ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 4, Issue 2, pp: (257-272), Month: April - June 2016, Available at: www.researchpublish.com

Effectiveness Analysis of Internship Program in Rwanda from July 2014 to June 2015: Kaplan Meier and Cox Regression Models

[1]MURASIRA GERARD, [2]Dr. Joseph Kyalo Mung'atu

[1,2]Department of Statistics & Actuarial Science, Jomo Kenyatta University of Agriculture and Technology

Abstract: **The study conducted on internship program that is under coordination of National Capacity Building Secretariat (NCBS), data was collected from fresh graduates, who have applied for internship program on the period from July 2014 to June 2015. In our study the sampled individuals are interns who were placed in different institutions and finalized the program before completing the period of 6 months, the list was extracted in internship database. The total number of 7000 graduates or students were hosted in different institutions in Rwanda in the period of 6 years, but the number that was used in our study is composed by 786 interns who hosted and completed the internship in the period of July 2014-2015, among them 119 were found a job and are conserved as censored cases to our study. The Kaplan Meier and Cox regression models will be used to analysis data collected. (IBM SPSS Statistics V20.0) software was be used to do the analysis in this study. In this research, we have introduced section, the literature review is the second while the third is the data collection and the forth as data analysis and interpretation of data analyzed. The final is the conclusion of the whole research and recommendations to address issues on internship program. The work plan and budget for this research were developed; the research is worth 1,280,000 Rwandan francs that was utilized to collect data, communication, stationary, publishing, etc.**

Keywords: **Analysis of Internship Program in Rwanda internship program that is under coordination of National Capacity.**

1. INTRODUCTION

1.1 Background of the Study:

In Rwanda, the majority of the population is very youth. Since 2009, the Rwanda Development Board (RDB), through its Department of Human Capital and Institutional Development (HCID), with the collaboration of the Youth Employment Steering Committee, has initiated the internship program destined to the University and TVET graduates (Internship Impact Assessment Report, RDB, December 2013, Pag.2).

There is an explicit, connection between poverty reduction and decent, productive work for young people. In Rwanda, the majority of the population is very young: 67 % of the population is under 25 year old.

Rwanda is committed to finding solutions and has taken the initiative of becoming a Youth Employment Network (YEN) Lead Country.

RWANDA has recognized this link between decent, productive employment for young people and poverty reduction. The Government approved the national employment policy and endorsed one of its implementing action plan; the National Action Plan Promoting Youth Employment on March 14th, 2007 demonstrating a clear commitment for youth employment issues.

Youth employment programs and activities are recognized as essential for the improvement of employability and crucial to Rwanda's sustained development. One activity highlighted in the National Employment Policy, and identified as an area of opportunity for increasing youth employability is the Internship Program within the Ministries and Public or Private Institutions of Rwanda and abroad

The objectives of this program are to provide the graduates with the capacity reinforcement by joining the theory to the practice, the professionalism development, the effective work competencies and employability skills creation.

Internships provide real world experience to those looking to explore or gain the relevant knowledge and skills required to enter into a particular career field. Internships are relatively short term in nature with the primary focus on getting some on the job training and taking what's learned in the classroom and applying it to the real world. Interns generally have a supervisor who assigns specific tasks and evaluates the interns overall work. For internships for credit, usually a faculty sponsor will work along with the site supervisor to ensure that the necessary learning is taking place. Interns can be high school or college students or even adults interested in trying out a new career.

In 2013 the first internship impact assessment/evaluation was conducted in order to check the impact of the program. The survey aimed to assess the impact of the program on the graduates' employment. It was conducted in the five provinces of the country.

Currently a number of **7,000** graduates or students were hosted in different institutions in Rwanda in the period of 6 years, each intern had to be hosted in the period of 6 months or less, and supervised by the institutional staff who is working in the area that the intern supposed to do the practice of knowledge acquired during his/her studies.

1.2 Statement of the problem:

Since 2009, the internship program was started to be implemented, the number of **7,000** graduates were placed in different institutions (public, private and civil society). The purpose of internship is to initiate the fresh graduates on the jobs and become perfect employees in the institutions that they will work with. The interns were placed based on the availability of the places respect to their training areas in the high learning institutions, the person in charge of placement respect also the order of applicants in the system and availability of the applicants/graduates.

The applicants has to wait for until the availability of place, sometimes the applicants should be placed while he/she has already found the job, he/she is automatically replaced by the following candidates in the system.

Students who have dedicated themselves to complete their studies lack opportunities, to gain experience and often fail to find productive employment. "Internships or apprenticeships provide tangible work experience required by most employers" For students attending universities, a high youth unemployment rate with high education sends a negative message about education's importance.

A Youth Internship program in public or private institutions of Rwanda and abroad would increase the transition between school and the workforce. This program could contribute significantly to improving the employability whilst also helping address the human resource constraints which some institutions may be facing. An Internship program would be an opportunity for youth to gain such kills and experience, Furthermore, while education provides students mainly with theoretical knowledge, skills training and experience build confidence. Interns will be backed up by senior, staff to work effectively towards hosting agencies, objectives and will also get facilitation fees.

Internship is a critical component of the academic and professional development. "Given the social, economic, and political costs of youth unemployment and underemployment, inaction is not an option. Reinforced skills and trainings will allow young people to raise their employability and help to improve the welfare of the Country.

It is a challenge to know if there is any relationship among flesh graduates who applied for internship in NCBS database, flesh graduates hosted in different institutions (Public, Private and Civil society) and interns who got jobs during internship exercise in those institutions. From above paragraph, we want to know exactly if there is a relationship among the three given categories of flesh graduates.

1.3 Rationale of the study:

The findings from this research are interest and relevance to long-term investing and speculation as the results and findings can help policy makers, analysts or organizations to gain an in-depth understanding of the best of internship program in Rwanda in terms of gaining working experience to the flesh graduates.

It will help the private sector and civil society to understudy their contribution on the labor market in Rwanda with hosting flesh graduates as interns,

1.4 Research questions:

a. Is there a relationship between interns and graduates applied for internship and interns found jobs?

b. Do interns who get job increase in number during on job training?

1.5 Objectives:

1.5.1 General objective:

The general objective of the study is to determine the effectiveness of the internship program to the youth graduates by using Kaplan Meier and Cox regression models

1.5.2 Specific objectives:

a. To determine probability of getting job for interns who were hosted in different institutions during internship program.

b. To determine the contribution of internship program to the national employment

c. To compare models for quality results for internship effectiveness

1.6 Scope of the study:

The present study was carried out among the total number of hosted interns and who found jobs in different institutions during internship program (they have already applied for internship through online system (www.ncbs.gov.rw/intership) from July 1014 to June 2015 while it was under control and management of National Capacity Building Secretariat (NCBS).

Data on interns used in this report were extracted in the internship database, where the system administrator in NCBS is updating the system when there is an additional/change information on the interns.

When an inters found a job, him/herself may inform NCBS on the situation or the supervisors have to report to NCBS the current information to interns in his/her institution.

2. LITERATURE REVIEW

Within the literature we will discuss the theoretical designs and models used in the evaluation of internship program for its effectiveness to the implementing institutions or to the country in general.

The government of Rwanda has initiated the internship program to promote youth to acquire practical skills and to initiate the interns on employment area. The government of Rwanda approved the national Employment policy and endorsed one of its implementation action plans. The National Action Plan promoting youth employment on March 14th 2007 demonstrating a clear commitment for youth employment issues.

In 2009 Ministry of public service and labor established a guide document for internship that is Labor and employment services.

In his research "Determinants of internship effectiveness for university students in Hong Kong, April 2010" he said that Internship has been viewed as an effective approach to equip university students with preliminary job knowledge and experience, thus enhancing their employability in the competitive labor market. Focusing on identifying and evaluating the determinants of internship effectiveness This study aims to explore the extent to which the characteristics of flesh graduates interns and organization practices account for their internship success and how flesh graduated applicants are accessing internship program in Rwanda or abroad.

The internship program has 12 steps, if they are well developed and applicable as planned, it was supposed to be achievement and the impact is clear to everybody.

The variations among the terms used to describe outside-the-classroom learning opportunities can be condensed into one phrase: experiential education. Juliet Miller, from the U.S. Department of Education, defines an experiential education as "all programs that are designed to expand the setting of learning experiences beyond the traditional school environment to

occupational and community settings and these programs use planned experiences…to promote cooperation between traditional educational institutions and business, industry, labor, government and community groups to support learning" (p. 3). Some examples of experiential education include service learning internship, academic service-learning, community service-learning, cooperative education program, applied practicum, experiential learning, and, most commonly internship program, Sir David Cox's 1972 paper took a different approach to standard parametric survival analysis and extended the methods of the non-parametric Kaplan-Meier estimates to regression type arguments for life-table analyses. Cox advanced to prediction of survival time in individual subjects by only utilizing variables covarying with survival and ignoring the baseline hazard of individuals. Cox did this by making no assumptions about the baseline hazard of individuals and only assumed that the hazard functions of different individuals remained proportional and constant over time.

In internship program for Rwanda, there are many steps that are respected for both sides, the flesh graduates hosted for internship program are automatically removed on the waiting list of applicants, and appeared on the list of interns, during internship interns may found a job because of on job training that he/she is receiving from the supervisor in the host institution, in that case, those who found jobs are considered as censored cases from the time frame of the cohort "t"

Kaplan Meier and Cox regression analysis are commonly employed in Health sector and social science fields. It is also common for interpretation of results to typically reflect on hazard rate, function within censoring processes, often resulting in very limited interpretations of variable importance. The frame work will be used to show the relationship between interns, fresh graduates, fresh graduates applied and interns hosted in different institutions

2.1 Internship Outcomes:

By far, most research on internships focuses on student learning as the major outcome. Internships may also help students acquire job relevant skills (Garavan & Murphy, 2001) such as writing skills (Freedman & Adam, 1996; Winsor, 1990), and help students put abstract concepts into context (Bowers & Nelson, 1991). A study of service learning, another situation which puts students into real-world settings, found that problem solving, critical thinking, and theoretical skills are improved in nonacademic settings (Matthews & Zimmerman, 1999).

Finally, internship experiences can overcome presumed shortcomings such as the lack of specific preparation, sometimes called "deficit reduction theory" (Herr & Cramer, 1988). Beyond skills training, internships also help socialize and acculturate (Tovey, 2001). They improve career decision making and perceptions of self-efficacy (Brooks, Cornelius, Greenfield, & Joseph, 1995; Taylor, 1988). Students who have completed an internship display greater ambition (Pedro, 1984). In the same vein, other research indicated that internships reduce reality shock for students (Paulson & Baker, 1999; Taylor, 1988).

We propose that the outcomes from the internship may be of three types. Paralleling the outcomes that have been found in other knowledge transfer activities, such as expatriate assignments (e.g., Edstrom & Galbraith, 1977; Hocking, Brown, & Harzing, 2004), we argue that outcomes of interest include (1) organizational benefits from the completion of the internship project, (2) enhanced capabilities of the company and the university, and, at the student level, (3) skill development and career enhancement. For the employing firm, the important immediate benefits may include project completion, efficiencies due to using cheaper labor, potential screening of and recruitment of the intern and, perhaps, an inflow of ideas (e.g., best practices) from the university to the company. Longer term benefits may include a stronger tie with the university, thereby encouraging students to apply for future internships, as well as a continued inflow of ideas. For the university, completion of the project will likely result in student satisfaction and possibly student placement. Longer term, one might expect to see the university enhance its capabilities by having greater knowledge about the challenges companies address as well as a reputation for placing students in good internship positions.

For the student, the outcomes will mostly focus on skill development and career opportunities. Satisfaction from the internship experience, possible employment, and less career shock may be likely proximal outcomes. Longer term outcomes will be perhaps better career decisions and career prospects. We have developed a model of possible determinants of internship effectiveness, by combining the limited knowledge about internships with research into personnel and knowledge transfers between organizations. While the paucity of research into internships encourages a model-building effort, the lacks of extensive empirical data also suggest there will be benefits from testing at least some of the relationships proposed in the model. In particular, as part of our theory-building effort, it would be beneficial to

ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 4, Issue 2, pp: (257-272), Month: April - June 2016, Available at: www.researchpublish.com

explore the impact of the two ideas imported from the personal and knowledge transfer literature: Internships involve three actors and are a process. Specifically, we propose two sets of relationships. In the first, we argue that relationships among internship variables will reflect the different levels of analysis of the respective actors. The second is that the internship process variables will have a relationship to outcome variables that is separate from the effect of the antecedent variables on outcome variables. In order to make this second contribution, we now turn to data that provide preliminary evaluation of these two basic ideas as well as of some parts of the full model. In this way, we can provide guidance to future researchers as they attempt to evaluate the full model

3. RESEARCH METHODOLOGY

This research was based on quantitative, research methods. In this research, we collected numerical data. This is closely connected to the final part of the definition: analysis using mathematically based methods. In order to be able to use mathematically based methods, our data have to be in numerical form. Our data is the number of graduates applied for internship program, graduates hosted in institutions and interns found jobs during internship for both monthly basis, therefore data can be analyzed by using quantitative methods

The research has an objective of using statistical analysis to check a relationship between different groups of graduates after being hosted as intern who gained experience that is required on the labor markets in Rwanda.

In our study, we used the techniques of documentary and observations. Hence, there we sampled the interns hosted in different institutions from July 2014 to June 2015 as sampled period.

In the research, we drafted template to collect data and SPSS used to analyse data extracted in the system. It relied on the understanding of the past and the present observation and we predicted to the future and the come up with suggestions and recommendations.

Then, we also used the multi linear regression model in the analyzing data. Kaplan Meier and Cox regression models stems from the desire to reify our natural ability to visualize the shape of data.

3.1 Definitions and concepts:

Time-to-event is a clinical course duration variable for each subject having a beginning and an end anywhere along the time line of the complete study; i.e, it may begin when the subject is enrolled into a study or when treatment begins, and ends when the end-point (event of interest) is reached or the subject is censored from the study (more on this later). This duration is known as serial time.

In the most clinical trials, individual subjects may enter or begin the study (zero time) and reach end-point at vastly differing points along the trial calendar, in our case the flesh graduates may register in the system at different time, access or get out the internship differently.

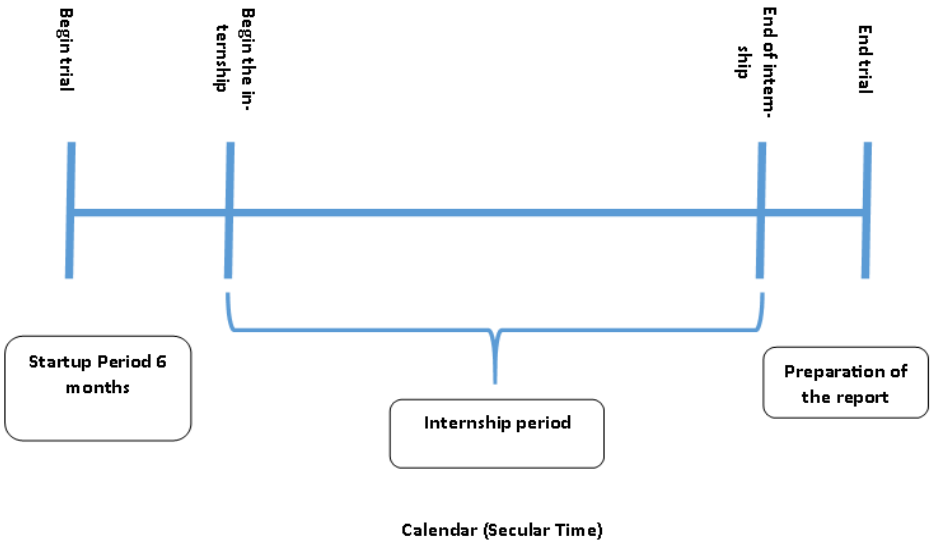

Figure.1: Internship trial

ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 4, Issue 2, pp: (257-272), Month: April - June 2016, Available at: www.researchpublish.com

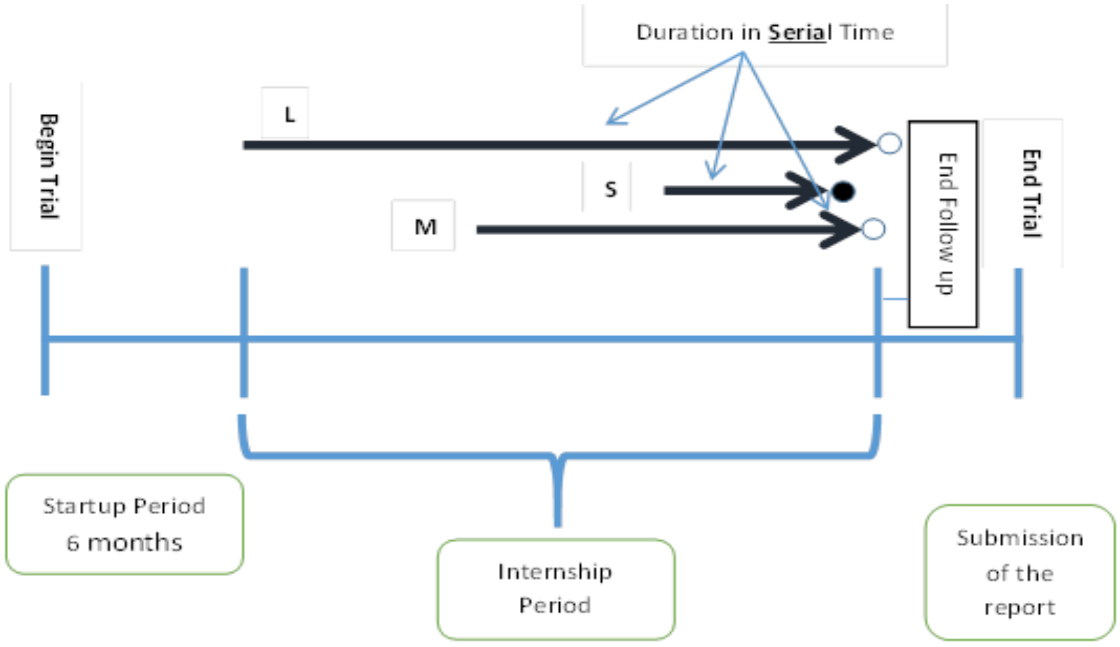

Figure.2: Internship trial with serial time

This is the subject of entering trial of internship program and ending at different time. L, M and S are indicating the serial times as following: long, Medium and Long serial times

3.2 Prevalence and incidence:

Prevalence is a frequently used epidemiological measure of how commonly a disease or condition occurs in a population. Prevalence measures how much of some disease or condition there is in a population at a particular point in time.

The prevalence is calculated by dividing the number of persons with the disease or condition at a particular time point by the number of individuals examined. For example, in the study of effectiveness of internship program in NCBS from 2014-2015, 797 individuals were registered in the database for internship of six months and have done the internship program in different institutions. Of these 797 people, 119 drop out the program without completing the six months as planned for that program due to the found of other occupations/ change of duties because after having experience. Thus the prevalence of having a job was 519/6139 = 0.14931. Prevalence is often expressed as a percentage, calculated by multiplying the ratio by 100. The above study expresses prevalence as a percentage, thus the prevalence of dropping out the internship program for finding a job is 14.93% (or rounded is 15%). Another common way of expressing prevalence, particularly if the prevalence is low, is as the number of cases per 100,000 of the population. For example, it is easier to state the result as `66 cases per 100,000 people' than to say the prevalence is 0.00066.

The incidence of a disease or an event, etc. is another epidemiological measure. Incidence measures the rate of occurrence of new cases of a disease or condition. Incidence is calculated as the number of new cases of a disease or condition in a specified time period (usually a year) divided by the size of the population under consideration who are initially disease free. For example, the incidence of finding a job for interns in the Rwanda in July 2015 could be calculated by finding the number of new cases interns recruited during July 2015 and dividing that number by the population of the Rwanda. As this incidence rate would be very small again we tend to consider number of cases per 100,000 people.

3.3 Kaplan Meier models:

3.3.1 Introduction to Kaplan Meier:

In 1958, Edward L. Kaplan and Paul Meier collaborated to publish a seminal paper on how to deal with incomplete observations.[1] Subsequently, the Kaplan-Meier curves and estimates of survival data have become a familiar way of dealing with differing survival times (times-to-event), especially when not all the subjects continue in the study

The Kaplan-Meier method is a nonparametric (actuarial) technique for estimating time-related events (the survivorship function). Ordinarily it is used to analyze death as an outcome, in biostatistics, but in recent years these techniques have

also gained popularity in the social sciences or industrial statistics (an economist might measure the length of time people remain unemployed after a job loss or an engineer might measure the time until failure of machine parts A plot of the Kaplan-Meier estimate of the survival function is a series of horizontal steps of declining magnitude which, when a large enough sample is taken, approaches the true survival function for that population. It should be also used to measure the time on internship program for flesh graduates who were hosted in different institutions (public, Private and Civil Society) in Rwanda. The value of the survival function between successive distinct sampled observations is assumed to be constant. An important advantage of the Kaplan-Meier curve is that the method can take into account "censored" data — losses from the sample before the final outcome is observed (for instance, if a patient withdraws from a study).

Normally all interns are assignment to accomplish the period of 6 months in the hosted institution, the censoring will appear when he/she left before the completion due to different reasons; as following: finding a new job, being unhappy for the placement, death, etc…

The censored case should be represented on the plot, small vertical tick-marks indicate losses, where item or member of population data has been censored. When no truncation or censoring occurs, the Kaplan-Meier curve is equivalent to the empirical distribution. Kaplan Meier method presupposes a greater reduction in calculus volume than the actuarial method, because survival is estimated every time when the pre-established event for a subject occurs (employment in our case), thus neglecting the registrations lost of sight along the survey. The stages of Kaplan-Meier method are:

- listing the time when the pre-established event occurs, since subject's involvement in the survey (participation time);

- Finding for every participation time the number of subjects that continue to participate in the survey – those who did not achieve the pre-established event (employment in our case);

- establishing the number of subjects who achieved the pre-established event within *nx* time interval;

- The calculus of the probability of occurrence of the pre-established event, for each participation interval *(dx)* according to the formula: *qx=dx/nx*, where *x* is the participation duration;

- As for the actuarial method, the calculus of survival probabilities for *x* duration is: *px=1-qx*, and the cumulated survival probability is *Px=px(px-1)(px-2)....p2p1*.

The Kaplan-Meier technique is usually only useful as a method of preliminary evaluation, since it is purely a descriptive method for the evaluation of one variable. The survival curve of this method is scalar form because the proportion of subjects who have the chance to continue observation without the occurrence of the pre-established event changes exactly at the moments when the pre-established event is achieved. The survival level is of 100% from the curve origin until the moment of the first occurrence of the event (employment in our case), where it drops to the new calculated value, that constitutes a new level during which survival is constant, until the next event achieved. Therefore, every step corresponds to the occurrence of one or several pre-established events.

The Kaplan-Meier (KM) estimator, or product limit estimator, is the estimator used by most software packages because of the simplistic step approach. The KM estimator incorporates information from all of the observations available, both censored and uncensored, by considering any point in time as a series of steps defined by the observed survival and censored times. When there is no censoring, the estimator is simply the sample proportion of observations with event times greater than t. The technique becomes a little more complicated but still manageable when censored times are included

3.3.2 Explanation of Kaplan Meier model:

Let t_1, t_2, t_3, denote the actual times of death of the n individuals in the cohort. Also

Let d_1, d_2, d_3, denote the number of deaths that occur at each of these times, and let n_1, n_2, n_3,… be the corresponding number of patients remaining in the cohort. Note that $n_2 = n_1 - d_1$, $n_3 = n_2 - d_2$, etc. Then, loosely speaking, $S(t_2) = P(T > t_2)$ = "Probability of surviving beyond time t_2" depends conditionally on $S(t_1) = P(T > t_1)$ = "Probability of surviving beyond time t_1." Likewise, $S(t_3) = P(T > t_3)$ = "Probability of surviving beyond time t_3" depends conditionally on $S(t_2) = P(T > t_2)$ = "Probability of surviving beyond time t_2," etc. By using this recursive idea, we can iteratively build a numerical estimate $\hat{S}(t)$ of the true survival function S(t). Specifically,

ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 4, Issue 2, pp: (257-272), Month: April - June 2016, Available at: www.researchpublish.com

For any time t ∈ [0, t_1), we have S(t) = P(T > t) = "Probability of surviving beyond time t" = 1, because no deaths have as yet occurred. Therefore, for all t in this interval, let $\hat{S}(t)$ = 1.

For any two events A and B, P(A and B) = P(A) × P(B | A).

Let A = "survive to time t_1" and B = "survive from time t_1 to beyond some time t before t_2." Having both events occur is therefore equivalent to the event "A and B" = "survive to beyond time t before t_2," i.e., "T > t." Hence, the following holds.

For any time t ∈ [t_1, t_2), we have…

$$S(t) = (T > t) = \underbrace{P(survive in [0, t_1))}_{1} \; x \; \underbrace{P(survive in [t_1, t] / \; survive in [0, t_1))}_{\frac{n_1 - d_1}{n_1}}$$

then

$$\hat{S} = 1 \; x \; \frac{n_1 - d_1}{n_1}$$

Or

$$\hat{S}(t) = 1 \; x \; \frac{n_1 - d_1}{n_1}$$. Similarly,

For any time t ∈ [t_2, t_3), we have…

$$S(t) = (T > t) = \underbrace{P(survive in [t_1, t_2))}_{1 - \frac{d_1}{n_1}} \; x \; \underbrace{P(survive in [t_1, t] / \; survive in [0, t_1))}_{\frac{n_2 - d2}{n_2}}$$

then

$$\hat{S}(t) = \left(1 - \frac{d_1}{n_1}\right)\left(1 - \frac{d_2}{n_2}\right) \quad etc.$$

In general, for t ∈ [t_j, t_{j+1}), j = 1, 2, 3,…, we have…

$$\hat{S}(t) = \left(1 - \frac{d_1}{n_1}\right)\left(1 - \frac{d_2}{n_2}\right) \ldots \left(1 - \frac{d_j}{n_j}\right) \quad = \quad \prod_{i=1}^{j} \left(1 - \frac{d_j}{n_j}\right)$$

This is known as the Kaplan-Meier estimator of the survival function S(t). (Theory developed in 1950s, but first implemented with computers in 1970s.) Note that it is not continuous, but only piecewise-continuous (actually, piecewise-constant, or "step function").

The KM estimator is a nonparametric estimator of the survivor function S(t).

$$\hat{S}(t) = \Pi_{t_{s \le t}} \left[1 - \frac{d_j}{n_j}\right] \qquad \textbf{(3.1)}$$

where d_j is the number of individuals who experience the event at time t(j), and n_j is the number of individuals who have not yet experienced the event at that time and are therefore still at risk for experiencing it.

The KM estimator consists of the product of a number of conditional probabilities resulting in an estimated survival function S(t) in the form of a step function. Using PROC LIFETEST we can compute and plot the survival curve of a single group or we can compare survival in subgroups.

3.3.3 Kaplan Meier Test:

The test of results analyzed using Kaplan Meier have to be tested for significance purpose, the test has to be based on three methods, There are three statistical tests that can be selected in SPSS Statistics that test whether the survival functions are equal. These are the log rank test (Mantel, 1966), Breslow test (Breslow, 1970; Gehan, 1965) and the Tarone-Ware test (Tarone & Ware, 1977),

All three tests compare a weighted difference between the observed number of events (i.e., the resumption of smoking) and the number of expected events at every time point, but differ in how they calculate the weight. We discuss the differences between these three statistical tests and which test to choose in our enhanced Kaplan-Meier guide.

It is fairly common to find that all three tests will lead you to the same conclusion (i.e., they will all reject the null hypothesis or they all won't), but which test you choose should depend on how you expect the survival distributions to differ so as to make best use of the different weightings each test assigns to the time points (i.e., increase statistical power). Unfortunately, you cannot rely on there being one best test it will depend on your data. If you choose the approach of picking a particular test, you will need to do this before analyzing your data. You shouldn't run all of them and then simply pick the one that happens to have the "best" p-value for your study (Hosmer et al., 2008; Kleinbaum & Klein, 2012)

The **log rank** method is testing the null hypothesis that there is no difference in the overall survival distributions between the groups (e.g., intervention groups) in the population. To test this null hypothesis, the log rank test calculates a $\chi 2$-statistic (the "Chi-Square" column), which is compared to a $\chi 2$-distribution with two degrees of freedom (the "df" column). In order to determine whether the survival distributions are statistically significantly different, you need to consult the "Sig." column which contains the p-value for this test. You can see that the significance level of this test is .000. This does not mean that $p = .000$, but that $p < .0005$

The **Gehan-Breslow-Wilcoxon** method gives more weight to deaths at early time points.

This often makes lots of sense, but the results can be misleading when a large fraction of patients are censored at early time points. In contrast, the log-rank test gives equal weight to all time points. The Gehan-Wilcoxon test does not require a consistent hazard ratio, but does require that one group consistently have a higher risk than the other

Tarone-Ware assumes identical distribution of background and compliance populations, and requires equal variances. Also, as with the other tests, the Tarone-Ware two-sample test also requires temporal stability and lack of spatial variability. To perform this test, the two data sets (for example, background and compliance data) are combined and the distinct (unique) detect values ordered from lowest to highest. The number of values (including nondetects) less than or equal to each ordered value is computed for compliance, background, and combined data. The Tarone-Ware statistic is then calculated using equations found in some statistical references, including the Unified Guidance.

Computed Tarone-Ware statistic (TW) greater than the tabulated critical value at the selected significance level, indicates that variables are statistically different (greater) at that significance level Cox's Proportional Hazards Regression model

3.3.4 Introduction:

Sir David Cox's 1972 paper took a different approach to standard parametric survival analysis and extended the methods of the non-parametric Kaplan-Meier estimates to regression type arguments for life-table analyses. Cox advanced to prediction of survival time in individual subjects by only utilizing variables covarying with survival and ignoring the baseline hazard of individuals. Cox did this by making no assumptions about the baseline hazard of individuals and only assumed that the hazard functions of different individuals remained proportional and constant over time.

When there are several explanatory variables, and in particular when some of these are continuous, it is much more useful to use a regression method such as Cox rather than a KM approach.

The log rank test cannot be used to explore (and adjust for) the effects of several variables, such as age and disease duration, known to affect survival. Adjustment for variables that are known to affect survival may improve the precision with which we can estimate the treatment effect. The regression method introduced by Cox is used to investigate several variables at a time.

ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 4, Issue 2, pp: (257-272), Month: April - June 2016, Available at: www.researchpublish.com

It is also known as proportional hazards regression analysis. Briefly, the procedure models or regresses the survival times (or more specifically, the so-called hazard function) on the explanatory variables.

3.3.5 The purpose of purpose of cox model:

As other some models, cox model is based on the analysis of survival analysis and its purpose is to simultaneously explore the effects of several variables on survival . It is a statistical technique that statisticians normally are using to analysis survival data. During analysis using this model other effects from other variables to the population under assessment are to be isolated

3.3.6 The heuristics behind the model:

Suppose we have 3 individuals, labeled A, B, and C, who are at risk for some particular event. Then we are told a week later that one of them had the event. If we know that each individual had the same risk for that event, what can we say about the probability that it was A (or B, or C)?

Knowing that each had the same risk tells us that each person is equally likely to have had the event, therefore: P(A had the event) = 1 / 3

The same holds for both B and C, i.e. P(B) = P(C) = 1/3

Now suppose (instead) that we are told that the risk for individual A was twice as high as B and C. Now what can we say about the probability of it being A (or B, etc.)? One way to think of this problem is to label the number of "risk balls" that correspond to the risks of each individual. If A has twice as much risk, then it has twice as many balls

If these balls were randomly mixed and blindly chosen, what is the probability that the ball is labeled 'A', 'B', or 'C'? The answer to this question gives us the probability that A had the event.

We know that P(A) = ½ whereas P(B) = P(C) = ¼ .

Now suppose that the risk of an event is proportional to the person's age. If we know the ages of these people, we can calculate the probability that it was A, B, or C.

In a sense, this is how the "pmf" for the proportional hazard model works. It assigns the probability that a particular individual has an event by dividing the risk (actually, the hazard) for that person by the sum of all the hazards for all of the people who are at risk.

That is, $$P(A) = \frac{h(A)}{h(A) + h(B) + h(C)} \qquad (3.2)$$

Of course, in statistics, the effects of the various covariates on the probability of individuals having an event are unknown because their associated parameters, bi, are unknown. In fact, the investigator is actually more interested in estimating the bi's for a given dataset. This estimation problem can use a likelihood function, and inference can rely on the asymptotic properties of the MLEs for the bi's. Before we move on to this problem, let's take a closer look at the proportional hazards assumption.

When there are several explanatory variables, and in particular when some of these are continuous, it is much more useful to use a regression method such as Cox rather than a KM approach

The hazard function is the probability that an individual will experience an event (for example dropping out the internship program) with a small time interval.

3.3.7 Introduction definition:

$$h_i(t) = h_0(t) . \exp\{ \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} \} \qquad (3.3)$$

That is, the hazard for individual, i, is based on two parts. $h_0(t)$ is called the baseline hazard function. It can be any complicated function of t as long as $h_0(t) \geq 0$. Fortunately, it doesn't have to be specified.

ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 4, Issue 2, pp: (257-272), Month: April - June 2016, Available at: www.researchpublish.com

The other part of the hazard function involves the covariates, which are in the exponent along with the unknown parameters. Note that this term does not involve a time variable. Therefore, the ratio of the hazards of two individuals does not depend on time, i.e. $h_0(t)$. This will simplify the problem of estimating the β_i's

$$\frac{h_i(t)}{h_j(t)} = \frac{h_0(t).\exp\{\beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik}\}}{h_0(t).\exp\{\beta_1 x_{j1} + \beta_2 x_{j2} + ... + \beta_k x_{jk}\}} = \exp\{\beta_1(x_{i1} - x_{j1}) + ... + \beta_k(x_{ik} - x_{jk})\} \quad \textbf{(3.4)}$$

Remember that the likelihood function is a function of the unknown parameters, given the outcomes of the random variables. If the events for the individuals are independent of each other, then the likelihood is a product of the likelihoods for all individuals in the sample:

$$PL = \prod_{i=1}^{n} L_i \quad \textbf{(2.5)}$$

The underlining or baseline hazard is the hazard when all covariates equal zero.

$$h_i(t,x) = h(t,0)e^{\beta' x} \quad \textbf{(3.6)}$$

h(t,0) is the baseline hazard rate at time t for covariate vector 0. A subject's hazard at time t is proportional to the baseline hazard ho(t). The proportionality factor depends on the covariate vector for an individual. If all covariate values are homogenous, then it gets subsumed into the baseline hazard function.

The probability that an individual dies, leaves, etc., at time t_i, is given by:

$$\frac{e^{\beta' x_j}}{\sum e^{\beta' x_j}} \quad \textbf{(3.7)}$$

The conditioning eliminates the baseline hazard function.

Researchers favor Cox's proportional hazards modeling because of the robust semi-parametric method of calculating the probabilities of survival while simultaneously adjusting for other possibly influential variables. Other attractive features of Cox modeling include: the relative risk type measure of association, no parametric assumptions, the use of the partial likelihood function, and the creation of survival function estimates.

4. RESULTS AND DISCUSSION

It is a section that summaries the data analysis and results interpreted of data collected on internship program (from internship database, the database is managed by National Capacity Building Secretariat) on fresh graduates who were hosted for internship program in different institutions (Public, Private, and Civil Society) in Rwanda. The data collection concerned to all interns hosted from July 2014 to June 2015, and who have finalized it in that period. Next step was to clean and organize data using excel sheet (Microsoft Excel 2013), codification and enter them into SPSS for analysis. The findings were analyzed by SPSS (IBM SPSS Statistics V20.0) and they are presented in tables, charts, graphs.

Long rank, Breslow and Tarone-wore statistical test were used to test the results as methods to compare the survival distribution

Here below results from analysis using Kaplan Meier and cox regression model are given:

Kaplan-Meier:

TABLE.1: Kaplan Meier_ Means and Medians for Survival Time

Mean[a]			
Estimate	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound
482.349	11.248	460.303	504.394
a. Estimation is limited to the largest survival time if it is censored.			

ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 4, Issue 2, pp: (257-272), Month: April - June 2016, Available at: www.researchpublish.com

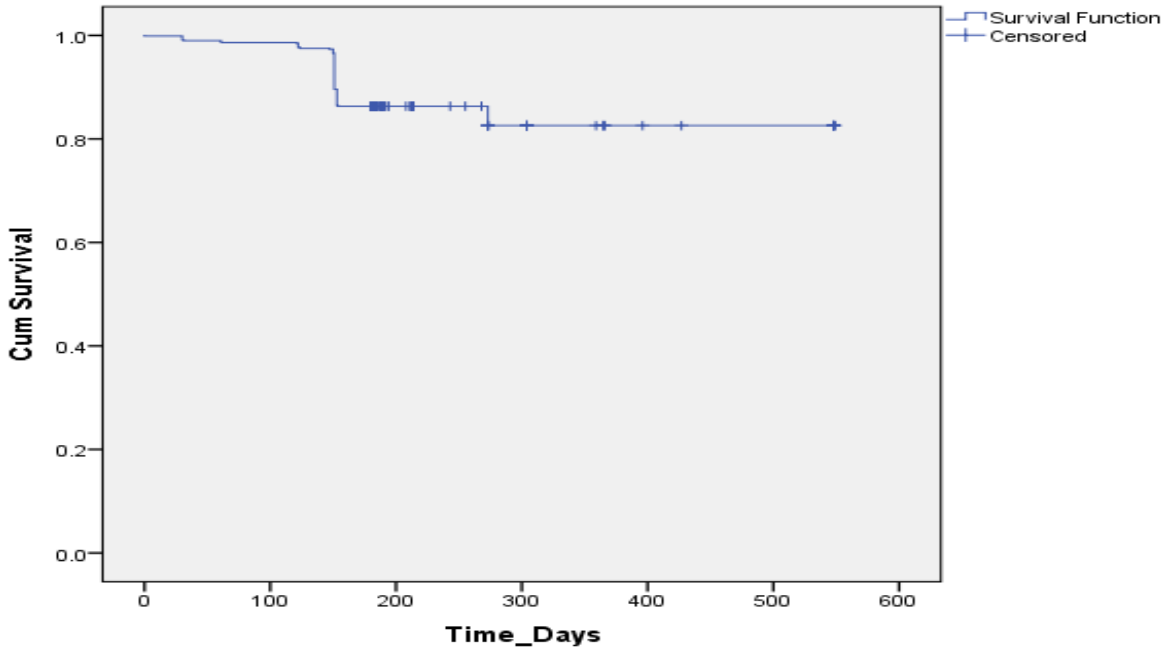

Figure.3: Kaplan Meier_ Survival Function

Normally the survival function has to fall down from the highest probability 1 to the lost zero. From above chart shows that the survival probability by survival function is belonging to 1 to 0.8 in that interval they are no many interns who got job during internship program until the end of six months as planned. At the end of the events the survival probability is not slopping to zero, it is closer to 0.8. i.e, being hosted as interns does not mean automatically finding job. When the probability is zero, it means that all interns found job during internship program.

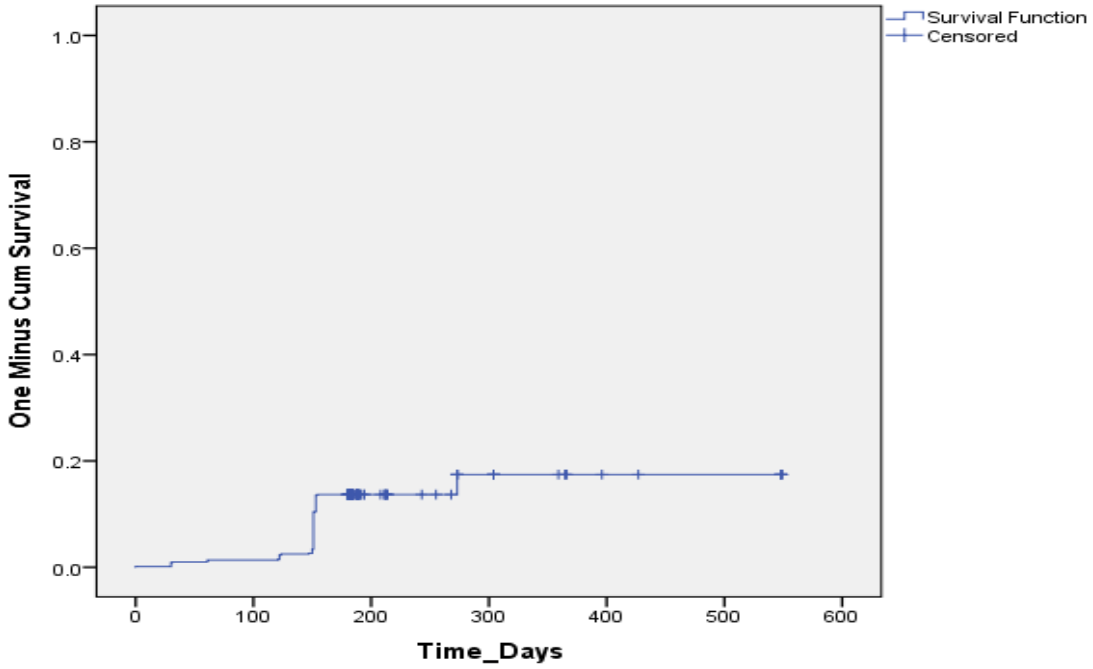

Figure.4: Kaplan Meier_ One Minus Survival Function

This chart shows that the probability of finding a job is varying (increase) in relation with time that interns are spending with on job training. It is noted the censoring started below 100 days, it was contracted before 180 days ends (equivalent to 6 months of internship period). i.e there were an number of interns who got jobs before completing internship program. The censoring after 180 days are the particular cases. The cumulative probability is not greater than 0.2 i.e there are not many interns who were censored from the event that is equivalent to the interns who got jobs during internship program.

ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 4, Issue 2, pp: (257-272), Month: April - June 2016, Available at: www.researchpublish.com

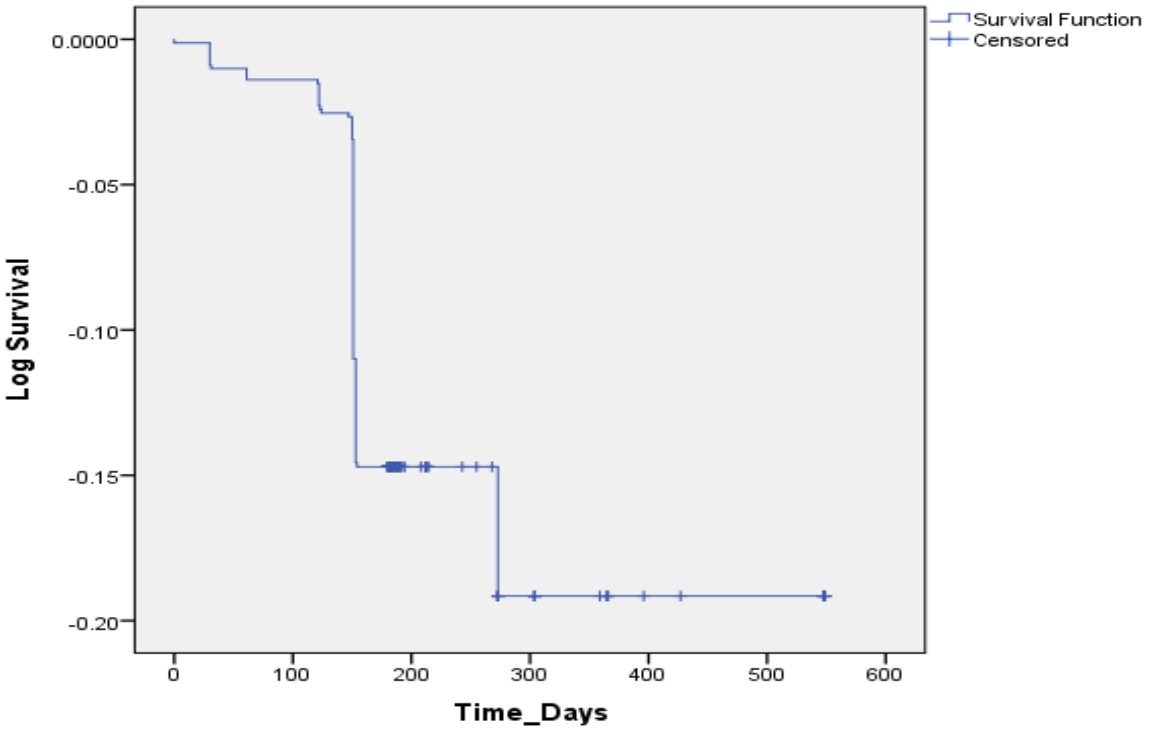

Figure.5: Kaplan Meier_ Log Survival Function

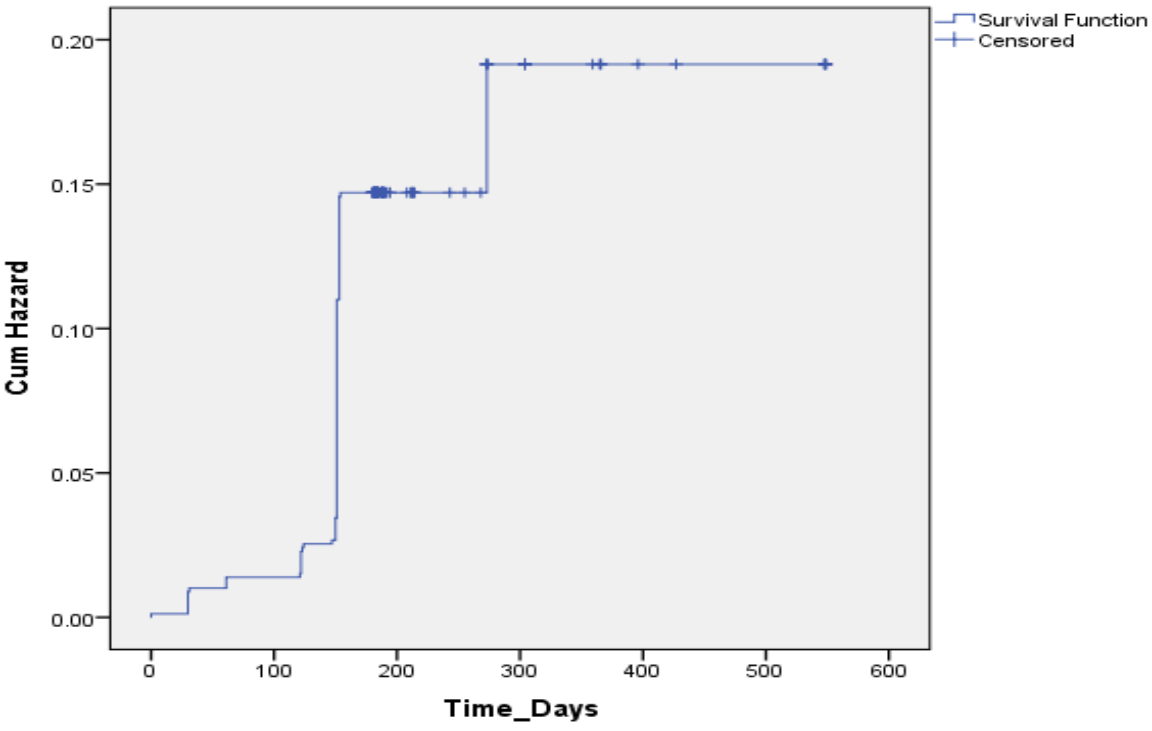

Figure.6: Kaplan Meier_ Hazard Function

The figure 6 and 5 long survival and hazard functions respectively to the increase in censoring or probability of interns finding a job for interns and who the interns were dropped out the internship for the purpose of finding jobs during internship program

4.1 Cox Regression Model:

The Kaplan Meier model has to be considered with the cox regression model in order to come up with quality result for this research. Charts bellows are detailing the results as follows:

ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 4, Issue 2, pp: (257-272), Month: April - June 2016, Available at: www.researchpublish.com

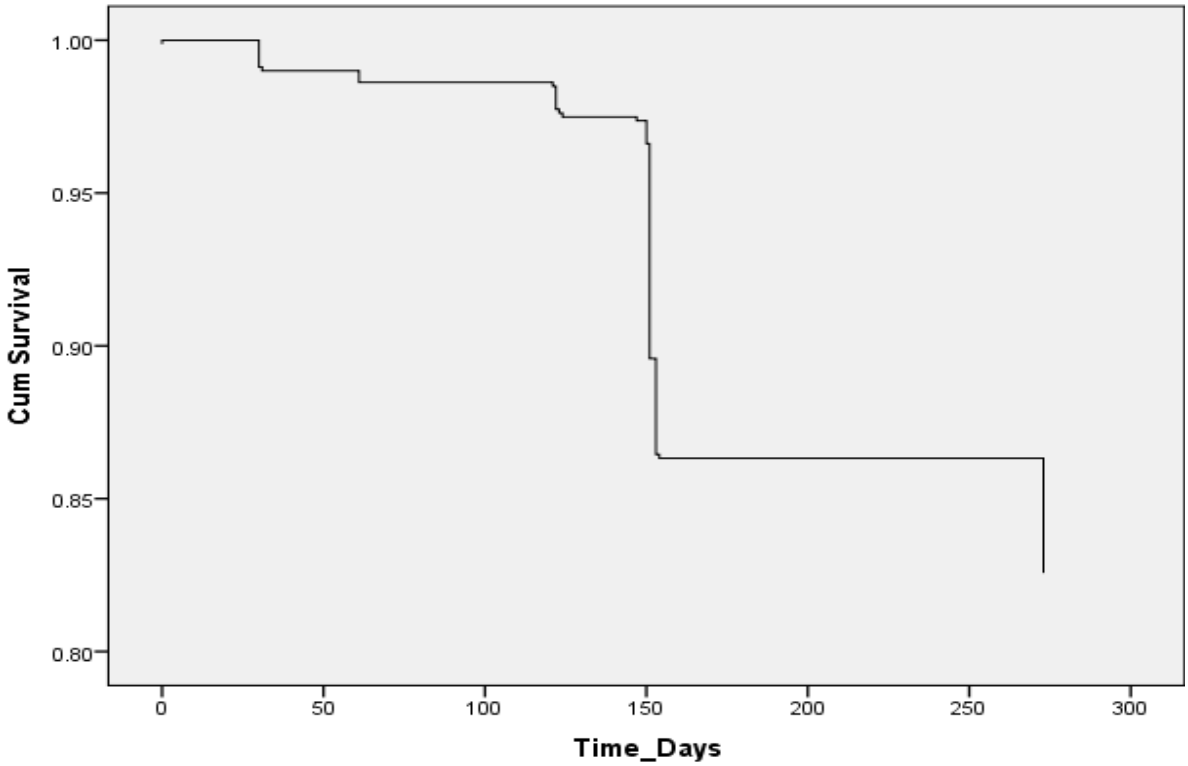

Figure.7: Cox Regression _Survival Function at mean of covariates

The decrease of curve is determining the interns who were censored from the internship program due to the finding of jobs for them during internship program. It noted that around 150 days there were many interns who dropped out the system (censored) i.e. many interns were employed for structure positions in different institutions. The event ends after 250 days.

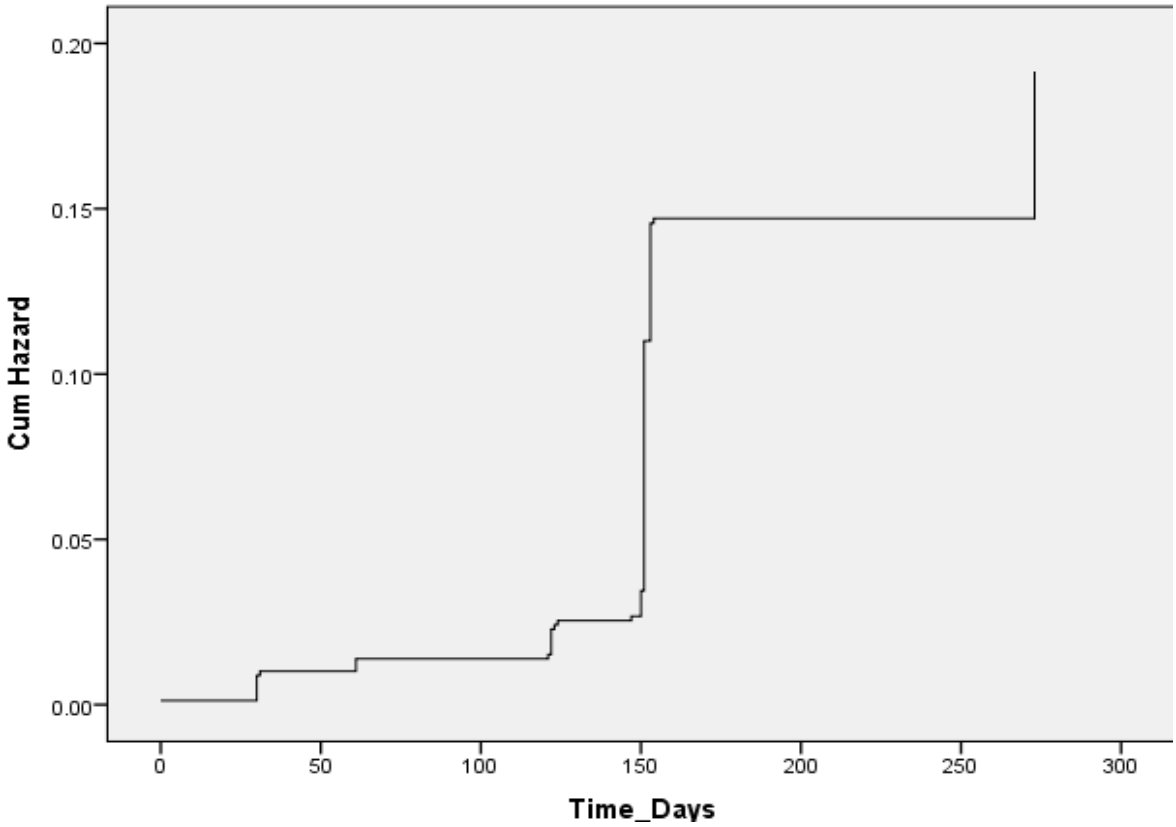

Figure.8: Cox Regression_ Hazard Function at mean of covariates

ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 4, Issue 2, pp: (257-272), Month: April - June 2016, Available at: www.researchpublish.com

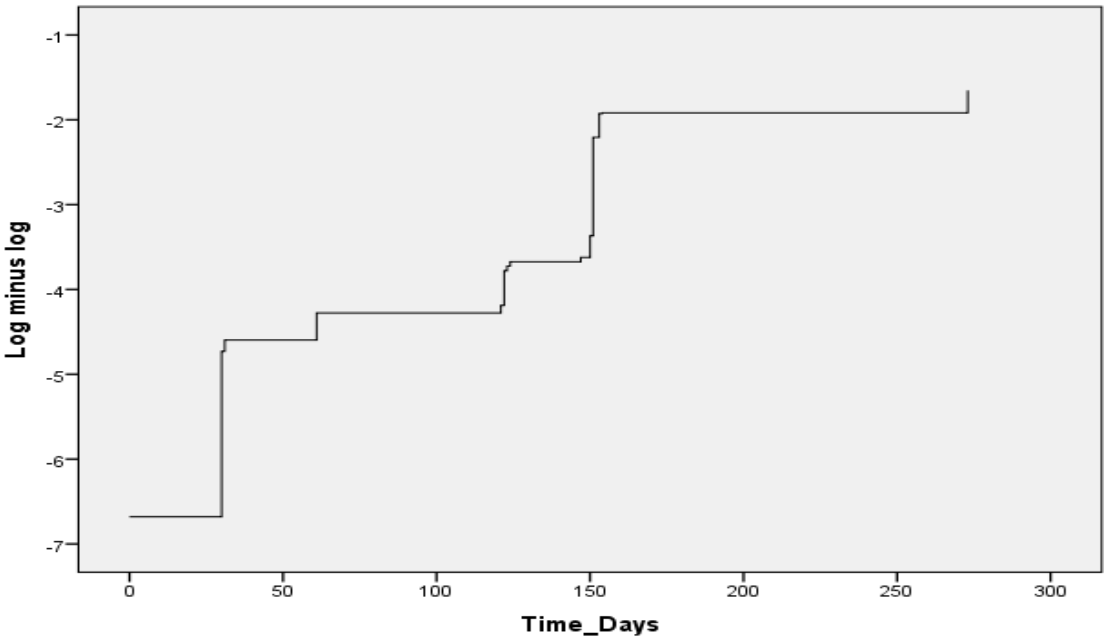

Figure.9: Cox Regression_ LML Function at mean of Covariates

The figures 11 and 12 indicate on how the increase of interns who got jobs in consideration of the time variation. The total probability of interns who were employed is not over 0.2 (see the figure 11).

5. SUMMARY, CONCLUSION AND RECOMMENDATION

5.1 Conclusion:

The survival function shows that the probability of finding a job is varying between 1 and 0,8 it means that the probability of not finding a job during internship is still high. Normal the surviving probability is tending to zero and slop of the figure is falling down in relation of time variation, at the end of event, the probability of surviving is about zero, it implies to the big number of censored cases.

For this case, the event is closes at the probability of 0.8 i.e the number of interns who got job during internship program is small based on the number of interns in our study; it is very indicative in the one minus survival function where the probability of being censored is not great the then 0,2. This findings is the same for the analysis by using Kaplan Meier and cox regression models

From above findings there is an effectiveness of internship to the fresh graduates who are in the system but the effectiveness is still low because of the high probability of surviving at the end of the event.

The two models; Kaplan Meier and Cox regression are giving the approximately results, the censoring/drop out of interns in the internship program

5.2 Recommendations:

The objective of the internship program for flesh graduates is on job training that has to increase experience of working in their specialized areas that is a tool of increasing competiveness to the labor market and the probability of finding job. Form above findings we propose recommendation as follows:

1. All sectors/institutions has to increase the hosting of interns because it is a time for increase an experience and application of the theory learnt during academic courses (requesting more interns in NCBS).
2. The supervision for interns has to be reinforced because of the small probability of finding a job.
3. Interns have to continue and strengthen the program with considering the supervision that we help them to increase their experience during job training.
4. The further research beyond the period of on job training has to be conducted for more findings (post internship assessment) to the effectiveness of internship in Rwanda for flesh graduates.

REFERENCES

[1] Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis. Part I: basic concepts and first analyses.Br J Cancer. 2003; 89:232–238.

[2] Clark TG, Stewart ME, Altman DG, Gabra H, Smyth J. A prognostic model for ovarian cancer. Br J Cancer.2001;85: 944–952.

[3] Collett D. Modelling Survival Data in Medical Research. London: Chapman and Hall/CRC; 1994.

[4] Cox DR. Regression models and life tables (with discussion) J R Statist Soc B. 1972;34:187–220.

[5] Daniel Jackel, EVALUATING THE EFFECTIVENESS OF AN INTERNSHIP PROGRAM, Dec. 2011

[6] Kay R, Kinnersley N. On the use of the accelerated failure time model as an alternative to the proportional hazards model in the treatment of time to event data: a case study in influenza. Drug Inf J. 2002;36:571–579.

[7] Stare J, Heinzl H, Harrell F. On the use of Buckley and James least squares regression for survival data2000. New approaches in applied statistics: Metodološki zvezki 16

[8] Natrella M.G. "Experimental Statistics", National Bureau of Standards, Washington DC (1963)

[9] Neter J., Wasserman W. "Applied linear statistical models", R.D. Irwin Inc., Homewood, Illinois (1974)

[10] Gunst R.F., Mason R.L. "Regression analysis and its application", Marcel Dekker Inc., NewYork (1980)

[11] Shoemaker D.P., Garland C.W., Nibler J.W. "Experiments in Physical Chemistry", The McGraw-Hill Companies Inc. (1996)

[12] Wen Kein, Determinants of internship effectiveness for university students in Hong Kong, , Page 2, April 2010

[13] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2394368/ 30th August 2015

[14] http://internships.about.com/od/internshipsquestions/g/internshipdefin.htm, 6th June 2015