_i} - 1\right)\right] e^{x_i \beta}\right\}^{\delta_i}$$

$$\times \left\{1 - \exp\left[-\frac{\alpha}{\gamma}\left(e^{\gamma t_i} - 1\right)\right]\right\}^{\exp(x_i \beta) - \delta_i}$$

so that the log likelihood function is

$$l(\alpha, \gamma, t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i \ln \alpha + \sum_{i=1}^{n} \delta_i \gamma t_i - \sum_{i=1}^{n} \frac{\delta_i \alpha}{\gamma}\left(e^{\gamma t_i} - 1\right)$$

$$+ \sum_{i=1}^{n} \left(e^{x_i \beta} - \delta_i\right) \ln \left\{1 - \exp\left[-\frac{\alpha}{\gamma}\left(e^{\gamma t_i} - 1\right)\right]\right\}$$

Gamma Distribution

When the lifetime random variable follows a Gamma distribution, the baseline distribution function is given by

$$F_0(t) = \frac{\gamma(\alpha, \omega t)}{\Gamma(\alpha)}, \quad t > 0; \alpha, \omega > 0$$

where $\gamma(\alpha, t)$ is the incomplete Gamma function and $\Gamma(\alpha)$ is the complete Gamma function. The baseline Reversed Hazard Rate of T is then obtained as

$$\lambda_0(t) = \frac{\omega^\alpha t^{\alpha-1} e^{-\omega t}}{\gamma(\alpha, \omega t)}$$

In the presence of the covariates X, we have the following

$$\lambda(t|X) = \frac{\omega^\alpha t^{\alpha-1} \exp(-\omega t + x_i \beta)}{\gamma(\alpha, \omega t)}$$

$$F(t|X) = \left[\frac{\gamma(\alpha, \omega t)}{\Gamma(\alpha)}\right]^{\exp(x_i \beta)}$$

$$f(t|X) = \frac{\omega^\alpha t^{\alpha-1} \exp(-\omega t + x_i \beta)}{\gamma(\alpha, \omega t)} \left[\frac{\gamma(\alpha, \omega t)}{\Gamma(\alpha)}\right]^{\exp(x_i \beta)}$$

From these, the likelihood function for the Gamma distribution is obtained as

$$L(\alpha, \omega, t) = \prod_{i=1}^{n} \left[\frac{\omega^\alpha t_i^{\alpha-1} \exp(-\omega t_i + x_i \beta)}{\gamma(\alpha, \omega t_i)}\right]^{\delta_i} \left[\frac{\gamma(\alpha, \omega t_i)}{\Gamma(\alpha)}\right]^{\exp(x_i \beta)}$$

so that the log likelihood function is

$$l(\alpha, \omega, t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i \alpha \ln \omega + (\alpha - 1) \sum_{i=1}^{n} \delta_i \ln t_i - \sum_{i=1}^{n} \delta_i \omega t_i$$
$$+ \sum_{i=1}^{n} \left(e^{x_i \beta} - \delta_i\right) \ln \left[\gamma(\alpha, \omega t_i)\right] - \sum_{i=1}^{n} e^{x_i \beta} \ln \left[\Gamma(\alpha)\right]$$

Note that the Exponential distribution is a special case of this result.

Generalized Gamma Distribution

When the lifetime random variable follows a Generalized Gamma distribution, the baseline distribution function is given by

$$F_0(t) = \frac{\gamma\left[\frac{\alpha}{\omega}, (\lambda t)^\omega\right]}{\Gamma\left(\frac{\alpha}{\omega}\right)}, \quad t > 0; \alpha, \omega, \lambda > 0$$

where $\gamma(\alpha, t)$ is the incomplete Gamma function and $\Gamma(\alpha)$ is the complete Gamma function. The baseline Reversed Hazard Rate of T is then obtained as

$$\lambda_0(t) = \frac{\omega \lambda^\alpha t^{\alpha-1} e^{-(\lambda t)^\omega}}{\gamma\left[\frac{\alpha}{\omega}, (\lambda t)^\omega\right]}$$

In the presence of the covariates X, we have the following

$$\lambda(t|X) = \frac{\omega\lambda^{\alpha}t^{\alpha-1}e^{-(\lambda t)^{\omega}+x_i\beta}}{\gamma\left[\frac{\alpha}{\omega}, (\lambda t)^{\omega}\right]}$$

$$F(t|X) = \left[\frac{\gamma\left[\frac{\alpha}{\omega}, (\lambda t)^{\omega}\right]}{\Gamma\left(\frac{\alpha}{\omega}\right)}\right]^{\exp(x_i\beta)}$$

$$f(t|X) = \frac{\omega\lambda^{\alpha}t^{\alpha-1}e^{-(\lambda t)^{\omega}+x_i\beta}}{\gamma\left[\frac{\alpha}{\omega}, (\lambda t)^{\omega}\right]}\left[\frac{\gamma\left[\frac{\alpha}{\omega}, (\lambda t)^{\omega}\right]}{\Gamma\left(\frac{\alpha}{\omega}\right)}\right]^{\exp(x_i\beta)}$$

From these, the likelihood function for the Generalized Gamma distribution is obtained as

$$L(\alpha, \omega, \lambda, t) = \prod_{i=1}^{n}\left[\frac{\omega\lambda^{\alpha}t_i^{\alpha-1}e^{-(\lambda t_i)^{\omega}+x_i\beta}}{\gamma\left[\frac{\alpha}{\omega}, (\lambda t_i)^{\omega}\right]}\right]^{\delta_i}\left[\frac{\gamma\left[\frac{\alpha}{\omega}, (\lambda t_i)^{\omega}\right]}{\Gamma\left(\frac{\alpha}{\omega}\right)}\right]^{\exp(x_i\beta)}$$

so that the log likelihood function is

$$l(\alpha, \omega, \lambda, t) = \sum_{i=1}^{n}\delta_i x_i\beta + \sum_{i=1}^{n}\delta_i\alpha\ln(\omega\lambda) + (\alpha-1)\sum_{i=1}^{n}\delta_i\ln t_i - \sum_{i=1}^{n}\delta_i(\lambda t_i)^{\omega}$$

$$+ \sum_{i=1}^{n}(e^{x_i\beta} - \delta_i)\ln\left[\gamma\left[\frac{\alpha}{\omega}, (\lambda t_i)^{\omega}\right]\right] - \sum_{i=1}^{n}e^{x_i\beta}\ln\left[\Gamma\left(\frac{\alpha}{\omega}\right)\right]$$

Inverse Gamma Distribution

When the lifetime random variable follows a Inverse Gamma distribution, the baseline distribution function is given by

$$F_0(t) = \frac{\gamma(\alpha, t)}{\Gamma(\alpha)}, \quad t > 0; \alpha, \omega > 0$$

where $\gamma(\alpha, t)$ is the incomplete Gamma function and $\Gamma(\alpha)$ is the complete Gamma function. The baseline Reversed Hazard Rate of T is then obtained as

$$\lambda_0(t) = \frac{\omega^{\alpha}t^{-\alpha-1}e^{-\omega/t}}{\gamma(\alpha, t)}$$

In the presence of the covariates X, we have the following

$$\lambda(t|X) = \frac{\omega^{\alpha}t^{-\alpha-1}e^{-\omega/t}}{\gamma(\alpha, t)}\exp(x_i\beta)$$

$$F(t|X) = \left[\frac{\gamma(\alpha, t)}{\Gamma(\alpha)}\right]^{\exp(x_i\beta)}$$

$$f(t|X) = \frac{\omega^\alpha t^{-\alpha-1} e^{-\omega/t}}{\gamma(\alpha, t)} \exp(x_i \beta) \left[\frac{\gamma(\alpha, t)}{\Gamma(\alpha)} \right]^{\exp(x_i \beta)}$$

From these, the likelihood function for the Inverse Gamma distribution is obtained as

$$L(\alpha, \omega, t) = \prod_{i=1}^{n} \left[\frac{\omega^\alpha t^{-\alpha-1} e^{-\omega/t + x_i \beta}}{\gamma(\alpha, t)} \right]^{\delta_i} \left[\frac{\gamma(\alpha, t)}{\Gamma(\alpha)} \right]^{\exp(x_i \beta)}$$

so that the log likelihood function is

$$l(\alpha, \omega, t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i \alpha \ln \omega - (\alpha + 1) \sum_{i=1}^{n} \delta_i \ln t_i - \sum_{i=1}^{n} \frac{\delta_i \omega}{t_i}$$

$$+ \sum_{i=1}^{n} \left(e^{x_i \beta} - \delta_i \right) \ln \left[\gamma(\alpha, t_i) \right] - \sum_{i=1}^{n} e^{x_i \beta} \ln \left[\Gamma(\alpha) \right]$$

Weibull Distribution

When the lifetime random variable follows a Weibull distribution, the baseline distribution function is given by

$$F_0(t) = 1 - e^{-\left(\frac{t}{\gamma}\right)^\alpha}, \quad t > 0; \alpha, \gamma > 0$$

The baseline Reversed Hazard Rate of T is then obtained as

$$\lambda_0(t) = \frac{\left(\frac{\alpha}{\gamma}\right) \left(\frac{t}{\gamma}\right)^{\alpha-1} e^{-\left(\frac{t}{\gamma}\right)^\alpha}}{1 - e^{-\left(\frac{t}{\gamma}\right)^\alpha}}$$

In the presence of the covariates X, we have the following

$$\lambda(t|X) = \frac{\left(\frac{\alpha}{\gamma}\right) \left(\frac{t}{\gamma}\right)^{\alpha-1} e^{-\left(\frac{t}{\gamma}\right)^\alpha + x_i \beta}}{1 - e^{-\left(\frac{t}{\gamma}\right)^\alpha}}$$

$$F(t|X) = \left[1 - e^{-\left(\frac{t}{\gamma}\right)^\alpha} \right]^{\exp(x_i \beta)}$$

$$f(t|X) = \left[\left(\frac{\alpha}{\gamma}\right) \left(\frac{t}{\gamma}\right)^{\alpha-1} e^{-\left(\frac{t}{\gamma}\right)^\alpha + x_i \beta} \right] \left[1 - e^{-\left(\frac{t}{\gamma}\right)^\alpha} \right]^{\exp(x_i \beta) - 1}$$

From these, the likelihood function for the Weibull distribution is obtained as

$$L(\alpha, \gamma, t) = \prod_{i=1}^{n} \left[\left(\frac{\alpha}{\gamma}\right) \left(\frac{t_i}{\gamma}\right)^{\alpha-1} e^{-\left(\frac{t_i}{\gamma}\right)^{\alpha} + x_i\beta} \right]^{\delta_i} \left[1 - e^{-\left(\frac{t_i}{\gamma}\right)^{\alpha}} \right]^{\exp(x_i\beta) - \delta_i}$$

so that the log likelihood function is

$$l(\alpha, \gamma, t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i \ln \alpha - \sum_{i=1}^{n} \delta_i \alpha \ln \gamma + (\alpha - 1) \sum_{i=1}^{n} \delta_i \ln t_i$$
$$- \sum_{i=1}^{n} \delta_i \left(\frac{t_i}{\gamma}\right)^{\alpha} + \sum_{i=1}^{n} \left(e^{x_i\beta} - \delta_i\right) \ln \left[1 - e^{-\left(\frac{t}{\gamma}\right)^{\alpha}} \right]$$

Generalized Inverse Weibull Distribution

When the lifetime random variable follows a Generalized Inverse Weibull distribution, the baseline distribution function is given by

$$F_0(t) = e^{-\gamma\left(\frac{\lambda}{t}\right)^{\alpha}}, \quad t > 0; \alpha, \gamma, \lambda > 0$$

The baseline Reversed Hazard Rate of T is then obtained as

$$\lambda_0(t) = \alpha\gamma\lambda^{\alpha}t^{-(\alpha-1)}$$

In the presence of the covariates X, we have the following

$$\lambda(t|X) = \alpha\gamma\lambda^{\alpha}t^{-(\alpha-1)}e^{x_i\beta}$$
$$F(t|X) = \left[e^{-\gamma\left(\frac{\lambda}{t}\right)^{\alpha}}\right]^{\exp(x_i\beta)}$$
$$f(t|X) = \alpha\gamma\lambda^{\alpha}t^{-(\alpha-1)}e^{x_i\beta}\left[e^{-\gamma\left(\frac{\lambda}{t}\right)^{\alpha}}\right]^{\exp(x_i\beta)}$$

From these, the likelihood function for the Generalized Inverse Weibull distribution is obtained as

$$L(\alpha, \gamma, \lambda, t) = \prod_{i=1}^{n} \left[\alpha\gamma\lambda^{\alpha}t_i^{-(\alpha-1)}e^{x_i\beta}\right]^{\delta_i} \left[e^{-\gamma\left(\frac{\lambda}{t_i}\right)^{\alpha}}\right]^{\exp(x_i\beta)}$$

so that the log likelihood function is

$$l(\alpha, \gamma, \lambda, t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i \ln \alpha + \sum_{i=1}^{n} \delta_i \ln \gamma + \sum_{i=1}^{n} \delta_i \alpha \ln \lambda$$

$$-(\alpha - 1) \sum_{i=1}^{n} \delta_i \ln t_i - \sum_{i=1}^{n} \gamma \left(\frac{\lambda}{t_i}\right)^{\alpha} e^{x_i \beta}$$

Modified Weibull Distribution

When the lifetime random variable follows a Modified Weibull distribution [21], the baseline distribution function is given by

$$F_0(t) = 1 - \exp(-\gamma t^{\alpha} e^{\lambda t}), \quad t > 0; \alpha, \gamma, \lambda > 0$$

The baseline Reversed Hazard Rate of T is then obtained as

$$\lambda_0(t) = \frac{\gamma(\alpha + \lambda t)t^{\alpha-1} \exp(\lambda t) \exp(-\gamma t^{\alpha} e^{\lambda t})}{1 - \exp(-\gamma t^{\alpha} e^{\lambda t})}$$

In the presence of the covariates X, we have the following

$$\lambda(t|X) = \frac{\gamma(\alpha + \lambda t)t^{\alpha-1} \exp(\lambda t) \exp(-\gamma t^{\alpha} e^{\lambda t}) \exp(x_i \beta)}{1 - \exp(-\gamma t^{\alpha} e^{\lambda t})}$$

$$F(t|X) = \left[1 - \exp\left(-\gamma t^{\alpha} e^{\lambda t}\right)\right]^{\exp(x_i \beta)}$$

$$f(t|X) = \gamma(\alpha + \lambda t)t^{\alpha-1} \exp(\lambda t) \exp(-\gamma t^{\alpha} e^{\lambda t}) \exp(x_i \beta) \left[1 - \exp(-\gamma t^{\alpha} e^{\lambda t})\right]^{\exp(x_i \beta)-1}$$

From these, the likelihood function for the Modified Weibull distribution is obtained as

$$L(\alpha, \gamma, \lambda, t) = \prod_{i=1}^{n} \left[\gamma(\alpha + \lambda t_i)t_i^{\alpha-1} \exp(\lambda t_i) \exp(-\gamma t_i^{\alpha} e^{\lambda t_i}) \exp(x_i \beta)\right]^{\delta_i}$$
$$\times \left[1 - \exp(-\gamma t_i^{\alpha} e^{\lambda t_i})\right]^{\exp(x_i \beta)-\delta_i}$$

so that the log likelihood function is

$$l(\alpha, \gamma, \lambda, t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i \ln \gamma + \sum_{i=1}^{n} \delta_i \ln(\alpha + \lambda t_i) + (\alpha - 1) \sum_{i=1}^{n} \delta_i \ln t_i + \sum_{i=1}^{n} \delta_i \lambda t_i$$
$$- \sum_{i=1}^{n} \delta_i \gamma t_i^{\alpha} e^{\lambda t_i} + \sum_{i=1}^{n} \left(e^{x_i \beta} - \delta_i\right) \ln \left[1 - \exp(-\gamma t_i^{\alpha} e^{\lambda t_i})\right]$$

Flexible Weibull Distribution

When the lifetime random variable follows a Flexible Weibull distribution [3], the baseline distribution function is given by

$$F_0(t) = 1 - \exp\left(-e^{\alpha t - \frac{\gamma}{t}}\right), \quad t > 0; \alpha, \gamma > 0$$

The baseline Reversed Hazard Rate of T is then obtained as

$$\lambda_0(t) = \frac{\left(\alpha + \frac{\gamma}{t^2}\right) \exp\left(\alpha t - \frac{\gamma}{t}\right) \exp\left(-e^{\alpha t - \frac{\gamma}{t}}\right)}{1 - \exp\left(-e^{\alpha t - \frac{\gamma}{t}}\right)}$$

In the presence of the covariates X, we have the following

$$\lambda(t|X) = \frac{\left(\alpha + \frac{\gamma}{t^2}\right) \exp\left(\alpha t - \frac{\gamma}{t}\right) \exp\left(-e^{\alpha t - \frac{\gamma}{t}}\right) \exp(x_i \beta)}{1 - \exp\left(-e^{\alpha t - \frac{\gamma}{t}}\right)}$$

$$F(t|X) = \left[1 - \exp\left(-e^{\alpha t - \frac{\gamma}{t}}\right)\right]^{\exp(x_i \beta)}$$

$$f(t|X) = \left(\alpha + \frac{\gamma}{t^2}\right) \exp\left(\alpha t - \frac{\gamma}{t}\right) \exp\left(-e^{\alpha t - \frac{\gamma}{t}}\right) e^{x_i \beta} \left[1 - \exp\left(-e^{\alpha t - \frac{\gamma}{t}}\right)\right]^{\exp(x_i \beta) - 1}$$

From these, the likelihood function for the Flexible Weibull distribution is obtained as

$$L(\alpha, \gamma, t) = \prod_{i=1}^{n} \left[\left(\alpha + \frac{\gamma}{t_i^2}\right) \exp\left(\alpha t_i - \frac{\gamma}{t_i}\right) \exp\left(-e^{\alpha t_i - \frac{\gamma}{t_i}}\right) \exp(x_i \beta)\right]^{\delta_i}$$
$$\times \left[1 - \exp\left(-e^{\alpha t_i - \frac{\gamma}{t_i}}\right)\right]^{\exp(x_i \beta) - \delta_i}$$

so that the log likelihood function is

$$l(\alpha, \gamma, t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i \ln\left(\alpha + \frac{\gamma}{t_i^2}\right) + \sum_{i=1}^{n} \delta_i \left(\alpha t_i - \frac{\gamma}{t_i}\right) - \sum_{i=1}^{n} \delta_i e^{\alpha t_i - \frac{\gamma}{t_i}}$$
$$+ \sum_{i=1}^{n} \left(e^{x_i \beta} - \delta_i\right) \ln\left[1 - \exp\left(-e^{\alpha t_i - \frac{\gamma}{t_i}}\right)\right]$$

Power Generalized Weibull Distribution

When the lifetime random variable follows a Power Generalized Weibull distribution [1], the baseline distribution function is given by

$$F_0(t) = 1 - \exp\left[1 - \left(1 + \left(\frac{t}{\lambda}\right)^\alpha\right)^{\frac{1}{\gamma}}\right], \quad t > 0; \alpha, \gamma, \lambda > 0$$

The baseline Reversed Hazard Rate of T is then obtained as

$$\lambda_0(t) = \frac{\frac{\alpha}{\gamma\lambda^\alpha}t^{\alpha-1}\left[1 + \left(\frac{t}{\lambda}\right)^\alpha\right]^{\frac{1}{\gamma}-1}\exp\left[1 - \left(1 + \left(\frac{t}{\lambda}\right)^\alpha\right)^{\frac{1}{\gamma}}\right]}{1 - \exp\left[1 - \left(1 + \left(\frac{t}{\lambda}\right)^\alpha\right)^{\frac{1}{\gamma}}\right]}$$

In the presence of the covariates X, we have the following

$$\lambda(t|X) = \frac{\frac{\alpha}{\gamma\lambda^\alpha}t^{\alpha-1}\left[1 + \left(\frac{t}{\lambda}\right)^\alpha\right]^{\frac{1}{\gamma}-1}\exp\left[1 - \left(1 + \left(\frac{t}{\lambda}\right)^\alpha\right)^{\frac{1}{\gamma}}\right]\exp(x_i\beta)}{1 - \exp\left[1 - \left(1 + \left(\frac{t}{\lambda}\right)^\alpha\right)^{\frac{1}{\gamma}}\right]}$$

$$F(t|X) = \left\{1 - \exp\left[1 - \left(1 + \left(\frac{t}{\lambda}\right)^\alpha\right)^{\frac{1}{\gamma}}\right]\right\}^{\exp(x_i\beta)}$$

$$f(t|X) = \frac{\alpha}{\gamma\lambda^\alpha}t^{\alpha-1}\left[1 + \left(\frac{t}{\lambda}\right)^\alpha\right]^{\frac{1}{\gamma}-1}\exp\left[1 - \left(1 + \left(\frac{t}{\lambda}\right)^\alpha\right)^{\frac{1}{\gamma}}\right]\exp(x_i\beta)$$

$$\times \left\{1 - \exp\left[1 - \left(1 + \left(\frac{t}{\lambda}\right)^\alpha\right)^{\frac{1}{\gamma}}\right]\right\}^{\exp(x_i\beta)-1}$$

From these, the likelihood function for the Power Generalized Weibull distribution is obtained as

$$L(\alpha, \gamma, \lambda, t) = \prod_{i=1}^{n}\left\{\frac{\alpha}{\gamma\lambda^\alpha}t_i^{\alpha-1}\left[1 + \left(\frac{t_i}{\lambda}\right)^\alpha\right]^{\frac{1}{\gamma}-1}\exp\left[1 - \left(1 + \left(\frac{t_i}{\lambda}\right)^\alpha\right)^{\frac{1}{\gamma}}\right]\exp(x_i\beta)\right\}^{\delta_i}$$

$$\times \left\{1 - \exp\left[1 - \left(1 + \left(\frac{t_i}{\lambda}\right)^\alpha\right)^{\frac{1}{\gamma}}\right]\right\}^{\exp(x_i\beta)-\delta_i}$$

so that the log likelihood function is

$$l(\alpha, \gamma, \lambda, t) = \sum_{i=1}^{n}\delta_i x_i\beta + \sum_{i=1}^{n}\delta_i\ln\alpha - \sum_{i=1}^{n}\delta_i\ln\gamma - \sum_{i=1}^{n}\delta_i\alpha\ln\lambda + (\alpha-1)\sum_{i=1}^{n}\delta_i\ln t_i$$

$$+ \left(\frac{1}{\gamma} - 1 \right) \sum_{i=1}^{n} \delta_i \ln \left[1 + \left(\frac{t_i}{\lambda} \right)^{\alpha} \right] + \sum_{i=1}^{n} \delta_i \left[1 - \left(1 + \left(\frac{t_i}{\lambda} \right)^{\alpha} \right)^{\frac{1}{\gamma}} \right]$$

$$+ \sum_{i=1}^{n} \left(e^{x_i \beta} - \delta_i \right) \ln \left\{ 1 - \exp \left[1 - \left(1 + \left(\frac{t_i}{\lambda} \right)^{\alpha} \right)^{\frac{1}{\gamma}} \right] \right\}$$

Marshal–Olkin Distribution

When the lifetime random variable follows a Marshal–Olkin distribution [26], the baseline distribution function is given by

$$F_0(t) = \frac{1 - e^{-(\lambda t)^{\alpha}}}{1 - (1 - \gamma)e^{-(\lambda t)^{\alpha}}}, \quad t > 0; \alpha, \gamma, \lambda > 0$$

The baseline Reversed Hazard Rate of T is then obtained as

$$\lambda_0(t) = \frac{\gamma \alpha \lambda (\lambda t)^{\alpha - 1} e^{-(\lambda t)^{\alpha}}}{\left[1 - (1 - \gamma)e^{-(\lambda t)^{\alpha}} \right] \left[1 - e^{-(\lambda t)^{\alpha}} \right]}$$

In the presence of the covariates X, we have the following

$$\lambda(t|X) = \frac{\gamma \alpha \lambda (\lambda t)^{\alpha - 1} e^{-(\lambda t)^{\alpha}} e^{x_i \beta}}{\left[1 - (1 - \gamma)e^{-(\lambda t)^{\alpha}} \right] \left[1 - e^{-(\lambda t)^{\alpha}} \right]}$$

$$F(t|X) = \left[\frac{1 - e^{-(\lambda t)^{\alpha}}}{1 - (1 - \gamma)e^{-(\lambda t)^{\alpha}}} \right]^{\exp(x_i \beta)}$$

$$f(t|X) = \frac{\gamma \alpha \lambda (\lambda t)^{\alpha - 1} e^{-(\lambda t)^{\alpha}} e^{x_i \beta} \left[1 - e^{-(\lambda t)^{\alpha}} \right]^{\exp(x_i \beta) - 1}}{\left[1 - (1 - \gamma)e^{-(\lambda t)^{\alpha}} \right]^{\exp(x_i \beta) + 1}}$$

From these, the likelihood function for the Marshal–Olkin distribution is obtained as

$$L(\alpha, \gamma, \lambda, t) = \prod_{i=1}^{n} \left\{ \frac{\gamma \alpha \lambda (\lambda t_i)^{\alpha - 1} e^{-(\lambda t_i)^{\alpha}} e^{x_i \beta} \left[1 - e^{-(\lambda t_i)^{\alpha}} \right]^{\exp(x_i \beta) - 1}}{\left[1 - (1 - \gamma)e^{-(\lambda t_i)^{\alpha}} \right]^{\exp(x_i \beta) + 1}} \right\}^{\delta_i}$$

$$\times \left\{ \frac{1 - e^{-(\lambda t_i)^{\alpha}}}{1 - (1 - \gamma)e^{-(\lambda t_i)^{\alpha}}} \right\}^{(1 - \delta_i) \exp(x_i \beta)}$$

so that the log likelihood function is

$$l(\alpha, \gamma, \lambda, t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i \ln \gamma + \sum_{i=1}^{n} \delta_i \ln \alpha + \sum_{i=1}^{n} \delta_i \alpha \ln \lambda + (\alpha - 1) \sum_{i=1}^{n} \delta_i \ln t_i$$

$$- \sum_{i=1}^{n} \delta_i \, (\lambda t_i)^{\alpha} + \sum_{i=1}^{n} \left(e^{x_i \beta} - \delta_i \right) \ln \left[1 - e^{-(\lambda t_i)^{\alpha}} \right]$$

$$- \sum_{i=1}^{n} \left(e^{x_i \beta} + \delta_i \right) \ln \left[1 - (1 - \gamma) e^{-(\lambda t_i)^{\alpha}} \right]$$

References

1. Bagdonavicius, V., Nikulin, M.: Accelerated Life Models: Modeling and Statistical Analysis. CRC Press, London (2001)
2. Barlow, R.E., Marshall, A.W., Proschan, F.: Properties of probability distributions with monotone hazard rate. Ann. Math. Stat. **5**, 375–389 (1963)
3. Bebbington, M., Lai, C.D., Zitikis, R.: A flexible weibull extension. Reliab. Eng. Syst. Saf. **92**(6), 719–726 (2007)
4. Beer, L., Mattson, C.L., Bradley, H., Skarbinski, J., et al.: Understanding cross-sectional racial, ethnic, and gender disparities in antiretroviral use and viral suppression among hiv patients in the united states. Medicine **95**(13), e3171 (2016)
5. Block, H.W., Savits, T.H., Singh, H.: The reversed hazard rate function. Probab. Eng. Inf. Sci. **12**(01), 69–90 (1998)
6. Chakraborty, H., Sen, P.K., Helms, R.W., Vernazza, P.L., Fiscus, S.A., Eron, J.J., Patterson, B.K., Coombs, R.W., Krieger, J.N., Cohen, M.S.: Viral burden in genital secretions determines male-to-female sexual transmission of hiv-1: a probabilistic empiric model. AIDS **15**(5), 621–627 (2001)
7. Chakraborty, H., Iyer, M., Duffus, W.A., Samantapudi, A.V., Albrecht, H., Weissman, S.: Disparities in viral load and cd4 count trends among hiv-infected adults in south carolina. AIDS Patient Care STDS **29**(1), 26–32 (2015)
8. Chakraborty, H., Weissman, S., Duffus, W.A., Hossain, A., Varma Samantapudi, A., Iyer, M., Albrecht, H.: Hiv community viral load trends in south Carolina. Int. J. STD AIDS **28**(3), 265–276 (2017)
9. Chakraborty, H., Hossain, A., Latif, M.A.: A three-state continuous time markov chain model for HIV disease burden. J. Appl. Stat. **46**(9), 1671–1688 (2019)
10. Cohen, M.S., Chen, Y.Q., McCauley, M., Gamble, T., Hosseinipour, M.C., Kumarasamy, N., Hakim, J.G., Kumwenda, J., Grinsztejn, B., Pilotto, J.H., et al.: Prevention of HIV-1 infection with early antiretroviral therapy. N. Engl. J. Med. **365**(6), 493–505 (2011)
11. Cox, D.R.: Regression models and life-tables. J. R. Stat. Soc. Ser. B (Methodological) **34**(2), 187–220 (1972)
12. Di Crescenzo, A.: Some results on the proportional reversed hazards model. Stat. Probab. Lett. **50**(4), 313–321 (2000)
13. Dieffenbach, C.W.: Preventing hiv transmission through antiretroviral treatment mediated virologic suppression: aspects of an emerging scientific agenda. Curr. Opin. HIV AIDS **7**(2), 106–110 (2012)
14. Eisinger, R.W., Dieffenbach, C.W., Fauci, A.S.: HIV viral load and transmissibility of HIV infection: undetectable equals untransmittable. J. Am. Med. Assoc. **321**(5), 451–452 (2019)
15. Gill, C.J., Griffith, J.L., Jacobson, D., Skinner, S., Gorbach, S.L., Wilson, I.B.: Relationship of HIV viral loads, cd4 counts and heart use to health-related quality of life. J. Acquir. Immune Defic. Syndr. **30**(5), 485–492 (1999)
16. Gupta, R.C., Wu, H.: Analyzing survival data by proportional reversed hazard model. Int. J. Reliab. Appl. **2**(1), 1–26 (2001)
17. Henderson, R., Diggle, P., Dobson, A.: Joint modelling of longitudinal measurements and event time data. Biostatistics **1**(4), 465–480 (2000)
18. Kleinschmidt, I., Schwabe, C., Benavente, L., Torrez, M., Ridl, F.C., Segura, J.L., Ehmer, P., Nchama, G.N.: Marked increase in child survival after four years of intensive malaria control. Am. J. Trop. Med. Hyg. **80**(6), 882–888 (2009)
19. Kong, M.C., Nahata, M.C., Lacombe, V.A., Seiber, E.E., Balkrishnan, R.: Association between race, depression, and antiretroviral therapy adherence in a low-income population with HIV infection. J. Gen. Intern. Med. **27**(9), 1159–1164 (2012)

20. Kundu, D., Gupta, R.D.: A class of bivariate models with proportional reversed hazard marginals. Sankhya B **72**(2), 236–253 (2010)
21. Lai, C., Xie, M., Murthy, D.: A modified weibull distribution. IEEE Trans. Reliab. **52**(1), 33–37 (2003)
22. Langebeek, N., Gisolf, E.H., Reiss, P., Vervoort, S.C., Hafsteinsdóttir, T.B., Richter, C., Sprangers, M.A., Nieuwkerk, P.T.: Predictors and correlates of adherence to combination antiretroviral therapy (art) for chronic HIV infection: a meta-analysis. BMC Med. **12**(1), 142 (2014)
23. Leung, K.M., Elashoff, R.M., Afifi, A.A.: Censoring issues in survival analysis. Annu. Rev. Public Health **18**(1), 83–104 (1997)
24. Lindsey, J.: A study of interval censoring in parametric regression models. Lifetime Data Anal. **4**(4), 329–354 (1998)
25. Manolescu, L., Marinescu, P.: Sex differences in HIV-1 viral load and absolute cd4 cell count in long term survivors HIV-1 infected patients from giurgiu, romania. Rom. Rev. Lab. Med. **21**(2), 217–224 (2013)
26. Marshall, A.W., Olkin, I.: A new method for adding a parameter to a family of distributions with application to the exponential and weibull families. Biometrika **84**(3), 641–652 (1997)
27. Muthulingam, D., Chin, J., Hsu, L., Scheer, S., Schwarcz, S.: Disparities in engagement in care and viral suppression among persons with HIV. JAIDS J. Acquir. Immune Defic. Syndr. **63**(1), 112–119 (2013)
28. Nanda, A.K., Das, S.: Dynamic proportional hazard rate and reversed hazard rate models. J. Stat. Plan. Inference **141**(6), 2108–2119 (2011)
29. Nanda, A.K., Shaked, M.: The hazard rate and the reversed hazard rate orders, with applications to order statistics. Ann. Inst. Stat. Math. **53**(4), 853–864 (2001)
30. Olatosi, B.A., Probst, J.C., Stoskopf, C.H., Martin, A.B., Duffus, W.A.: Patterns of engagement in care by HIV-infected adults: South Carolina, 2004–2006. AIDS **23**(6), 725–730 (2009)
31. Passamonti, F., Rumi, E., Arcaini, L., Boveri, E., Elena, C., Pietra, D., Boggi, S., Astori, C., Bernasconi, P., Varettoni, M., et al.: Prognostic factors for thrombosis, myelofibrosis, and leukemia in essential thrombocythemia: a study of 605 patients. Haematologica **93**(11), 1645–1651 (2008)
32. Reif, S., Geonnotti, K.L., Whetten, K.: Hiv infection and aids in the deep south. Am. J. Public Health **96**(6), 970–973 (2006)
33. Reif, S., Whetten, K., Wilson, E., Gong, W.: HIV/aids epidemic in the south reaches crisis proportions in last decade (2011)
34. Rosholm, M.: An analysis of the processes of labor market exclusion and (re-) inclusion (2001)
35. Rosholm, M.: HIV/aids epidemic in the south reaches crisis proportions in last decade (2002)
36. Roy, D.: A characterization of model approach for generating bivariate life distributions using reversed hazard rates. J. Jpn. Stat. Soc. **32**(2), 239–245 (2002)
37. Sankaran, P., Gleeja, V.: On bivariate reversed hazard rates. J. Jpn. Stat. Soc. **36**(2), 213–224 (2006)
38. South Carolina Department of Health and Environmental Control (SCDHEC): An epidemiologic profile of HIV and AIDS in South Carolina (2015)
39. Sparling, Y.H., Younes, N., Lachin, J.M., Bautista, O.M.: Parametric survival models for interval-censored data with time-dependent covariates. Biostatistics **7**(4), 599–614 (2006)
40. Thiébaut, R., Jacqmin-Gadda, H., Babiker, A., Commenges, D.: Joint modelling of bivariate longitudinal data with informative dropout and left-censoring, with application to the evolution of cd4+ cell count and HIV rna viral load in response to treatment of HIV infection. Stat. Med. **24**(1), 65–82 (2005)
41. Tripathi, A., Youmans, E., Gibson, J.J., Duffus, W.A.: The impact of retention in early HIV medical care on viro-immunological parameters and survival: a statewide study. AIDS Res. Hum. Retroviruses **27**(7), 751–758 (2011)
42. US Department of Health & Human Services: Viral Load (2020). https://www.aids.gov/hiv-aids-basics/index.html
43. Variyath, A., Sankaran, P.: Parametric regression models using reversed hazard rates. J. Probab. Stat. **2014**, 1–5 (2014)
44. Weis, K.E., Liese, A.D., Hussey, J., Gibson, J.J., Duffus, W.A.: Associations of rural residence with timing of HIV diagnosis and stage of disease at diagnosis, South Carolina 2001–2005. J. Rural Health **26**(2), 105–112 (2010)
45. Weissman, S., Duffus, W.A., Iyer, M., Chakraborty, H., Samantapudi, A.V., Albrecht, H.: Rural-urban differences in HIV viral loads and progression to aids among new HIV cases. South. Med. J. **108**(3), 180–188 (2015)

46. World Health Organization: WHO Fact Sheets: HIV/AIDS (2020). https://www.who.int/news-room/fact-sheets/detail/hiv-aids

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.