International Journal of Data Science and Analysis
2019; 5(5): 86-91
http://www.sciencepublishinggroup.com/j/ijdsa
doi: 10.11648/j.ijdsa.20190505.12
ISSN: 2575-1883 (Print); ISSN: 2575-1891 (Online)

Research/Technical Note

Joint Survival Model of CD4 Outcome for HIV/TB Coinfected: Data from Kenya AIDS Indicator Survey

Bernadette Ikandi*, Samuel Musili Mwalili, Anthony Wanjoya

Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Email address:

bntsyomiti@gmail.com (B. Ikandi), samuel.mwalili@gmail.com (S. M. Mwalili), awanjoya@gmail.com (A. Wanjoya)

*Corresponding author

To cite this article:

Bernadette Ikandi, Samuel Musili Mwalili, Anthony Wanjoya. Joint Survival Model of CD4 Outcome for HIV/TB Coinfected: Data from Kenya AIDS Indicator Survey. *International Journal of Data Science and Analysis.* Vol. 5, No. 5, 2019, pp. 86-91.
doi: 10.11648/j.ijdsa.20190505.12

Received: September 24, 2019; **Accepted**: October 22, 2019; **Published**: October 28, 2019

Abstract: HIV infection leads to immune deficiency, increasing the risk of TB in people with HIV. HIV/TB co-infection increases the risk of death from TB or other opportunist infections. CD4 cell counts (cells/mm3) along with viral load are measures of treatment failure. This study purposed to apply shared frailty model in analyzing the survival and hazard rates of the TB/HIV co-infected persons. This work is very important because co-morbidity with TB and HIV is a rambling cause of death in Africa. The research employed a bivariate Gamma Frailty model to get the correlation amongst the HIV/TB outcomes to necessitate valid and reliable statistical inferencing. A survival frailty model on the CD4 counts is developed and fitted to factor in the unobserved heterogeneity that might occur in some observations. Ignoring some unobserved or unmeasured effects gives misguided estimates of survival. Thus, correcting these overdispersion or under-dispersion helps adjust these frailties. Frailty model provided a solid statistical analysis to CD4 data accounting for TB/HIV co-infection. The study also carried out some simulations along with the standard errors to compare the true values of the parameters. From the simulation findings, it is evident that precision and coverage improves with increase in sample size. Data used in this paper is from Kenya AIDS Indicator Survey (2012) which comprised of 648 HIV-positive patients, 10978 HIV-negative, and 2094 whose status was unknown. From the results, it is evident that the survival rate for the HIV positive individuals who are TB negative, with CD4 $\leq$ 310 is higher, at 0.9963 than that of the TB positive persons, at 0.975. The research finding points TB/HIV co-infection as a key factor for predicting immunological failure as measured by CD4 counts. The Kenyan government, and in particular the ministry of health should develop policies that mandate TB diagnosis among the PLHIV and linkage to TB treatment for the positive cases.

Keywords: HIV/TB Coinfection, CD4 Counts, Heterogeneity, Frailty Model, People Living with HIV

1. Introduction

The resurgence of tuberculosis and HIV infections is a big threat to the Kenyan population. According to Kaplan et al., human immunodeficiency virus (HIV) leads to an explosive upsurge of TB incidences, hence raising a great concern for mortality due to HIV/TB co-infections [1]. Coupled with the fact that HIV/AIDS is incurable and lowers immunity when superimposed with TB it adversely lowers survivorship [2]. Therefore, there is a dare need to use an adequate statistical model to analyze the survival rate for TB/HIV co-infected individuals.

In response to this need, shared frailty models using R program were used to model the correlation amongst the observations. HIV/TB coinfection is a significant factor in predicting immunological failure as indicated by CD4 counts [2]. Tuberculosis (TB) is an opportunistic infection which strikes more severely in persons with weak immune system than the healthier systems. According Esmail et al., HIV weakens the immunology, hence escalating the risks of suffering from TB in persons with HIV [3]. In spite of having this knowledge, there is the need to model the correlation amongst the observations for the coinfections to make valid and reliable statistical inferences.

Frailty models factor in the unobserved heterogeneity that might occur in some observations [4]. Since some observations are frailer than others in data sets, there is, therefore, the need to introduce an extra parameter to the hazard rate to account for the random frailties. Research by Gasparini et al. reveals that heterogeneity lowers the power of detecting clinically relevant treatment variances [5]. Such heterogeneities affect the reporting and interpretation of the treatment effects.

Literature Review

Sometimes, the actual survival time for an individual may not be known, more so if the event of interest has not occurred. Such cases call for specialized methods in analyzing the data. The univariate analysis describes survival data concerning the factors being investigated, whereas ignoring the effects of others [6]. Statistical models help assess simultaneously all the factors which may affect a certain diagnosis to obtain an estimation of the effects for each constituent factor.

Frailty models explain the effects of the unobserved covariates in a proportional hazard model [7]. They are an extension of the Cox PH models giving a suitable method to introduce random effects, associations, and unobserved heterogeneities. Shared frailty model models data with subjects having mutual frailty within groups or clusters. Zarulli posits that, shared frailty models are conditionally independent models with mutual frailty amongst all the subjects in a group [8]. Hence, they create dependence between the even times to bring in the aspect of shared frailty. The assumption is that the different groups are independent [9]. If frailty is zero, then the groups are independent, otherwise, there is positive dependence between the events within the groups.

Various distributions can be selected for the frailty with the widest distribution being gamma distribution. According to Wang et al., the gamma distribution is commonly used as mixture distribution [10]. Its simplicity in deriving the closed formats for survival, density, and hazard functions make it very convenient for use. For gamma distributions, big values of the variance simply a great degree of heterogeneity among the groups and a stronger association within the clusters [11]. The small variance shows independence between the group individuals.

For a gamma distribution with parameters β and θ and random effects μ_i the density function of z is given by;

$$g(z;\theta,\beta) = \frac{\theta^{\beta} z^{\beta-1}\exp(-\theta z)}{\Gamma(\beta)}$$

Other distributions that are widely used for the frailty effects are positive stable distribution, normal, lognormal, Poisson, and the inverse Gaussian distributions [11].

HIV infected persons are said to be 30 times more prone to contracting TB than HIV-free persons. Klein et al. claim that this is because the same cells holding latent TB in check (CD4 lymphocytes) are made dysfunctional by the human immunodeficiency virus [12]. CD4 T cells compose the immune response and guard against bacteria, pathogens, and viruses. The mechanisms that promote susceptibility of persons with HIV to TB are inadequately understood, being prospectively linked to multi-factorial processes [6].

2. Methodology

2.1. Introduction

A bivariate gamma frailty model was considered to the model correlation amongst the HIV/TB outcomes to necessitate valid and reliable statistical inferencing [13]. The typical Cox proportional hazard models are not suitable in analyzing the recurrent events since they account only for single occurrences of the outcome in an individual.

2.2. Bivariate Frailty Model

Assume that the bivariate random variables T_{i1} and T_{i2} with covariate vector $X_{ij} = (X_{i1}, X_{i2})$ and i=1, 2, 3,..., n individuals under study. The study will assume survival times T_1 and T_2 for each group to share the same values of the covariates. Also, assume that the frailties are multiplicative on the baseline hazard function and that the survival times of the subjects T_1 and T_2 are independent for some frailty $U_i = u_i$. the conditional hazard model for the i^{th} group at survival time $t_{ij} > 0$, for a given frailty $U_i = u_i$. The conditional hazard model for the i^{th} group at survival time $t_{ij} > 0$, for a given frailty $U_i = u_i$ becomes;

$$h(t_{i1}, t_{i2}|U_i, X_i) = u_i h_0(t_{ij}) \exp(x' ij\beta)$$

where;

Ui are the unobserved common risks factors shared by all individuals in a group j

$h_0 t_{ij}$ is the common baseline hazard function

X_i is the vector of observable covariates

β is the vector of unknown regression coefficients

The model above is termed as a shared frailty model since the individuals in the same group share the same frailty factor. It brings about correlations between the survival times of individuals within the same groups. The U_i is mutual to the subjects in the cluster and creates dependence, which is usually positive.

The conditional hazard function for the i^{th} subjects at j^{th} survival time $t_{ij} > 0$ becomes;

$$H(t_{i1}, t_{i2}|U_i, X_i) = u_i H_0(t_{ij}) \exp(x' ij\beta)$$

The conditional survival function for the i^{th} subjects becomes:

$$S(t_{i1}, t_{i2}|U_i, X_i) = \exp\{-H_0(t_{i1}, t_{i2}|U_i, X_i)\} = \exp\{-u_i H_0(t_{i1}, t_{i2}) \exp(x'_{ij}\beta)\}$$

Assuming independence, the conditional survival function for the bivariate case at survival times $t_{i1} > 0$ and $t_{i2} > 0$ becomes;

$$S(t_{i1}, t_{i2}|U_i, X_i) = S(t_{i1}|U_i, X_i)\, S(t_{i2}|U_i, X_i) = \exp\{-u_i[H_{01}(t_{i1}) + H_{02}(t_{i2})] \exp(x'_{ij}\beta)\}$$

To obtain the bivariate survival function we integrate the above equation with probability function *f (ui);*

$$S(t_{i1}, t_{i2}|X_i) = \int_{Ui}^{\cdot} \exp\{-u_i[H_{01}(t_{i1}) + H_{02}(t_{i2})] \exp(x'_{ij}\beta)\} f(u_i) du_i$$

$$= E[\exp\{-u_i[H_{01}(t_{i1}) + H_{02}(t_{i2})] \exp(x'_{ij}\beta)\}]$$

$$= Lu_i[[H_{01}(t_{i1}) + H_{02}(t_{i2})] \exp(x'_{ij}\beta)]$$

Where L(.) is the Laplace Transformation of the U distributions. Thus, the bivariate survival function has been written in terms of the Laplace transformations of the frailty distribution, obtained as the aggregate integrated conditional hazard.

2.3. Estimation in Semi-Parametric Cox PH Model

Fitting the Cox PH model will help in estimating the vector regression coefficients, β. Using the Cox approach, we will obtain the partial likelihood function for β, which is independent of h_0 (t). Partial likelihood approach makes inferences about the regression parameters in the presence of nuisance parameters h_0 (t) in the Cox proportional hazard model [14]. In this case, the study will obtain partial likelihood functions based on the Cox PH model.

Consider T_1, T_2,..., T_n, the survival time for n individuals. The ordered death rates for k individuals will be $t_{i1} < t_{i2} <... < t_{ij} <... < t_{ik}$. Let R ($t_{ij}$) be the risk before t_{ij}.

Then, the conditional likelihood that the i^{th} subject dies at time t_{ij} given that one individual from the risk set R (t_{ij}) dies at t_{ij} will be:

P (individual i dies at t_{ij}) / P (single death occurs at t_{ij}) = $\frac{h(t_{ij})}{\sum h(t_{ij})}$

$$= \frac{h_0(t_{ij})\exp(\beta' x_{ij}(t_{ij}))}{\sum h_0(t_{ij})\exp(\beta' x_{ij} t_{ij})}$$

$$= \frac{\exp(\beta' x_{ij}(t_{ij}))}{\sum \exp(\beta' x_{ij} t_{ij})}$$

The partial likelihood function for the Cox PH model will be given by;

$$L(\beta) = \prod_{i=1}^{n} \frac{\exp(\beta' x_{ij}(t_{ij}))}{\sum \exp(\beta' x_{ij} t_{ij})}$$

where $x_{ij}(t_{ij})$ is the vector of covariates for individual i who dies at t_{ij} [15]. However, the likelihood function is only for uncensored individuals. Let δ_i be the event indicator, such that;

$$\delta_{ij} = \begin{cases} 1; & t_{ij} \leq w_i \\ 0; & t_{ij} > w_i \end{cases}$$

The likelihood function in the equation above becomes;

$$L(\beta) = \prod_{i=1}^{n} \left\{\frac{\exp(\beta' x_{ij}(t_{ij}))}{\sum \exp(\beta' x_{ij} t_{ij})}\right\}^{\delta_i}$$

The partial likelihood applies when there are zero ties in the data i.e., no individuals with the same event times.

3. Results

3.1. Simulation Results

Table 1. *Summary of the Simulation Results (n=100).*

Parameter	True	Estimate	MSE	RMSE	MAE	MAPE	COVERAGE
β_1	0.405	0.397	0.021	0.146	0.117	126	94
β2	1.253	1.226	0.036	0.189	0.150	104	96

Table 2. *Summary of the Simulation Results (n=250).*

Parameter	True	Estimate	MSE	RMSE	MAE	MAPE	COVERAGE
β_1	0.405	0.407	0.008	0.090	0.070	108.6	97
β2	1.253	1.266	0.012	0.111	0.088	99.6	96

As can be seen in Table 1 the model fits well even for the sample size n=100, especially for β2. The precision and coverage improve with the increase in sample size, n = 250 (see Table 2). This is demonstrated by attenuation of MAPE values to the null (100%) after increasing the sample size.

3.2. Results of Real-data Analysis

The analysis was carried out using R-Software using various packages. The values reported were estimated under 95% confidence intervals. Table 3 presents a summary of the demographic profile of the respondents.

Table 3. *Socio-demographics and outcome characteristics of respondents.*

	HIV infected only (n=411)	TB/HIV (n=53)	Overall (n=474)
Sex			
Men	117 (28.5%)	23 (43.4%)	142 (30%)
Women	294 (71.5%)	30 (56.6%)	332 (70%)
Age Group (years)			
15-24	51 (12.4%)	0 (0%)	52 (11%)

	HIV infected only (n=411)	TB/HIV (n=53)	Overall (n=474)
25-29	78 (19%)	5 (9.4%)	85 (17.9%)
30 -39	124 (30.2%)	19 (35.8%)	146 (30.8%)
40 -49	104 (25.3%)	10 (18.9%)	117 (24.7%)
50 -59	44 (10.7%)	16 (30.2%)	61 (12.9%)
60 -64	10 (2.4%)	3 (5.7%)	13 (2.7%)
Residence			
Rural	233 (56.7%)	25 (47.2%)	265 (55.9%)
Urban	178 (43.3%)	28 (52.8%)	209 (44.1%)
Consistence condom use			
No, did not use condoms every time	226 (55%)	483 (37.7%)	253 (53.4%)
Yes, use condoms every time with the last partner	76 (18.5%)	14 (26.4%)	91 (19.2%)
Missing	109 (26.5%)	19 (35.8%)	130 (27.4%)
CD4			
Mean (SD)	557 (369)	483 (289)	550 (364)
Median (Min, Max)	496 [7.00, 2610)	437 [34.0, 1340]	491 [7.00, 2610]
Log viral load			
Mean (SD)	4.51 (0.767)	4.65 (1.04)	4.52 (0.793)
Median [Min, Max]	4.54 [2.74, 6.69]	4.98 [2.89, 6.10]	4.57 [2.74, 6.69]
Missing	137 (33.3%)	33 (62.3%)	174 (36.7%)

Among the HIV-infected persons, 42.5% had CD4 counts per microliter ≤ 350 counts per microliter, while 18.42% had ≤ 200 CD4 cells. Thus, 57.5% of the HIV-infected met the CD4 count threshold for Antiretroviral Therapy (ART) initiations. It was evident that, of the 648 HIV-infected persons, the average count of the CD4 cells was 455.7.

The Kaplan Meier CD4 curve was obtained as shown below: HIV-TB Positive In this category, results were obtained for two TB cases: TB-uninfected and TB-infected.

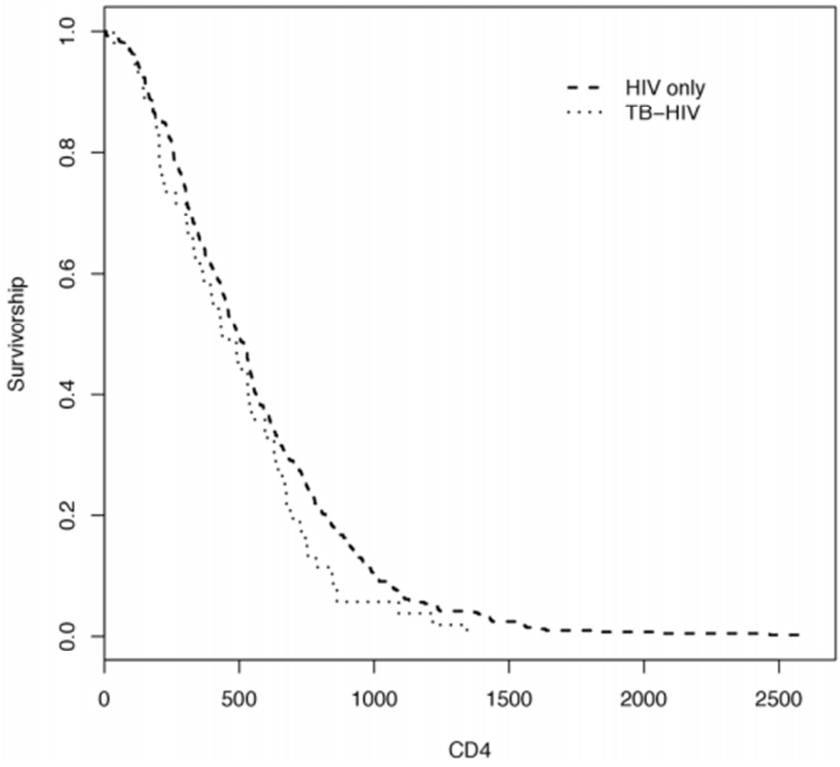

Figure 1. *Kaplan Meier CD4 curve.*

The best model was selected using the Akaike Information Criterion (AIC). As seen in Table 4 the best model has Sex only as of the covariate with the smallest AIC of 4450. From the Cox regression model results where we included the interaction term with HIV TB status, the model shows that HIV only as compared TB-HIV have had higher CD4 cells count.

Table 4. *Model Selection using AIC.*

Model	AIC
Sex	4450
Age Group	4462
Residence	4455
Sex + Age Group	4459
Sex + Residence	4452
Sex + Age Group + Residence	4461

A summary of the survival for the HIV-infected persons who were also TB infected was obtained as shown in Table 5. Likewise, a summary of the survival for the HIV-TB infected people was also obtained as illustrated by Figure 1.

Table 5. *Parameter estimates from the Cox regression model stratified for TB-HIV status and Shared Frailty Model.*

Model	Variable	HR	95% CI Lower	95% CI Upper
Cox	Sex (Women)	0.782	0.6381	0.9534
Shared	Sex (Women)	0.794	0.6517	0.986
	δ	2.583		
	ρ	-0.427		

From the results, it is evident that the survival rate for the HIV positive individuals who are TB negative, with CD4≤ 310 is higher, at 0.9963 than that of the TB positive persons, at 0.975. As shown in Table 5, the regression coefficients of sex from the Cox regression model stratified by TB-HIV, and that of the shared frailty model is generally in the same direction. It the hazard ratio (HR) show that men have higher CD4 cell counts than females.

4. Discussions

The main method applied in this research project is the shared frailty survival model. Shared frailty model, incorporates the unobserved frailty, are random effect models for time variables with the frailty having multiplicative effects on the hazard. The frailty being an unobservable random variable varying over the sample balances the individual risk by safeguarding against underestimation or overestimation of the parameters. This model strengthens the accurate measures of covariates effects.

Most epidemiology or biomedical research studies have applied Cox proportional hazard model. The hazard function in this model is a product of baseline hazard function and the exponential to linear predictor function of time-independent covariates. Even though this model does not consider frailties, it provides robust and easy to interpret hazard ratios.

We performed a simulation study to assess the precision and coverage of the shared frailty model. Results from the simulation study indicated that the precision of the model improved with the sample size. The key finding the simulation study was that the model maintained the 95% coverage of all the parameter even for the smaller sample size. This demonstrated the flexibility and appropriateness of the model in modelling survival data.

In this project, we applied a shared frailty model to the national representative complex survey data, namely, the second Kenya AIDS indicator survey. Tough many explanatory variables were considered only sex was shown to the key predictor of the CD4 cell counts. This was based on the model with the smallest Akaike Information Criterion (AIC). The AIC is an estimator of the relative quality of statistical models for a given set of data. Prior studies have founded that TB/HIV co-infection severely impact on young women, more so those in low-income settings. This agrees with the findings from this research. The results from this research indicated that women had lower CD4 counts than males, which could be a result of a higher burden on women in terms of health-seeking behavior. A different research also agrees that TB is a number one infectious disease killer amongst women than in men, hence it is imperative that women to undergo screening whenever they experience the first signs of TB. They attribute the disparity to the fact that females have lower body mass indices, low hemoglobin levels, and high macroglobulin levels than males.

5. Conclusion

In this work, we presented an approach to model immunological markers with an example of CD4 count using a shared frailty survival model. This provided a solid statistical analysis to CD4 data accounting for TB/HIV co-infection. This is in contrast to traditional transmission dynamics cellular models that defined HIV and TB coinfections in an individual. This work is very important because co-morbidity with TB and HIV is a rambling cause of death in Africa. Through the research, we found out that TB/HIV co-infection is a key factor for predicting immunological failure as measured by CD4 counts. This is a clear pointer for urgent and immediate action by policymakers in handling TB scenarios amongst persons living with HIV (PLHIV). The Kenyan government, and more so the ministry of health (MOH) should develop policies that mandate TB diagnosis among the PLHIV and linkage to TB treatment for the positive cases.

References

[1] Kaplan, R., Hermans, S., Caldwell, J., Jennings, K., Bekker, L. G., & Wood, R. (2018). HIV and TB co-infection in the ART era: CD4 count distributions and TB case fatality in Cape Town. BMC infectious diseases, 18 (1), 356.

[2] Pinto, C. M., & Carvalho, A. R. (2017). The HIV/TB coinfection severity in the presence of TB multi-drug resistant strains. Ecological complexity, 32, 1-20.

[3] Esmail, H., Riou, C., du Bruyn, E., Lai, R. P. J., Harley, Y. X., Meintjes, G.,... & Wilkinson, R. J. (2018). The immune response to Mycobacterium tuberculosis in HIV-1-coinfected persons. Annual review of immunology, 36, 603-638.

[4] Gleiss, A., Gnant, M., & Schemper, M. (2018). Explained variation in shared frailty models. Statistics in medicine, 37 (9), 1482-1490.

[5] Gasparini, A., Clements, M. S., Abrams, K. R., & Crowther, M. J. (2018). Impact of model misspecification in shared frailty survival models. arXiv preprint arXiv: 1810.08140.

[6] Emura, T., Matsui, S., & Rondeau, V. (2019). Survival Analysis with Correlated Endpoints: Joint Frailty-Copula Models. Springer Singapore.

[7] Colchero, F., & Kiyakoglu, B. Y. (2019). Beyond the proportional frailty model: Bayesian estimation of individual heterogeneity on mortality parameters. Biometrical Journal.

[8] Zarulli, V. (2016). Unobserved heterogeneity of frailty in the analysis of socioeconomic differences in health and mortality. European Journal of Population, 32 (1), 55-72.

[9] Pitacco, E. (2019). Heterogeneity in mortality: a survey with an actuarial focus. European Actuarial Journal, 1-28.

[10] Wang, A., Chandra, K., & Jia, X. (2018). The analysis of left truncated bivariate data using frailty models. Scandinavian Journal of Statistics, 45 (4), 847-860.

[11] Martins, A., Aerts, M., Hens, N., Wienke, A., & Abrams, S. (2018). Correlated gamma frailty models for bivariate survival time data. Statistical methods in medical research, 0962280218803127.

[12] Klein, J. P., van Houwelingen, H. C., Ibrahim, J. G., & Scheike, T. H. (2016). Frailty Models. In Handbook of Survival Analysis (pp. 461-477). Chapman and Hall/CRC.

[13] Van den Berg, G. J., & Drepper, B. (2016). Inference for shared-frailty survival models with left-truncated data. Econometric Reviews, 35 (6), 1075-1098.

[14] Rondeau, V., Gonzalez, J. R., Mazroui, Y., Mauguen, A., Diakite, A., Laurent, A.,... & Sofeu, C. L. (2019). frailtypack-package: General Frailty models: shared, joint and nested frailty.

[15] Gupta, R. C. (2018). Association measures in the bivariate correlated frailty model. REVSTAT–Statistical Journal, 16 (2), 257-278.