Modeling of a Cure Rate Model for TB with HIV Co-Infection.

Oluwafemi Samson Balogun, M.Sc.[1*] and Prof. Emmanuel Teju Jolayemi[2]

[1]Department of Statistics and Operations Research, Modibbo Adama University of Technology, Yola, Adamawa State, Nigeria.
[2]Department of Statistics, University of Ilorin, Ilorin, Kwara State, Nigeria.

Email: balogun.os@mautech.edu.ng *

ABSTRACT

Several investigations allow the incorporation of cured individuals in the analysis for chronic diseases such as cancer, etc., but this research has been motivated by the fact that infectious diseases with possible co-infection do exist and not much research has been done in this area of epidemiology modeling. We therefore use a survival model that incorporates the cured patients in the analysis, which is called a cure rate model. In this research, we extended the model used by Chen *et al.,* 1999; Uddin *et al.,* 2006; Aljawadi *et al.,* 2011b to estimate the proportion of patients cured in cases of left censored data using exponential distribution based on the Bounded Cumulative Hazard (BCH). The analysis provided the analytical solution and simulation study for the cure rate parameter using R package.

(Key terms: epidemiology, cure fraction, BCH model, bounded cumulative hazard, left censored, EM algorithm)

INTRODUCTION

Tuberculosis (TB) is very prevalent in the Nigerian setting. This accounts for the fact that the Nigerian Government has declared that its management be free. In recent times, however, the achievement recorded in the TB management has drastically slowed down.

TB and Human Immunodeficiency Virus/ Acquired Immunodeficiency Syndrome (HIV/AIDS) are inextricably linked through complex biological and social pathways. The two diseases accelerate each other's progression and perpetuate synergistic rates of morbidity and mortality (Williams, 2016). While HIV/AIDS and TB can individually be the major causes for concern as stand-alone public health threats, the combination of the two diseases has proven to have a far greater impact on the epidemiologic progression and consequently on the impact it has on the global health scene. The dual infection has been termed "accursed duet" (Jaiswal *et al.,* 2012).

The escalating global burden of co-infection has prompted public health institutions to consider an alternate, integrated of healthcare. In place of vertical health programs for infectious disease control, the World Health Organization (WHO) has prioritized integration of TB and HIV healthcare services as a means to decrease the incidence and impact of both infections.

Integrated strategies involve improvement in TB and HIV testing, prevention, treatment, and supportive services through coordinated and collaborative efforts between individual TB and HIV programs (WHO, 2004). However, countries experiencing the highest burden of co-infection are also the most resource poor. Planning and execution of informed integrated interventions in these setting is, at once, imperative and onerous as a result of multi-level challenges and deficiencies (WHO, 2008a).

Central to this is the phenomenological experience of illness, and of healthcare, that are relevant from the perspective of infected individuals, particularly in relation to their social, cultural, political, historical, and economic contexts. Unless we understand how individuals live with TB and HIV and how this relates to their decisions for TB and HIV care, we might fail to devise a successful and sustained public health response to curb the dual epidemic.

Infectious diseases (like measles, hepatitis, TB, HIV/AIDS, etc.) have continued to cause large outbreaks of health failures all over the world, even in countries that have achieved high

vaccination due to awareness and early detection of the chronic bacteria/virus. Most infectious diseases have been discovered to have a possible co-infection (TB patients co-infected with HIV/AIDS, TB co-infected with malaria and hepatitis co-infected with jaundice, etc.). Most often, some people also adhere strictly to self-medication at home completely unsupervised and this has contributed to the high prevalence of the diseases. Most people have attributed the main causes of the prevalence to lack of proper education about the diseases and low detection rates at early stages. Financial constraints are also a major factor as most people prefer the treatment of the disease locally or traditionally than exposing themselves to hospitals for it treatment.

TB is transferable from parents to children with the highest rate of mortality especially in developing countries. Attempts to understand the epidemiology of TB with advent of HIV started as far back as early 1980s. TB management in the presence of HIV is complicated by several factors. The large pill burden could be confusing to the patient, association of the drugs to meals could also be a challenge, while the side effects may impede adherence. The challenges of managing drug interactions and overlapping toxicities when treating co-infected patients makes it imperative that treatment of such patients are coordinated between the TB and HIV care providers to ensure optimal treatment for both diseases. It is important that health workers are on alert for interactions between these two diseases hence the need to learn how to manage conditions arising from them competently.

Hence, this study is been motivated by the fact that while infectious diseases with possible co-infection do exist in Nigeria, not much research has been done in this area of its epidemiology modeling. The aim of this research is therefore to extend an existing model suitable to help us manage the disease and its possible co-infections.

With rapid developments in the fields of medical and health science, we now encounter more survival studies where some patients are expected to be cured. Survival models that give accounts for the cure are important for understanding prognosis in potentially terminal diseases. Traditional parametric survival model such as Weibull or Gamma (Cox and Oakes, 1984) do not account for the probability of cure.

Cure rate is a survival model that incorporate the cured fraction (non-susceptible) and the uncured fraction (susceptible). The main interest of the model is to determine the proportion of the cured patients, the survival distribution (or, failure time distribution) of the uncured individuals and the effects of the treatment on the survival (Datta, 2013).

Recently, much attention has been devoted to formulating parametric survival models incorporating a cure fraction a non-zero tail probability of the survival function. These have focused on cancer-relapse trials, including breast cancer, non-Hodgkin lymphoma, leukaemia, prostate cancer, melanoma, and head and neck cancer and from other chronic diseases have used the cure rate model to analyze such data, where due to recent advances in therapy and treatment, a significant proportion of patients are expected to be "cured", that is to remain disease free even after really long follow-ups (Gu *et al.,* 2010).

Cure rate incorporating a cured fraction is defined as a non-zero tail-probability of the survival function, adjust for this feature of the data and date back to the mixture model by Berkson and Gage (1952) (mixture-cure model, in short) and has been extensively discussed by several authors, including Farewell (1982), Farewell (1986), Gray and Tsiatis (1989), Maller and Zhou (1996), Ewell and Ibrahim (1997) and Stangl and Greenhouse (1998).

In this model, we assume that a certain fraction π of the population is "cured" and the remaining $1-\pi$ are not cured. In the model, the survivor function for the entire population is given by $S_1(t)$, where:

$$S_1(t) = \pi + (1-\pi) S^*(t) \qquad (1)$$

Such that $S^*(t)$ denotes the survival function for the non-cured group in the population. Common choices for $S^*(t)$ are the exponential and Weibull distributions. The model (1) is called the standard cure rate model.

In this research, the aim is to use an analytic solution and a simulation study to extend the works of (Chen *et al.,* 1999; Uddin *et al.,* 2006; Aljawadi *et al.,* 2011b) to develop a model suitable to help us model the cure rate model for a disease with its possible co-infection.

In Daftary (2011), the study shows integration effort should consider the social contexts of TB/HIV co-infection, social consequence of patients health decisions and paradigms within which such efforts are set in the design and execution of successful intervention.

In Janida *et al.,* (2015), the study showed Abuja has one of the highest proportions of TB-HIV co-infection rates in Sub-Saharan Africa and it also shows the outcome of patients with statistically significant higher mortality.

This model has been applied in many areas such as health, criminology, reliability, and economics. For example, Freireich *et al.,* (1963) studied leukemia patients and compared the survival experiences of the patients given them the 6-MP compared to a control group. It was discovered that all patients in the control group experienced symptoms of the disease during the period of the study, while more than half of the patients from the treatment group did not exhibit any signs of remission, thus resulting in a large number of immunes.

Maller and Zhou (1996) used a study on recidivism times of prisoners released from prisons in Western Australia. The data revealed that a significant proportion of the prisoners were likely to return to prison. The result of the analysis from the study indicated that the prison programs and other factors such as the age, job status and marital status have an effect on the recidivism rates.

Struthers and Farewell (1989) the progression of AIDS in HIV –positive individuals was modeled with cure rate model. It was noted that the model that allowed the proportion of immunes provided a better fit to the data used than the model which did not account for long-term survivor.

In reliability, Nelson (1982) in his study observed the life of insulation on electric motors which were operated at various levels of temperature. He discovered that the motors lasted almost indefinitely when operated under low temperature, and it broke down quickly at higher temperature. He applied mixture model to capture the immune components in his study.

According to Barriga *et al.,* (2015), the authors studied the time until the event of default on a Brazilian personal loan portfolio and where the authors let N follow a geometric distribution and be a cumulative density function of the inverse Weibull distribution.

Also in the area of credit risk modelling in Oliveira and Louzada (2014b) the authors applied the model given by Chen *et al.,* (1999) to study the time span to full recovery of non-performing loans in a portfolio of personal loans. The authors compared the parameter obtained from two follow-up studies of a set of non-performing loans. The first follow-up is related to the time span to the default occurrence, while the second is related defaulted loans. The authors found a significant relationship between the default and recovery processes. They suggested in the article that in times of higher risk of default, it is also likely to have a decrease in the recovery rates of non-performing loans.

Yu *et al.,* (2004) mixture cure rate model was used for group survival data and they observed that the estimate of the cure fraction can be quite sensitive to the length of follow up time and the choice of latency distribution (failure time distribution). They also investigated the effects of various distributions such as the lognormal, log-logistic, Weibull and generalized gamma and they concluded that the estimate of the cure fraction was robust with the generalized gamma distribution.

In Chen *et al.,* (1999), the authors proposed a new Bayesian model for survival data with surviving fraction. The model has a proportional hazards structure with the cure rate depending naturally on covariates. One main difference between the Bayesian approach and the cure rate approach is that the former models the entire population as a proportional hazards model while in the latter model only the non-cured group has proportional hazard model and both model can be obtained from one another. The Bayesian model as suggested by the authors is computational attractive.

In Li *et al.,* (2016) proposed flexible cure rate models in analysing univariate right-censored based on the assumption that the logarithm of survival time follows generalized extreme value

(GEV) distribution with spatial and non-linear covariate effects. The model proposed is very flexible but relatively complex. In practice, it is useful to have the methods for the data screening so that the necessity of introducing the proposed components in the model can be examined before actually fitting the full model.

We explored the need for a more flexible model by comparing the fitted survival functions under the usual Weibull model and Kaplan Meir method in one of the Figure used in their current study. Future research need to investigate the data characteristics that require inclusion of the proposed modelling component all at once. The extension to the latent activation model (Cooner *et al.,* 2007) can further improve the flexibility of the cure rate model under different or even unknown medical mechanism. Also, the modelling approaches used in the study can be easily applied to a multivariate setting, such as the spatial survival data with multiple causes of death.

One extension of their work could be to explore possible latent risk factor distribution other than Poisson, such as negative binomial distribution (Cancho *et al.,* 2011) or the geometric distribution. A theoretical challenge is the posterior propriety under improper priors on the cure rate using the GEV distribution. In their study, they impose no constraint on the range of the scale or the shape parameter in the GEV distribution. Posterior propriety can be achieved with limited parameter values in the one-parameter GEV distribution in the survival analysis (Roy *et al.,* 2013; Roy and Dey, 2014). Further investigations are needed to set conditions for the posterior propriety with their three-parameter GEV distribution in the cure rate model.

Chen (2016) in his work concluded that parametric mixture cure rate models possess the flexibility to accommodate varying treatment effects introduced by therapies with different mechanisms of actions. The proposed method provides immuno-oncology clinical trial researchers with a useful tool for continuous event monitoring and prediction of analysis times, such that informed decisions with quantifiable risks can be made for better resource and logistic planning. It will also ensure that efficacious treatments are made available to patients in a timely manner.

Uddin *et al.,* (2006) deals with the analysis of cure rate estimation under uncensored data. They propose cure model using maximum likelihood method (MLE). The analysis from the study showed that the cure rate estimator converges to the true parameter when considered for both cured and non-cured group.

The analysis also shows that the estimating equation converges to the true equation of the parameter when only the uncured group is considered. Aljawadi *et al.,* (2011c) propose an analytical approach for parametric estimation of the cure fraction in cancer clinical trials based on the bounded cumulative hazard (BCH) model with covariates involved in the data set. The analysis is constructed by means of the exponential distribution in the case of left censoring and within the framework of the expectation maximization (EM) algorithm.

The analysis provided the analytical solution and a simulation study for the cure rate parameter. The results demonstrate that cure fraction estimation based on the proposed procedure was more attractive when censoring rate is low than when it is high.

MATERIAL AND METHOD USED

In this work, we propose an extension to the standard cure rate model proposed by (Chen *et al.,* 1999; Uddin *et al.,* 2006; Aljawadi *et al.,* 2011b) suitable to manage a disease with co-infection in Nigeria.

Assumptions of the Model

1. The model is derived from a natural biological motivation
2. It has proportional hazards structure through the cure rate parameter, and thus has an appealing interpretation
3. It is computationally attractive. In particular, by introducing latent variables, we are able to efficiently sample from the posterior distribution of the parameters.
4. The model has a mathematical relationship with the standard cure rate. Specifically, it was shown that any standard cure rate model can be written as the proposed model and vice versa.

The BCH Model

For individual in the population, let N denote the unobservable number of causes of the event of interest for this individual. The time for the i-th cause to produce the event of interest is denoted by $Z_i, i=1,2,...,N$. We assume that cumulative density function (cdf) $F(z)$ and Survival function $S(z)=1-F(z)$. We also assume that N is independent of $Z_1, Z_2, ...$ the observable time to event is defined by $T=\min[Z_1, Z_2..., Z_N], if\ N\geq 1 and\ T=\infty if\ N=0, with\ P(T=\infty / N=0)=1$ under this step, the survival function for the population is given:

$$\begin{aligned} S_{pop}(t) &= P(N=0)+P(Z_1>t,...Z_N>t / N\geq 1) \\ &= \exp(-\theta)+\sum_{k=1}^{\infty} S(t)^k \frac{\theta^k}{k!}\exp(-\theta) \\ &= \exp(-\theta+\theta S(t)) \\ &= \exp(-\theta F(t)) \end{aligned} \qquad (2)$$

Since $S(\infty)=\exp(-\theta) and\ F(\infty)=1$ then Equation 2 is an improper survival function. Therefore, the fraction π can be defined as follows:

$$\pi = S(\infty)=P(N=0)=\exp(-\theta) \qquad (3)$$

As $\theta\to\infty, \pi\to 0, whereas\ as\ \theta\to 0, \pi\to 1 (i.e., 0\leq\pi\leq 1)$. It should be notified that the first derivative of $S(t)$ with respect to t is:

$$\frac{\partial S}{\partial t}=\theta f(t)\exp(-\theta F(t))$$

Since $1-S(t)=F(t)$ and accordingly:

$$-\frac{\partial S}{\partial t}=f(t)$$

Then $\partial S/\partial t$ is an improper survival function and therefore $f(t)$ is an improper probability density function as well. The standard cure rate model becomes:

$$S(t)=\exp(-\theta)+(1-\exp(-\theta))S^*(t) \qquad (4)$$

The Likelihood Function for Proposed Model

The likelihood function is considered using the left censoring type of input data, in order to analyze such data, let α_i be the indicator of left censoring , c_i be a cure indicator and d_i be the disease status where for the i^{th} patient:

$$\alpha_i = \begin{cases} 0: censored \\ 1: otherwise \end{cases} \qquad c_i = \begin{cases} 0: cured \\ 1: otherwise \end{cases} \; and \; d_i = \begin{cases} 0: TB \\ 1: TB/HIV \end{cases}$$

If α_i=1, then $c_i = 1, but\ if\ \alpha_i = 0, then\ c_i$ is not observed and it can be either one or zero, assuming that censoring is independent of failure times.

Suppose that T is a random variable with probability density function $f(t;\theta)$, θ is to be estimated and $t_1, t_2, ..., t_n$ is a random sample of size n, then the joint probability density function is given as:

$$L(t_1, t_2, ..., t_n; \theta) = \prod_{i=1}^{n} f(t_i; \theta) \tag{5}$$

In parametric maximum likelihood estimation of both the cumulative distribution $F(.)$ and the probability density function $f(.)$ for the entire population are known. Consequently, the complete log likelihood function is:

$$l_c = \log \prod_{i=1}^{n} \left[\left\{ f_u(t_i)(1-\pi) \right\}^{c_i} \right]^{\alpha_i} \left[\{\pi\}^{1-c_i} \left\{ (1-\pi)(1-S_u(t)) \right\}^{c_i} \right]^{1-\alpha_i} \tag{6}$$

The study employs the exponential distribution for $S_u(t)\ and\ f_u(t)$ such that:

$S_u(t) = \ell^{-\lambda t}$ *and* $f_u(t) = \lambda \ell^{-\lambda t}$

However, in the case of left censoring the survival function of the uncured patients becomes

$S_u(t) = 1 - \ell^{-\lambda t}$

Where $S_u(t)\ and\ f_u(t)$ are the p.d.f. and the survival function for the uncured patients, respectively. Therefore, the log likelihood function becomes:

$$l_c = \log \prod_{i=1}^{n} \left[\left\{ \lambda \ell^{-\lambda t} \left(1 - \ell^{-\theta}\right) \right\}^{c_i} \right]^{\alpha_i} \left[\left\{ \ell^{-\theta} \right\}^{1-c_i} \left\{ \left(1 - \ell^{-\theta}\right)\left(1 - \ell^{-\lambda t}\right) \right\}^{c_i} \right]^{1-\alpha_i}$$

Considering a situation where we have a co-infective disease, we introduce $d_i = 0\ for\ TB\ and\ d_i = 1\ for\ TB/HIV$ into the likelihood function to have:

$$= (1-d_i) \sum_{i=1}^{m_1} \left[\left\{ \lambda \ell^{-\lambda t} \left(1 - \ell^{-\theta}\right) \right\}^{c_i} \right]^{\alpha_i} \left[\left\{ \ell^{-\theta} \right\}^{1-c_i} \left\{ \left(1 - \ell^{-\theta}\right)\left(1 - \ell^{-\lambda t}\right) \right\}^{c_i} \right]^{1-\alpha_i} +$$

$$(d_i) \sum_{i=1}^{m_2} \left[\left\{ \lambda \ell^{-\lambda t} \left(1 - \ell^{-\theta}\right) \right\}^{c_i} \right]^{\alpha_i} \left[\left\{ \ell^{-\theta} \right\}^{1-c_i} \left\{ \left(1 - \ell^{-\theta}\right)\left(1 - \ell^{-\lambda t}\right) \right\}^{c_i} \right]^{1-\alpha_i}$$

$$= \sum_{i=1}^{m_1} (1-d_i) c_i \alpha_i \left[\log \lambda - (\lambda t_i) + \log\left(1 - \ell^{-\theta}\right) \right] + \sum_{i=1}^{m_1} (1-d_i)(1-c_i) \left[(1-c_i)(-\theta) + c_i \left\{ \log\left(1 - \ell^{-\theta}\right) + \log\left(1 - \ell^{-\lambda t_i}\right) \right\} \right]$$

$$+\sum_{i=1}^{m_2} d_i c_i \alpha_i \left[\log\lambda - (\lambda t_i) + \log\left(1-\ell^{-\theta}\right)\right] + \sum_{i=1}^{m_2} (d_i)(1-c_i)\left[(1-c_i)(-\theta) + c_i\left\{\log\left(1-\ell^{-\theta}\right) + \log\left(1-\ell^{-\lambda t_i}\right)\right\}\right]$$

$$= \sum_{i=1}^{m_1}(1-d_i)c_i\alpha_i \log\lambda - \sum_{i=1}^{m_1}(1-d_i)c_i\alpha_i\lambda t_i + \sum_{i=1}^{m_1}(1-d_i)c_i\alpha_i \log\left(1-\ell^{-\theta}\right) - \theta\sum_{i=1}^{m_1}(1-d_i)(1-c_i)(1-\alpha_i)$$

$$+\sum_{i=1}^{m_1}(1-d_i)(1-\alpha_i)c_i \log\left(1-\ell^{-\theta}\right) + \sum_{i=1}^{m_1}(1-d_i)(1-\alpha_i)\log\left(1-\ell^{-\lambda t_i}\right) + \sum_{i=1}^{m_2} d_i c_i \alpha_i \log\lambda - \sum_{i=1}^{m_2} d_i c_i \alpha_i \lambda t_i$$

$$+\sum_{i=1}^{m_2} d_i c_i \alpha_i \log\left(1-\ell^{-\theta}\right) - \theta\sum_{i=1}^{m_2} d_i (1-c_i)(1-\alpha_i) + \sum_{i=1}^{m_2} d_i (1-\alpha_i) c_i \log\left(1-\ell^{-\theta}\right) + \sum_{i=1}^{m_2} d_i (1-\alpha_i)\log\left(1-\ell^{-\lambda t_i}\right)$$

$$= \sum_{i=1}^{m_1}(1-d_i)c_i\alpha_i \log\lambda - \sum_{i=1}^{m_1}(1-d_i)c_i\alpha_i\lambda t_i + \sum_{i=1}^{m_1}(1-d_i)c_i \log\left(1-\ell^{-\theta}\right) - \theta\sum_{i=1}^{m_1}(1-d_i)(1-c_i)(1-\alpha_i)$$

$$+\sum_{i=1}^{m_1}(1-d_i)(1-\alpha_i)\log\left(1-\ell^{-\lambda t_i}\right) + \sum_{i=1}^{m_2} d_i c_i \alpha_i \log\lambda - \sum_{i=1}^{m_2} d_i c_i \alpha_i \lambda t_i + \sum_{i=1}^{m_2} d_i c_i \log\left(1-\ell^{-\theta}\right) - \theta\sum_{i=1}^{m_2} d_i (1-c_i)(1-\alpha_i)$$

$$+\sum_{i=1}^{m_2} d_i (1-\alpha_i)\log\left(1-\ell^{-\lambda t_i}\right) \qquad (7)$$

The solution of $\frac{\partial l_c}{\partial\theta} = 0$ *and* $\frac{\partial l_c}{\partial\lambda} = 0$ are the desired estimates of θ *and* λ where:

$$\frac{\partial l_c}{\partial\theta} = \left[(1-\alpha_i)(1-c_i)\left\{\sum_{i=1}^{m_1}(1-d_i) + \sum_{i=1}^{m_2}(d_i)\right\}\right] + \frac{1}{\left(1-\ell^{-\theta}\right)}\left[c_i\left\{\sum_{i=1}^{m_1}(1-d_i) + \sum_{i=1}^{m_2}(d_i)\right\}\right] \qquad (8)$$

$$\frac{\partial l_c}{\partial\lambda} = \frac{\left[c_i\alpha_i\left\{\sum_{i=1}^{m_1}(1-d_i) + \sum_{i=1}^{m_2}(d_i)\right\}\right]}{\lambda} - c_i\alpha_i t_i\left\{\sum_{i=1}^{m_1}(1-d_i) + \sum_{i=1}^{m_2}(d_i)\right\} +$$

$$c_i(1-\alpha_i)\left\{\sum_{i=1}^{m_1}(1-d_i) + \sum_{i=1}^{m_2}(d_i)\right\}\left(\frac{t_i}{\ell^{-\lambda t_i}-1}\right) \qquad (9)$$

solving Eqn. 8 *implies*:

$$\theta = \log\left[\frac{\sum_{i=1}^{m_1}(1-d_i)c_i + \sum_{i=1}^{m_2}(d_i)c_i}{\sum(1-d_i)(1-\alpha_i)(1-c_i) + \sum(d_i)(1-\alpha_i)(1-c_i)} + 1\right] \qquad (10)$$

$$g_i = \alpha_i + (1-\alpha_i)\left[\frac{\left(1-\ell^{-\theta}\right)S_u(t_i)}{\left(\ell^{-\theta}\right)S_u(t_i)}\right]$$

for censored individuals $\alpha_i = 0$ and hence the g_i can be re-written as follows:

$$g_i = \left[\frac{\left(1-\ell^{-\theta}\right) S_u\left(t_i\right)}{\ell^{-\theta} + \left(1-\ell^{-\theta}\right) S_u\left(t_i\right)} \right]$$

$$g_i = \left[\frac{\left(1-\ell^{-\theta}\right)\left(1-\ell^{-\lambda t_i}\right)}{\ell^{-\theta} + \left(1-\ell^{-\theta}\right)\left(1-\ell^{-\lambda t_i}\right)} \right]$$

for simplicity, let p_i to be the probability of cured patients such that $p_i = E\left(1-c_i\right) = 1-g_i$

$$= 1 - \left[\frac{\left(1-\ell^{-\theta}\right)\left(1-\ell^{-\lambda t_i}\right)}{\ell^{-\theta} + \left(1-\ell^{-\theta}\right)\left(1-\ell^{-\lambda t_i}\right)} \right]$$

$$= \left[\frac{1}{1+\left(\ell^{-\theta}-1\right)\left(1-\ell^{-\lambda t_i}\right)} \right]$$

EM Algorithm

From Equation **7**, for $d_i = 0$ *for TB* , $d_i = 1$ *for TB / HIV* and $n = m_1$ *or* m_2

$$= \sum_{i=1}^{m} c_i \alpha_i \log\lambda - \sum_{i=1}^{m} c_i \alpha_i \lambda t_i + \sum_{i=m+1}^{n} c_i \log\left(1-\ell^{-\theta}\right) - \theta \sum_{i=m+1}^{n} \left(1-\alpha_i\right)\left(1-c_i\right) + \sum_{i=m+1}^{n} \left(1-\alpha_i\right)\log\left(1-\ell^{-\lambda t_i}\right)$$

$$E\left(l_c / \alpha_i, c_i, t_i\right) = m\log\lambda - \lambda\sum_{i=1}^{m} t_i - \theta \sum_{i=m+1}^{n} \left(1-c_i\right) + m\log\left(1-\ell^{-\theta}\right) + \log\left(1-\ell^{-\theta}\right) \sum_{i=m+1}^{n} c_i + \sum c_i \log\left(1-\ell^{-\lambda t_i}\right)$$

$$\sum_{i=m+1}^{n} \left(1-c_i\right), \sum_{i=m+1}^{n} c_i, \sum_{i=m+1}^{n} c_i \log\left(1-\ell^{-\lambda t_i}\right)$$

are the sufficient statistic for the parameters vector $\left(\lambda, \theta\right)^T$. It follows that the log-likelihood based on complete data is linear incomplete data sufficient statistic and E step requires the computation of:

$$E_{\lambda,\theta}\left[\sum_{i=m+1}^{n} \left(1-c_i\right)\right],\ E_{\lambda,\theta}\left[\sum_{i=m+1}^{n} c_i\right] \text{ and } E_{\lambda,\theta}\left[\sum_{i=m+1}^{n} c_i \log\left(1-\ell^{-\lambda t_i}\right)\right]$$

Let

$$S_1 = E_{\lambda,\theta}\left[\sum_{i=m+1}^{n}\left(1-c_i\right)\right] = \left(n-m\right)p_i$$

$$= \left(n-m\right)\left[\frac{1}{1+\left(\ell^{-\theta}-1\right)\left(1-\ell^{-\lambda t_i}\right)}\right] \tag{11}$$

$$S_2 = E_{\lambda,\theta}\left[\sum_{i=m+1}^{n} c_i\right] = \sum_{i=m+1}^{n}\left(1-p_i\right)$$

$$= \sum_{i=m+1}^{n}\left[1-\frac{1}{1+\left(\ell^{-\theta}-1\right)\left(1-\ell^{-\lambda t_i}\right)}\right] \tag{12}$$

$$S_3 = E_{\lambda,\theta}\left[\sum_{i=m+1}^{n} c_i \log\left(1-\ell^{-\lambda t_i}\right)\right] = \sum_{i=m+1}^{n}\left(1-p_i\right)t_i$$

$$= \sum_{i=m+1}^{n} c_i\left[1-\frac{1}{1+\left(\ell^{-\theta}-1\right)\left(1-\ell^{-\lambda t_i}\right)}\right]t_i \tag{13}$$

For the M step we can use the complete data maximum likelihood estimate of $\left(\lambda,\theta\right)^T$ given Eqn **3** and **4** and then substituting the expectations derived in the E-step for the complete data sufficient statistic such that on the sufficient statistic, the maximum likelihood Eqn of θ implies:

$$\theta^{t+1} = \log\left[\frac{\sum\left(1-d_i\right)c_i + \sum d_i c_i}{\sum\left(1-d_i\right)\left(1-\alpha_i\right)\left(1-c_i\right) + \sum d_i\left(1-\alpha_i\right)\left(1-c_i\right)} + 1\right]$$

For $d_i = 0 \;\; for\ TB \;and\; d_i = 1 \;\; for \;\; TB/HIV$, we have

$$\theta^{t+1} = \log\left[\frac{\sum_{i=1}^{m} c_i + \sum_{i=m+1}^{n} c_i}{\sum_{i=1}^{m}\left(1-\alpha_i\right)\left(1-c_i\right) + \sum_{i=m+1}^{n}\left(1-\alpha_i\right)\left(1-c_i\right)} + 1\right]$$

$$= \log\left[\frac{m+\sum_{i=m+1}^{n} c_i}{\sum_{i=m+1}^{n}\left(1-\alpha_i\right)\left(1-c_i\right)} + 1\right]$$

$$= \log\left[\frac{m+S_2}{S_1} + 1\right] \tag{14}$$

Equation **9** could be written as follows:

$$\frac{\partial l_c}{\partial \lambda} = \frac{c_i\alpha_i\sum_{i=1}^{m_1}(1-d_i)+c_i\alpha_i\sum_{i=1}^{m_2}d_i}{\lambda} - c_i\alpha_i t_i\sum_{i=1}^{m_1}(1-d_i) - c_i\alpha_i t_i\sum_{i=1}^{m_2}d_i + c_i(1-\alpha_i)\sum_{i=1}^{m_1}(1-d_i) + c_i(1-\alpha_i)\sum_{i=1}^{m_2}d_i\left(\frac{t_i}{\ell^{-\lambda t_i}-1}\right)$$

For $d_i = 0 \;\; for\ TB \;\; and \;\; d_i = 1 \;\; for\ TB/HIV$, we have:

$$\frac{\partial l_c}{\partial \lambda} = \frac{m}{\lambda} - \sum_{i=1}^{m} t_i c_i + \sum_{i=m+1}^{n} c_i(1-\alpha_i)\left(\frac{t_i}{\ell^{-\lambda t_i}-1}\right) = 0$$

$$= \frac{m}{\lambda} - \sum_{i=1}^{m} t_i(1-p_i) + \sum_{i=m+1}^{n}(1-p_i)(1-\alpha_i)\left(\frac{t_i}{\ell^{-\lambda t_i}-1}\right) = 0 \qquad (15)$$

RESULTS AND DISCUSSION

In this study, exponential, binomial, and uniform distributions were used to generate the data which is composed of lifetime t, censoring, cure status and covariates vectors (α, c, Z), respectively.

In this simulation we generated 100, 200 and 300 individuals who are effected with tuberculosis (TB) and TB/HIV co-infection using R package, we considered the left censoring individuals are cured. Thus, the cure indicator c can be defined in the same manner of censoring indicator where:

$$c(i) = \begin{cases} 0 \;\; if\ time[i] = T(i) \\ 1 \;\; if\ time[i] = T(i) \end{cases}$$

The time is determined using the median time for each of the diseases, for TB we used 4 months and TB/HIV we used 6 months.

Regarding the covariates, only two were considered: gender and type of treatment which are generated from binomial distribution. In this simulation, we are interested in estimating the proportion of patients who are cured and uncured from the two diseases considered and also to estimate their respective variance, bias and MSE.

$$bias = \pi - E(\pi)$$

$$MSE = \text{var}\,iance(\hat{\pi}) + bias^2$$

From the simulation we were able to come up with the following tables:

Table 1: TB Patients.

Sample size (n)	Proportion of cured	Proportion of the uncured	Bias	Variance
100	29	71	0.00061	0.3502
200	20	80	0.00040	0.3142
300	16	84	0.00035	0.2609

Table 2: TB/HIV Patients.

Sample size (n)	Proportion of cured	Proportion of the uncured	Bias	Variance
100	27	73	0.00058	0.3174
200	21	79	0.00051	0.3011
300	24	76	0.00031	0.1694

SUMMARY

We were able to show from the model developed that it is consistent because as the sample size tends to infinity the variance tends to zero (decreases) and that the model is unbiased because the bias is very close to zero.

CONCLUSION

We were able to develop a model suitable to manage a disease with co-infection from an existing model and we are also able to estimate the proportions of the patients cured and uncured from the two diseases with their variance, bias using the parametric maximum likelihood

estimation methodology for the cure rate estimation based on bounded cumulative hazard model when the exponential distribution can be used to represent the survival function of the uncured patients.

CONFLICT OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

1. Aditi, D. 2013. "A Study of the Cure Rate Model with Case Weights and Weights and Time-Dependent Weights". A Masters' Thesis. Mathematics and Statistics, University of Guelph: Ontario, Canada.

2. Aljawadi, B. A., M.R.A. Bakar, N.A. Ibrahim, and H. Midi. 2011. "Parametric Estimation of the Cure Fraction Based on BCH Model using Left Censoring Data with Covariates". *Modern Applied Science*. 5(3):103-110.

3. Amrita, D. 2011. "Integrating Patients into Integrated Care: Perspectives from Individuals Co-infected with Tuberculosis and HIV". Ph.D. Thesis. Dalla Lana School of Public Health University of Toronto: Toronto, Canada.

4. Berkson, J. and R.P. Gage. 1952. "Survive Curve for Cancer Patients following Treatment". *Journal of the American Statistical Association*. 47(259): 501-515.

5. Cancho, V.G., J. Rodrigues, and M. de Castro. 2011. "A Flexible Model for Survival Data with a Cure Rate: A Bayesian Approach". *Journal of Applied Statistics*. 38:57-70.

6. Chen, M.H., J. Ibrahim, and D. Sinha. 1999. "A New Bayesian Model for Survival Data with a Surviving Fraction". *Journal of American Statistical Association*. 94:909-918.

7. Chen, T.T. 2016. "Predicting Analysis Times in Randomized Clinical Trials with Cancer Immunotherapy". *BMC Medical Research Methodology*. 16(12):1-10.

8. Cooner, F., S. Banerjee, B.P. Carlin, and D. Sinha. 2007. "Flexible Cure Rate Modeling under Latent Activation Schemes". *Journal of the American Statistical Association*. 102:560-572.

9. Cox, D.R. and D. Oakes. 1984. *Analysis of Survival Data*. Chapman and Hall: London, UK.

10. Ewell, M. and J.G. Ibrahim. 1997. "The Large Sample Distribution of the Weighted Log-Rank Statistics under General Local Alternatives". *Life Data Analysis*. 3:5-12.

11. Farewell, V.T. 1982. "The Use of Mixture Models for the Analysis of Survival Data with Long-Term Survivors". *Biometrics*. 38:1041-1046.

12. Farewell, V.T. 1986. "Mixture Models in Survival Analysis: Are they Worth the Risk?". *Canadian Journal of Statistics*. 14:257-262.

13. Freireich, E.J., E. Gehan, E. Frei, L.R. Schroeder, I.J. Wolman, R. Anbari, E.O. Burgert, S.D. Mills, D. Pinkel, O.S. Selawry, J.H. Moon, B.R. Gendel, C.L. Spurr, R. Storrs, F. Haurani, B. Hoogstraten, and S. Lee. 1963: "The Effect of 6-Mercaptopurine on the Duration of Steroid-Induced Remissions in Acute Leukemia; A Model For Evaluation of Other Potentially useful Therapy. *Blood*. 21:699-716.

14. Gray, R.J. and A.A. Tsiatis. 1998. "A Linear Rank Test for use when the Main Interest is in Difference in Cure Rate". *Biometrics*. 45: 899-904.

15. Gu, Y., D/ Sinha, and S. Banerjee. 2010. "Analysis of Cure Rate Survival Data under Proportional Odds Model". *Lifetime Data Analysis*. 17:123-134.

16. Jaiswal, R.K., S. Srivastav, and H. Mahajan. 2012. "Socio Demographic Profile of TB-HIV Co-Infected Patients in Bundelkhand Region, Uttar-Pradesh". *Natl. J. of Med. Res.* 2:149-151.

17. Janida, M.A., L. Lawson, O.O. Nnodu, M.N. Ajani, E.O. Umobong, C.C. Fredrick, R. Adelaiye, W.D. Nafarda, S.T. Abdurraham, J.N. Okechukwu, I.T. Tafida, I. Bashir, N.D. Baamlong, R.Q. Gurgel, M.A. Yassin, O. Oladimeji, and L.E. Cuevas. 2015. "Treatment Outcome of Patients Co-Infected with Tuberculosis and HIV in Abuja, Nigeria". *Nigerian Journal of Basic and Clinical Sciences*. 11(2):72-75.

18. Li, D., X. Wang, and D.K. Dey. 2016. "A Flexible Cure Rate Model for Spatially Correlated Survival Data Based on Generalized Extreme Value Distribution and Gaussian Process Priors". *Biometrical Journal*. (Accepted but unpublished)

19. Maller, R. and X. Zhou. 1996. *Survival Analysis with Long-Term Survivors*. John Wiley & Sons: New York, NY.

20. Nelson, W. 1982. *Applied Life Data Analysis*. John Wiley & Sons: New York, NY.

21. Oliveira, M.R. and F. Louzada. 2014a. An Evidence of Link between Default and Loss of Bank Loans from the Modeling of Competing

Risks. *Singaporean Journal of Business Economics and Management Studies*. 3(1):30-37.

22. Oliveira, M.R. and F. Louzada. 2014b. “Recovery Risk: Application of the Latent Competing Risk Model to Non-Performing Loans”. *Technologia de Credito*. 88:43-53.

23. Roy, D. and D.K. Dey. 2014. “Propriety of Posterior Distributions arising in Categorical and Survival Models under Generalized Extreme Value Distribution”. *Statistica*. 24:699-722.

24. Roy, D., V. Roy, and D.K. Dey. 2013. “Analysis of Survival Data with a Cure Fraction under Generalized Extreme Value Distribution”. Technical Report 49, Department of Statistics, University of Connecticut.

25. Stangl, D.K. and J.B. Greenhouse. 1998. “Assessing Placebo Response using Bayesian Hierarchical Survival Models”. *Life Data Analysis*. 4:5-28.

26. Struthers, C.A. and V.T. Farewell. 1989. “A Mixture Model for Time to AIDS Data with Left Truncation and Uncertain Origin”. *Biometrika*. 76:814-817.

27. Uddin, M.T., M.N. Islam, and Q.I.U. Ibrahim. 2006. “An Analytical Approach on Cure Rate Estimation Based on Uncensored Data”. *Journal of applied Sciences*. 6(3):548-552.

28. WHO. 2004. “Interim Policy on Collaborative TB/HIV Activities”. *Stop TB Department and Department of HIV/AIDS*. World Health Organization: Geneva, Switzerland.

29. WHO. 2008. *Global Tuberculosis Control 2008: Surveillance, Planning, Financing*. World Health Organization: Geneva, Switzerland.

30. Williams, F.E. 2016. “Pattern of Medicine use in the Management of HIV/AIDS Patients in Public Hospitals in Kwara State”. Ph.D. Thesis (Unpublished). University of Ilorin: Ilorin, Nigeria.

31. Yu, B., R.C. Tiwari, and K.Z. Cronin. 2004. “Cure Fraction Estimation from the Mixture Cure Rate Models for Grouped Survival Times”. *Statistics in Medicine*. 23:1733-1747.

ABOUT THE AUTHORS

Balogun Oluwafemi Samson, is a Lecturer in the Department of Statistics and Operations Research, Modibbo Adama University of Technology, Yola, Nigeria. He holds a Master of Science Degree in Statistics, (Ph.D. in view) from the University of Ilorin, Nigeria. His research interests include categorical data analysis, biostatistics, and modelling.

Jolayemi Emmanuel Teju, is a Professor and Lecturer in the Department of Statistics, University of Ilorin. His research interests include categorical data analysis and biostatistics.

SUGGESTED CITATION

Balogun, O.S. and E.T. Jolayemi. 2017. “Modeling of a Cure Rate Model for TB with HIV Co-Infection”. *Pacific Journal of Science and Technology*. 18(1):288-299.

Pacific Journal of Science and Technology