

RESEARCH

Open Access



Accelerated failure-time model with weighted least-squares estimation: application on survival of HIV positives

Yesuf Abdela Mustefa^{1*}  and Ding-Geng Chen²

Abstract

Background: Survival analysis is the most appropriate method of analysis for time-to-event data. The classical accelerated failure-time model is a more powerful and interpretable model than the Cox proportional hazards model, provided that model imposed distribution and homoscedasticity assumptions satisfied. However, most of the real data are heteroscedastic which violates the fundamental assumption and consequently, the statistical inference could be erroneous in accelerated failure-time modeling. The weighted least-squares estimation for the accelerated failure-time model is an efficient semi-parametric approach for time-to-event data without the homoscedasticity assumption, which is developed recently and not often utilized for real data analysis. Thus, this study was conducted to ascertain the better performance of the weighted least-squares estimation method over the classical methods.

Methods: We analyzed a REAL dataset on Antiretroviral Therapy patients we recently collected. We compared the results from classical methods of estimation for the accelerated failure-time model with the results revealed from the weighted least-squares estimation.

Results: We found that the data are heteroscedastic and indicated that the weighted least-square method should be used to analyze this data. The weighted least-squares estimation revealed more accurate, and efficient estimates of covariates effect since its confidence intervals were shorter and it identified more significant covariates. Accordingly, the survival of HIV positives was found to be significantly linked with age, weight, functional status, CD4 (Cluster of Differentiation agent 4 glycoproteins), and clinical stages.

Conclusions: The weighted least-squares estimation performed the best in providing more significant effects and precise estimates than the classical accelerated failure-time methods of estimation if data are heteroscedastic. Thus, we recommend future researchers should utilize weighted least-squares estimation rather than the classical methods when the homoscedasticity assumption is violated.

Keywords: Survival data analysis, Accelerated failure-time, Cox proportional-hazards regression, Weighted least-squares estimation, Heteroscedasticity

Background

Although the Cox proportional-hazards (PH) model [1] is the most employed technique in survival analysis

because of its reduced set of assumptions about the baseline hazard function, formulation of the accelerated failure-time (AFT) model [2] allows the derivation of a time ratio, which is more interpretable than a ratio of two hazards [3]. The AFT model doesn't require PH assumption which is seldom met assumption of the Cox PH model [3]. It also encompasses relatively a wide

* Correspondence: yesufabdela777@gmail.com

¹Computing and Analytics Directorate, Technology and Innovation Institute of Ethiopia, Addis Ababa, Ethiopia

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



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

range of survival time distributions and yields more powerful estimates than the Cox PH model, provided that method imposed assumptions satisfied [4]. Thus, the AFT model is more appealing in many ways [5].

Conventionally, the rank [6–13] and least-squares [14–17] are most often used methods of inference for the AFT model [18]. The classical AFT models in general are unified by the adoption of a log-linear representation with a particular survival time distribution for the error term [4]. These classical methods impose constant variance (homoscedasticity) assumption and hence, don't take heteroscedastic data into consideration. Consequently, rank or least-squares estimation (LSE) based inference for heteroscedastic data is not reliable [5, 18]. The coverage probabilities of the 95% confidence intervals (CI) of the estimated coefficients are considerably lower than the nominal level 0.95 because the variance estimates of the parameter estimators are mostly underestimated with the homogeneous assumption of variance. Besides, it results in the loss of efficiency for coefficient estimators. Moreover, the intercept estimation is inconsistent.

To incapacitate these negative aspects, Yu, et al. [18] proposed the weighted least-squares estimation (WLSE) for the AFT model. It is a semi-parametric approach to handle both homoscedastic and heteroscedastic data beyond the incorporation of censoring, which was shown to be more efficient theoretically as well as extensive Monte-Carlo simulation studies. Specifically, Yu, et al. [18] showed that the WLSE was more efficient than the classical AFT method with higher statistical power to detect the associated statistical significance with lower standard errors of estimates and narrower confidence intervals due to lower variances obtained from synthetic observations combined with the actual observations.

Therefore, this article is aimed to promote the utilization of this WLSE model in analyzing our HIV data to detect significant effects of covariates on HIV patients' survival for a more valid inference and conclusions. In doing so, we compared the classical AFT method with WLSE to ascertain the validity, detective ability, and efficiency of inference from WLSE based on real Antiretroviral Therapy (ART) dataset with more covariates. The magnitude of standard errors of estimates, the width of confidence intervals, and the number of significant covariates was considered as comparison criteria due to the WLSE has been shown to be efficient and can recover the true effect of the parameters and their variances.

Data and research method

Description of the dataset

The data was obtained from University of Gondar referral hospital ART database. Patients whose ART starting

date unrecorded were excluded. On the other hand, a total of 3042 patients with complete records of their baseline characteristics were targeted for the study. We used systematic random sampling since a complete list was available. To determine the sample size, we used Cochran's [19] formula $n = \frac{Z_{0.025}^2 \pi(1-\pi)/E^2}{1 + \frac{Z_{0.025}^2 \pi(1-\pi)/E^2}{N}}$ where: n is the number of patients required for the study; $Z_{0.025}^2$ is the upper 25th percentile point of the standard normal distribution and it is 1.96; π is the proportion of death in the study population which was 0.17; $E = 0.05$ was margin of error; and N was the total number of patients with complete records of baseline characteristics. Accordingly, a sample of 203 HIV patients who started ART between 2003 and 2009 were followed until April 2015.

The response variable was considered to be the length of time measured in months from ART initiation until time of death (censor).

Covariates of the study

There were six covariates recorded at the beginning of ART, namely "gender" with two levels (male and female); "functional status" with three levels (working, ambulatory, and bedridden); WHO "Clinical Stage" with four levels (I, II, III, and IV); Cluster of Differentiation Agent 4 (CD4) percent with two levels (12–15%, and 16–28%), age and weight. We used CD4 percent instead of CD4 count because it was recorded in the database. It was recorded in the form of class intervals for some of the patients and in the form of actual values for the others. Thus, we used the Struges [20] rule to determine the appropriate number of class groupings. However, we merged all percent categories greater than or equal to 16% since the percentages of death in those categories were very small. Similarly, we merged the first three categories of the WHO clinical stage.

Statistical methods

The log-rank test

We used the log-rank test to determine which covariates to select for further analysis. The log-rank test, developed by Mantel and Haenszel [21], is a non-parametric test for equality of survival functions in two or more groups. Let t_i be times where events are observed (assume these are ordered and there are D such times); d_{ik} be the number of observed events from group k at time t_i ; Y_{ik} be the number of subjects in group k that are at risk at time t_i , $d_i = \sum_{j=1}^n d_{ij}$, $Y_i = \sum_{j=1}^n Y_{ij}$, and n is the number of comparison groups. Then to test the hypothesis, a vector Z is computed, where the k^{th} element is $Z_k = \sum_{i=1}^D (d_{ik} - Y_{ik} \frac{d_i}{Y_i})$. The test statistic $Q = Z' \hat{\Sigma}^{-1} Z$ where $\hat{\Sigma}$ is

variance covariance matrix obtained from the data follows chi-square distribution with n degrees of freedom. Larger value of the test statistic or the corresponding smaller P -value provides strong evidence against the null hypothesis of no difference.

Assessment plots

We used plots to assess the time-varying nature of covariates; heteroscedasticity in the data; and the linearity assumption of covariates. Plots of coefficients of covariates over time help examine the time-varying nature of covariates [4]. Reference lines at zero falling within the 95% confidence limits provide evidence that covariates are not time-varying. Plot of the estimated variances versus the means can be used to assess heteroscedasticity in the data [18]. Any pattern suggests the presence of heteroscedasticity. Plot of the log-survival time against a covariate can be used to assess the linearity assumption and a straight Locally Weighted Scatterplot Smoothing (LOWESS) line close to a horizontal reference line shows no violation to the linearity assumption [4].

The classical AFT model

The AFT model regresses survival time \mathcal{T} on covariates X as follows,

$$T_i \equiv \log(\mathcal{T}_i) = \alpha_0 + \beta_0^T X_i + \epsilon_i, \quad i = 1, 2, \dots, n \text{ Eq. 1.}$$

Where α_0 is the true intercept; β_0 is the true p -dimensional vector of slope parameters; the index “ i ” corresponds to the study participant in the analysis and n is the total number of patients. Let $\tilde{X} = [1 \ X]$ and $\tilde{\beta}_0^T = [\alpha_0 \ \beta_0^T]$ where $\tilde{\beta}_0$ is the vector of coefficients for the AFT model [18].

In this model, let $\epsilon_i = \sigma e_i$. Then, e_i is the error term which is independently and identically distributed (IID) with unspecified distribution function F of mean 0 and variance 1. In other words, ϵ_i is an IID error term with a constant variance σ^2 (homoscedasticity) assumption and a particular survival time distribution [5]. The most commonly used survival distributions for AFT metrics are exponential (Exp), Weibull (Weib), log-logistic (Logl), lognormal (LN), and generalized gamma (GG) [3]. The exponential distribution is the special case of the Weibull distribution. Similarly, the generalized gamma distribution includes a wide range of family distribution as its special case. Its flexible hazard function allows for many possible shapes such as Weibull, exponential, and lognormal distributions with various values of the shape and the scale parameters [4]. Therefore, we used the GG AFT model for evaluating and selecting an appropriate model from its parametrically nested possible AFT models for the dataset by testing the shape and the scale parameters.

Moreover, we used information criterion statistics (ICS), $-2 \log \text{likelihood}$ ($-2LL$) for comparison of the performance of parametrically nested classical AFT models and Akaike information criterion (AIC) for comparison the performance of other alternative classical AFT models. Accordingly, the best model was the one with the smallest ICS indicating the minimum loss of information.

Finally, the performance of the best-fitted classical AFT model was compared with the model fitted by the WLSE. This comparison was based on either the magnitude of standard errors (SE) of estimates or the length of confidence intervals to determine the accuracy and efficiency of a given inference. The ability to detect more significant covariates was another interest of comparison.

The weighted least-squares method

The WLSE method frees the homoscedasticity assumption in Eq. 1, which is the most practical method for real data analysis.

Define $Z = \min \{T, C\}$ and $\delta = I(T \leq C)$, where C is the logarithm of the censoring time, T is as defined in Eq. 1, $I(\cdot)$ is the indicator function. Then, the triplet denoted by $\{Z_i, X_i, \delta_i\}$ represents values of a transformation of observed survival time (Z_i), covariates (X_i), and censoring indicator function (δ_i) for the i^{th} patient where C_i is assumed to be independent of T_i and X_i . The synthetic observation is defined to be $T_i^* = Z_i \delta_i + E(T_i | T_i > C_i)(1 - \delta_i)$ $i = 1, 2, \dots, n$, where, $E(\cdot)$ is the expectation function. The WLSE utilizes a weighted least-squares equation as in Yu, et al. [18] with synthetic observations weighted by the square root of their variances where the variances are estimated via the local polynomial regression. Thus, the weighted regression according to Yu, et al. [18] is as follows.

$$T_{i \text{ new}}^* = \alpha_0 x_{i \text{ new}} + \beta_0^T X_{i \text{ new}} + e_i, \quad i = 1, 2, \dots, n \text{ Eq. 2.}$$

Where $T_{i \text{ new}}^* = T_i^* / \sigma_{NPE}(\mu_i)$; $\sigma_{NPE}(\mu_i)$ is the non-parametric estimator (NPE) via the local polynomial regression of $\sigma^*(\mu_i)$, the square root of the variance of T_i^* ; $x_{i \text{ new}} = 1 / \sigma_{NPE}(\mu_i)$; $X_{i \text{ new}} = X_i / \sigma_{NPE}(\mu_i)$; $e_i = v_i$, and n are as defined in Eq. 1.

Based on the weighted least-squares regression in Eq. 2, all $\tilde{\beta}_0$, including α_0 , are slope parameters.

Results

Covariates with a probability of being significantly less than or equal to 10% in the log-rank test were potentially considered for further analysis besides age and weight. The P -value for gender is greater than 0.1(10% probability of inclusion) (Table 1). Thus, it was not further included in the models. Since age and weight were numeric, we tested them one at time (independently)

Table 1 Patients Characteristics

Covariates	Labels	Number of Patients	Number of Deaths	LRT P-Value
Gender	Female	118	21	1.000
	Male	85	14	1.000
FS	Working	136	15	0.002
	Ambulatory	55	14	0.002
CS	I, II, & III	154	13	0.000
	IV	49	22	0.000
CD4%	12–15%	162	33	0.100
	16–28%	41	2	0.100

FS: Functional Status; CS: Clinical Stage; CD4%: Cluster of Differentiation Agent Four in Percent; LRT: Log-Rank Test.

using all survival models considered in this study and we found them significant at 5% level of significance for further analysis. The averages of age and weight for the study patients were 31 years old and 45 Kg respectively.

We proceeded further with an assessment of the time-varying nature of covariates in Fig. 1. We used the Cox PH model for this test. We found that the reference line on zero differences falls within the 95% confidence intervals for age, weight, and CD4%, which confirms that these covariates are not time-varying.

We also examined the plot of the estimated variances versus the means from the weight least-squares method, as shown in Fig. 2. The data was found to be

heteroscedastic, which violates the assumption of the classical AFT models.

We further examined the linearity assumption using locally weighted scatter plot smoothing (LOWESS). As seen in Fig. 3, there were no violations to the covariates “age” and “weight”. Thus, we used the values of age and weight as they were rather than their transformed form. Furthermore, there was no significant interaction effect to be included in the models.

According to Yu, et al. [18], the WLSE is a valid inference in such cases. We summarized the results in Table 2. The reference categories are working for functional status; I, II, & III for clinical-stage; and 12–15% for CD4%. The level of significance is considered to be 5%.

In addition to the WLSE, we revealed the results from the classical methods namely, parametric AFT model, rank, and least-squares estimation (LSE) to ascertain the superiority of WLSE over the classical methods. We selected the appropriate parametric AFT model for the dataset among the exponential, Weibull, log-logistic, log-normal, and generalized gamma AFT models.

The shape parameter (Q) from the classical GG AFT model in Table 3, is not significantly different from 0 or 1. This indicates that the LN and the Weibull AFT models are likely to be appropriate among the special cases of GG AFT model. However, Q is estimated to be 0.411 and it is nearer to 0 than to 1. For this conclusion, we selected the LN AFT model over the Weibull. Moreover, we considered ICS in Table 4 for selecting the most appropriate model.

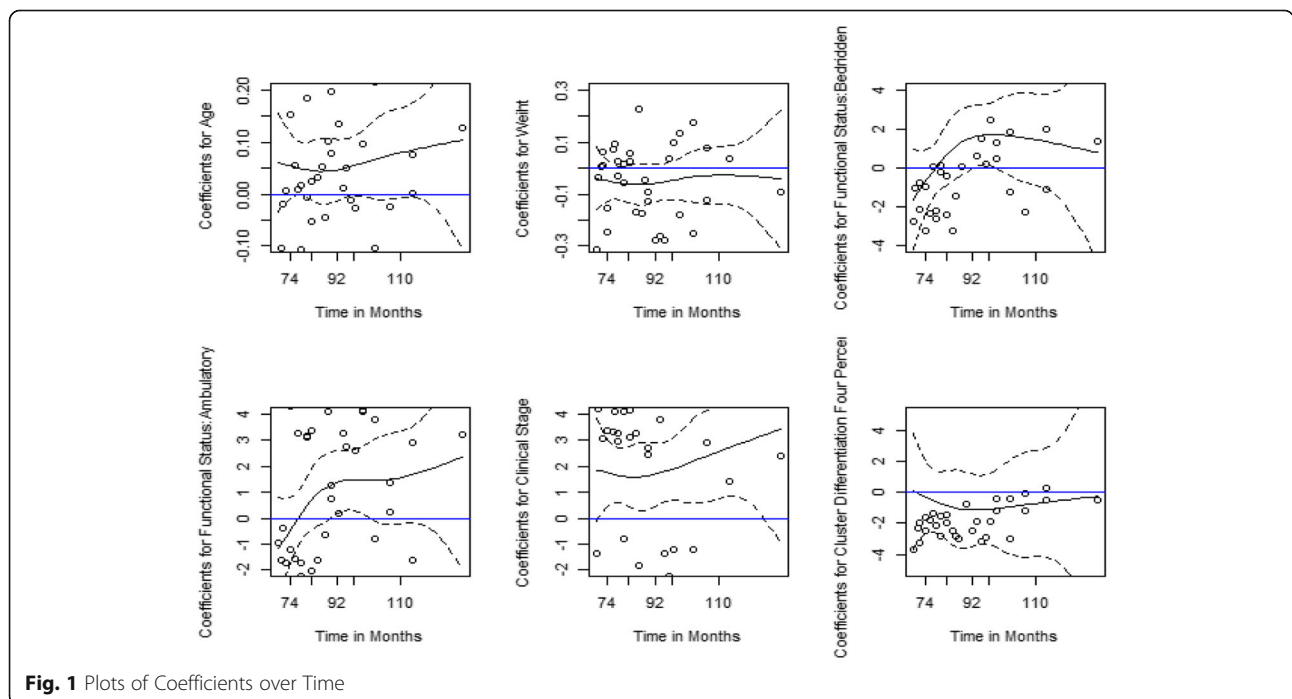
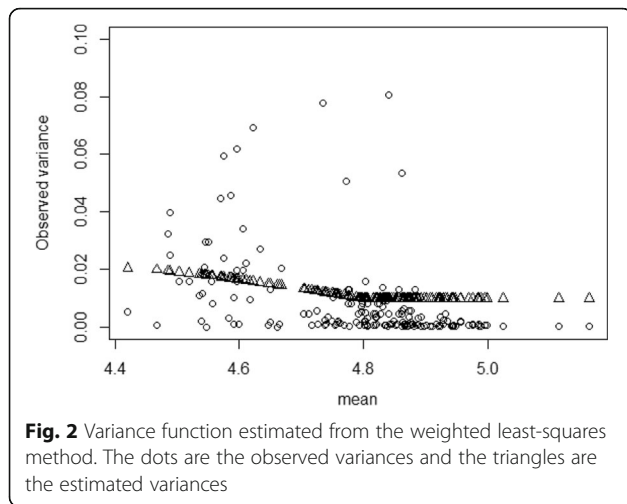


Fig. 1 Plots of Coefficients over Time



The smaller magnitude of the information criterion statistics (ICS) for LN AFT model led us to the conclusion that the LN AFT model is the best fit for the data. Therefore, we used the LN AFT model to represent the parametric AFT models for comparison. Furthermore, the estimation performance of the other parametric AFT models was the same as LN AFT model as in Table 5.

As seen in Table 5, the performance of the three classical methods was similar in this particular study. Accordingly, age, weight, and WHO clinical stage were significant based on these methods. However, an efficient method is obviously more likely to identify significant covariates. Thus, we found the LSE to be the most efficient among the classical methods in Table 5 since the standard errors (SE) associated with its estimates are relatively small. Similarly, LN AFT model was more efficient as compared to the rank method. The Cox PH model identified the same significant covariates as the classical AFT models. However, directions of the effects

Table 2 Weighted Least-squares Estimation

Covariates	Estimates	SE	95% CI	
			LCL	UCL
Constant	4.773	0.026020	4.722	4.824
Age	-0.00516	0.000010	-0.00518	-0.00514
Weight	0.00497	0.000005	0.00495	0.00498
BFS	-0.04742	0.013663	-0.07420	-0.02064
AFS	-0.02048	0.003439	-0.02721	-0.01374
CS (IV)	-0.25080	0.002214	-0.25513	-0.24646
CD4%(16–28)	0.10320	0.046816	0.01143	0.19496

Estimates: Coefficients Correspond to the log Median Survival Time; Estimate for Age corresponds to a year increase in Age; Estimate for Weight corresponds to a Kilogram increase in Weight; CI: Confidence Interval; LCL: Lower Confidence Limit; UCL: Upper Confidence Limit; BFS: Bedridden Functional Status and its reference category is Working Functional Status; AFS: Ambulatory Functional Status and its reference category is Working Functional Status; CS (IV): Clinical Stage IV and its reference category is Clinical Stage I-III; CD4%16–28: Cluster of Differentiation Agent Four Percent 16–28 and its reference category is Cluster of Differentiation Agent Four Percent 12–15.

differed in the Cox PH model since estimates were affected by the time-varying nature of clinical stage.

Nevertheless, we compared the performance of LSE with the WLSE since the LSE was the best among the classical methods. The results in Table 2 revealed that the WLSE is more accurate than the LSE. It resulted in efficient estimates of covariates effect on HIV patients' survival since the narrow CI indicates a relatively small standard error of estimates as theoretically showed in Yu, et al. [18]. It also identified more significant covariates since functional status and CD4 percent were additional significant covariates that were not identified by LSE. Though we considered the same set of covariates in all models at the beginning, we included only significant covariates in the final models as per their ability to identify significant effects. We interpreted the results of WLSE as follows.

Holding the effects of all other factors constant, the log survival time for a patient with an additional year of age decreases by 0.00516. In other words, the median survival time of a patient is $e^{0.00516} = 1.01$ times as compared to that of a patient with a single year older. This indicates that the survival probability of younger patients is better than older. Studies by [22–24] confirmed that the hazard of death increases with higher age intervals. These studies also confirmed that less weight, low level of CD4, higher clinical stage, and non-working functional status are associated with increased hazard rate. Similarly, we found that the logarithm of survival is 0.00497 more for a kilogram of additional weight. For patients with bedridden and ambulatory functional status, it is respectively 0.047 and 0.0205 less as compared to those with working functional status. Moreover, it is 0.251 less for patients at clinical stage IV than patients at lower stages and it is 0.1032 more

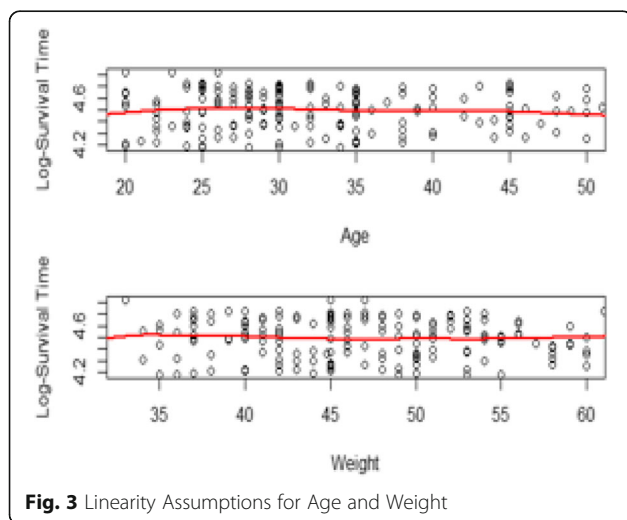


Table 3 Estimates of Shape Parameter (Q) from the Generalized Gamma AFT Model

Estimates	95%CI	
	LL	UL
0.411	-0.507	1.328

CI: Confidence Interval; LL: Lower Limit; UL: Upper Limit.

for patients with CD4 percent of 16–28% as compared to those with lower CD4 percent. In our study, gender of patients was not significantly associated with their survival at 5% level significant. In contrast, it was found to be significant according to the three studies we mentioned in this discussion. Note that, interpretations in terms of median time ratio can be obtained by exponentiation of the corresponding coefficient estimates as it is indicated for age.

Discussions

With many available statistical software packages, such as ‘survival’ and ‘flexsurv’ in R, STATA, SAS, Python, etc., AFT models have become popular in survival data analysis with a wide range of clinical and epidemiological applications. Among others, AFT models have been utilized in acute liver, cancer, and HIV AIDS studies [3, 22–25]. However, the validity of an inference depends on the realization of model assumptions. One of the assumptions in AFT models is constant variance assumption. According to Yu, et al. [18], a valid inference can be obtained from the WLSE than the classical AFT methods when the constant variance assumption is violated. Moreover, the WLSE for the AFT model is constructed operationally by the synthetic observations based on the transformation that allows the same conditional expectations as the logarithm of survival time [26] so that the transformation has no effect on the manner of interpretation for AFT model. Therefore, the degree of precision obtained from WLSE is superior to the classical methods without any difficulties imposed on the interpretation of effects. However, we relied on evidence from previous studies about the efficiency of the WLSE and utilized only magnitudes of standard errors, the width of confidence intervals, and the number of identified significant covariates as criteria of comparison. Thus, future researchers can consider more criteria than we used for comparison of the efficiency of WLSE with that of classical AFT methods.

Table 4 Information Criterion Statistics for Classical AFT Models

ICS	Exp	Weib	Logl	LN	GG
-2LL	479.7	380.6	379.9	379.8	379.2
AIC	485.7	390.6	389.9	389.8	391.2

ICS: Information Criteria Statistics; Exp: Exponential; Weib: Weibull; Logl: Log-logistic; LN: Log-Normal; GG: Generalized Gamma; LL: Log-likelihood; AIC: Akaike Information Criterion.

Table 5 Estimates from the Classical AFT and the Cox PH Models

Method	Covariates	Estimates	SE	95%CI	
				LCL	UCL
LN AFT	meanlog	4.832	0.1564	4.526	5.139
	sdlog	0.232	0.0286	0.182	0.295
	Age	-0.009	0.0029	-0.015	-0.003
	Weight	0.008	0.0032	0.002	0.014
Rank	CS (IV)	-0.278	0.0599	-0.395	-0.160
	Age	-0.009	0.0033	-0.015	-0.002
	Weight	0.008	0.0033	0.001	0.014
LSE	CS (IV)	-0.277	0.0572	-0.389	-0.165
	Age	-0.009	0.0027	-0.014	-0.004
	Weight	0.008	0.0026	0.003	0.013
Cox PH	CS (IV)	-0.272	0.0530	-0.376	-0.168
	Age	0.054	0.0177	0.019	0.089
	Weight	-0.047	0.0228	-0.092	-0.002
	CS (IV)	1.917	0.3684	1.195	2.639

Estimates: Coefficients Correspond to the log Median Survival Time for the First Three and log Hazard for the Cox PH; Estimate for Age corresponds to a year increase in Age; Estimate for Weight corresponds to a Kilogram increase in Weight; SE: Standard Error; CI: Confidence Interval; LCL: Lower Confidence Limit; UCL: Upper Confidence Limit; LN AFT: Log-Normal Accelerated Failure-Time; CS: Clinical Stage; LSE: Least Square Estimation; PH: Proportional-Hazards; meanlog: Mean of Log Survival Time; sdlog: Standard Deviation of Log Survival Time.

Conclusions and recommendations

We utilized AFT models based on the classical and WLSE methods on a real ART dataset we recently compiled with more covariates than considered in [18]. Among possible parametric AFT models, the lognormal AFT model fitted the data well. We compared the results from this model with the results from LSE and rank methods. From the classical methods, LSE was found to be the best; LN AFT the second; and rank the least. Consequently, we compared LSE with WLSE since WLSE was proved in Yu et al. [18] to be an efficient estimation. The width of confidence intervals from the WLSE was found to be shorter than that of the classical methods. The WLSE also detected more significant covariates. As a result, the WLSE performed best in providing more significant effects and precise estimates. However, the data was heteroscedastic. Thus, we recommend future researchers extend the application of WLSE to a homoscedastic real dataset with more covariates to ascertain its validity. They should utilize WLSE rather than the classical AFT methods when the homoscedasticity assumption is violated to obtain efficient estimates. Moreover, health workers should be more cautious when a patient is in advanced clinical stages, old in age, relatively lower in weight, in bedridden or ambulatory functional status, or with lower CD4 percent during ART initiation.

Abbreviations

AFS: Ambulatory Functional Status; AFT: Accelerated Failure-Time; AIC: Akaike Information Criterion; ART: Antiretroviral Therapy; BFS: Bedridden Functional Status; CD4: Cluster of Differentiation Agent 4; CI: Confidence Intervals; CS: Clinical Stage; EXP: Exponential; FS: Functional Status; GG: Generalized Gamma; ICS: Information Criterion Statistics; IID: Independently and Identically Distributed; Kg: Kilogram; LCL: Lower Confidence Limit; LL: Log-Likelihood; LN: Log-Normal; LogL: Log-logistic; LOWESS: Locally Weighted Scatterplot Smoothing; LSE: Least-Squares Estimation; NPE: Non-Parametric Estimator; PH: Proportional Hazards; SE: Standard Errors; UCL: Upper Confidence Limit; Weib: Weibull; WHO: World Health Organization; WLSE: Weighted Least-Squares Estimation

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13690-021-00617-0>.

Additional file 1.

Acknowledgements

Not Applicable.

Authors' contributions

Both authors YAM and DGC generated the idea, YAM as a corresponding author analyzed and interpreted the data, DGC contributed as a supervising author. Both authors read and approved the final manuscript.

Funding

Not Applicable.

Availability of data and materials

All data analyzed during this study are included in this published article as supplementary material with the file name "13690_2021_617_MOESM1_ESM.csv".

Declarations

Ethics approval and consent to participate

The study was cross-sectional that we made no experimental intervention beyond using secondary data from the database of University of Gondar Teaching Hospital in a retrospective manner. Thus, no consent to participate was applicable in this study. However, Ethical clearance was obtained from the College of Natural and Computational Science, University of Gondar.

Consent for publication

Not Applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Computing and Analytics Directorate, Technology and Innovation Institute of Ethiopia, Addis Ababa, Ethiopia. ²Department of Statistics, University of Pretoria, Pretoria, South Africa.

Received: 28 November 2020 Accepted: 21 May 2021

Published online: 31 May 2021

References

- Cox DR. Regression models and life-tables. *J R Stat Soc Series B*. 1972;34:187–220.
- Kalbfleisch DJ, Prentice LR. *The statistical analysis of failure time data*: Wiley; 1980.
- Khanal SP, Sreenivas V, Acharya SK. Accelerated failure time models: an application in the survival of acute liver failure patients in India. *IJSR*. 2014; 3(6):161–6.
- Collett D. *Modelling survival data in medical research*. 2nd ed. London: Chapman And Hall/CRC; 2003.
- Yu L, Liu L, Chen D-G. A homoscedasticity test for the accelerated failure time model. *Comput Stat*. 2018.
- Jin ZZ, Lin DY, Wei LJ, Ying ZL. Rank-based inference for the accelerated failure time model. *Biometrika*. 2003;90(2):341–53. <https://doi.org/10.1093/biomet/90.2.341>.
- Lai TL, Ying ZL. Rank regression methods for lefttruncated and right-censored data. *Ann Stat*. 1991b;19:531–56.
- Lai TL, Ying ZL. Linear rank statistics in regression-analysis with censored or truncated data. *J Multivar Anal*. 1992;40(1):13–45. [https://doi.org/10.1016/0047-259X\(92\)90088-W](https://doi.org/10.1016/0047-259X(92)90088-W).
- Lin DY, Ying ZL. Semiparametric inference for the accelerated life model with time-dependent covariates. *J Stat Plan Inference*. 1995;44(1):47–63. [https://doi.org/10.1016/0378-3758\(94\)00039-X](https://doi.org/10.1016/0378-3758(94)00039-X).
- Robin J, Tsiatis AA. Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika*. 1992;79:311–9.
- Tsiatis AA. Estimating regression parameters using linear rank-tests for censored-data. *Ann Stat*. 1990;18:354–72. Retrieved from <http://www.jstor.org/stable/2241547>.
- Ying ZL. A large-sample study of rank estimation for censored regression data. *Ann Stat*. 1993;21:76–99.
- Zhou M. Empirical likelihood analysis of the rank estimator for the censored accelerated failure time model. *Biometrika*. 2005;92(2):492–8. <https://doi.org/10.1093/biomet/92.2.492>.
- Buckley J, James I. Linear-regression with censored data. *Biometrika*. 1979; 66(3):429–36. <https://doi.org/10.1093/biomet/66.3.429>.
- Jin ZZ, Lin DY, Ying ZL. On leastsquares regression with censored data. *Biometrika*. 2006;93(1):147–61. <https://doi.org/10.1093/biomet/93.1.147>.
- Lai TL, Ying ZL. Large sample theory of a modified Buckley–James estimator for regression-analysis with censored-data. *Ann Stat*. 1991a;19:1370–402.
- Ritov Y. Estimation in a linear-regression model with censored-data. *Ann Stat*. 1990;18:303–28.
- Yu L, Liu L, Chen D-G. Weighted least-squares method for right-censored data in accelerated failure time model. *Int Biom Soc*. 2013;69(2):358–65. <https://doi.org/10.1111/biom.12032>.
- Cochran W. *Sampling techniques*. 3rd ed. New York: John Wiley & Sons; 1977.
- Sturges H. The choice of a class-interval. *J Am Stat Assoc*. 1926;21(153):65–6. <https://doi.org/10.1080/01621459.1926.10502161>.
- Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst*. 1959;719–48. <https://doi.org/10.1093/jnci/22.4.719>.
- Assefa T, Wencheko E. Survival analysis of patients under chronic HIV care and antiretroviral treatment. *Ethiop J Health Dev*. 2012;26(1):22–9.
- Ayalew J, Moges H, Sahu O, Worku A. Identifying factors related to the survival of AIDS patients under the follow-up of antiretroviral therapy (ART): the case of south Wollo. *Int J Data Envelop Anal *Oper Res**. 2014;1(2):21–7.
- Sieleunou I, Souleymanou M, Nenberger A-MS, Menten J, Boelaert M. Determinants of survival in AIDS patients on antiretroviral therapy in a rural Centre in the far-North Province, Cameroon. *Trop Med Int Health*. 2009;14(1): 36–43. <https://doi.org/10.1111/j.1365-3156.2008.02183.x>.
- Ghorbani N, Yazdani-Charati J, Anvari K, Ghorbani N. Application of Weibull accelerated failure time model on the disease-free survival rate of breast Cancer. *Iran J Health Sci*. 2016;4(2):11–8. <https://doi.org/10.18869/acadpub.jhs.4.2.11>.
- Delecroix M, Lopez O, Patilea V. Nonlinear censored regression using synthetic data. *Scand J Stat*. 2008;35(2):248–65. <https://doi.org/10.1111/j.1467-9469.2007.00591.x>.

Publisher's Note

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