

**COMPARISON OF HEPATITIS-A AND HEPATITIS-B SURVIVAL FUNCTION BY  
USING LIFE TABLE SURVIVAL ANALYSIS  
(Case study at Dr. M. Djamil Hospital Padang in 2008)**

Hazmira Yozza, Maiyastri, Harges Swastika

**Abstract**

Hepatitis is an inflammation of the liver, most commonly caused by a viral infection. There are five main type of hepatitis, referred to as types A, B, C, D and E. In this paper the survivor function of hepatitis A and hepatitis B patients are compared. Data is analyzed by using Life Table Survival Analysis. In general, hepatitis B viruses had better survival experienced than hepatitis A viruses, so the hepatitis A patients will recover faster than hepatitis B patients. The median survival time was 9,29 days for hepatitis A patients and 17,36 days for hepatitis B patients.

**Keywords:** *hepatitis, life table survival analysis, survivor function, median survival time*

**INTRODUCTION**

The liver is a vital organ present in human body. It lies below the diaphragm in the thoracic region of the abdomen. This organ plays a major role in metabolism, and has a number of functions in the body, including glycogen storage, decomposition of red blood cells, plasma protein synthesis, hormone production, and detoxification. It produces bile, an alkaline compound which aids in digestion, via the emulsification of lipids. It also performs and regulates a wide variety of high-volume biochemical reactions requiring highly specialized tissues, including the synthesis and breakdown of small and complex molecules, many of which are necessary for normal vital. This organ is necessary for survival; there is currently no way to compensate for the absence of liver function.

There are many diseases affecting the liver. One of those diseases is hepatitis. Hepatitis (plural hepatitides) implies injury to the liver characterized by the presence of inflammatory cell in the tissue of the organ. The condition can be self-limiting, healing on its own, or can progress to scarring of the liver. Hepatitis is acute when it lasts less than six months and chronic when it persists longer.

Hepatitis is most commonly caused by a viral infection. There are five main hepatitis viruses, referred to as types A, B, C, D and E. Hepatitis A and E are typically caused by ingestion of contaminated food or water. Hepatitis B, C and D usually occur as a result of parenteral contact with infected body fluids (e.g. from blood transfusions or invasive medical procedures using contaminated equipment). Hepatitis B is also transmitted by sexual contact. Hepatitis can also be due to toxins (notably alcohol), other infections or from autoimmune process.

The survival of hepatitis patients is interesting to be observed. This survival is described by recovery time of hepatitis patients. In this paper, we will construct the survivor function and hazard function of hepatitis A and hepatitis B patients that are cured at Dr. M.Djamil Hospital in Padang. Then, we will compare the resulting survivor function. Statistical analysis used is the survival analysis by using life table method.

**SURVIVAL ANALYSIS**

Survival analysis is a collection of statistical procedure for data analysis for which the outcome variables of interest is time until an event occurs. In survival analysis, we usually refer to the time variable as survival time, because it gives the time that an individual has survived over some follow-up period. It can be years, months, weeks or days from the beginning of the follow-up of an individual until an event occur.

We also typically refer to the event as a failure because the event of interest usually death, disease incidence or some other negative experience of interest that may happen to an individual.

However, survival time may be “time to return to work after an elective surgical procedure” in which case failure is a positive event.

Most survival analyses must consider a key analytical problem called censoring. In essence, censoring occurs when we have some information about individual survival time, but we don't know the survival time exactly. There are generally three reasons why censoring may occur: (1) the person does not experience the event before the study ends. (2) a person is lost to follow-up during the study period, and (3) a person withdraws from the study because of death (if death is not the event of interest) or some other reasons (eg., adverse drug reaction or other competing risk).

There are two fundamental functions considered in any survival analysis, i.e. the survivor function, denoted by  $S(t)$  and the hazard function, denoted by  $h(t)$ . The survivor function  $S(t)$  gives the probability that an object survives longer than some specified time  $t$ . As  $T$  denotes survival time and  $t$  is specified value of  $T$ , the survivor function can be expressed as :

$$S(t) = P(T > t)$$

Theoretically,  $t$  ranges from 0 to infinity and the survivor function can be graphed as a smooth-non decreasing curve as illustrated below. At time  $t = 0$ ,  $S(t) = S(0) = 1$ , that is, at the start of the study, since no one has gotten the event yet, the probability of surviving past time 0 is one. At time  $t = \infty$ ,  $S(t) = S(\infty) = 0$ ; that is, theoretically, if the study period increased without limit, eventually nobody would survive, so the survivor curve must eventually fall to zero.

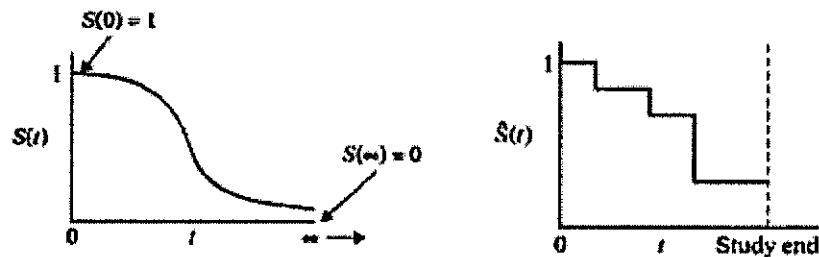


Figure 1. The Survivor Function (a) theoretical  $S(t)$  (b)  $\hat{S}(t)$  in practice

In practice, when using actual data, we usually obtain graph that are step functions, as illustrated here, rather than smooth curve. Moreover, because the study period is never infinite in length and there may be competing risks for failure, it is possible that not everyone studied gets the event. The estimated survivor function, denoted by a caret over the  $S$  in the graph, thus may not go all the way down to zero at the end of the study.

Also notice from the graph that one can obtain estimates of the median of survival time. The median of survival time is defined as the 50<sup>th</sup> percentile, which is the value of  $t$  at  $S(t) = 0.50$ . Graphically, the median is obtained by proceeding horizontally from the 0.5 point on the Y-axis until the survivor curve is reached, and then proceeding vertically downward until the X-axis is crossed at the median survival time.

The hazard function is sometimes called a conditional failure rate. It gives the probability that a person's survival time,  $T$ , will lie in the time interval between  $t$  and  $t + \Delta t$ , given that the survival time is greater than or equal to  $t$ . This function is formulated as :

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t}$$

This mathematical formula is difficult to explain in practical term

As with a survivor function, the hazard function  $h(t)$  can be graphed as  $t$  ranges over various values. The figure bellow illustrates three different hazard function. In contrast to a survivor function, the graph of  $h(t)$  does not have to start at 1 and go down to zero, but rather can start anywhere and go up and down in any direction over time. In particular, for a specified value of  $t$ , the hazard function  $h(t)$  has the following characteristics : (1) it is always non negative (2) it has no upper bound.

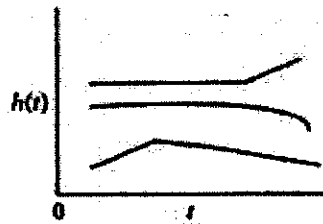


Figure 2. The Hazard Function

**Life Table Survival Analysis**

There are several technique for estimating the distribution of survival times from a sample, especially sample with censored data. The most straightforward way to describe the survival in a sample is to compute the *Life Table*. The life table technique is one of the oldest methods for analyzing survival data. This table can be thought of as an "enhanced" frequency distribution table. The distribution of survival times is divided into smaller time intervals. For each interval we can then compute the number and proportion of cases or objects that entered the respective interval "alive," the number and proportion of cases that failed in the respective interval (i.e., number of terminal events, or number of cases that "died"), and the number of cases that were lost or censored in the respective interval. For each interval, all people who have been observed at least that long are used to calculate the probability of a terminal event occurring in that interval. The probabilities estimated from each of the intervals are then used to estimate the overall probability of the event occurring at different time points.

The column in a life table is described below.

1. *Interval Start Time*. The first column gives the interval which the survival times and times to loss or withdrawal are distributed. The interval is from  $t_i$  up to but not including  $t_{i+1}$ ,  $i = 1, 2, \dots, s$ . The last interval has an infinite length and these length are assumed to be fixed.
2. *Number Entering Interval*. This is the number of objects who are observed in each interval. It is denoted as  $O_i$  and computed as :

$$O_i = O_{i-1} - W_{i-1} - d_{i-1}$$

$W_i$  and  $d_i$  will be explained later.

3. *Number Withdrawing During Interval* is denoted as  $W_i$  and defined as the number of censored data in the respective interval.
4. *Number Of Terminal Events* is denote by  $d_i$  and defined as the number of objects that experienced the event in the respective interval.
5. *Number of Cases at Risk*. This is defined as the number of objects who are exposed to risk in the respective interval. It is assumed that the time to loss or withdrawal are approximately uniformly distributed in the interval. Therefore, objects lost or withdrawn in the interval (or censored objects) are exposed to risk of experienced the event for one-half the interval. This is

denoted by  $r_i$  and computed by :

$$r_i = O_i - \frac{W_i}{2}$$

6. *Proportion Terminating*. This is defined as the propotion of objects that experience the event. It is denoted by  $q_i$ , and computes as the ratio of the number of terminal event in the respective interval, divided by the number of cases at risk in the interval, as follows :

$$q_i = \frac{d_i}{r_i}$$

7. *Proportion Surviving*. This is defined as the proporsion of cases that is survive at any interval. It is denoted by and calculated by :

$$p_t = 1 - q_t$$

8. *Cumulative Proportion Surviving*. This is an estimate of the survivors function at time  $t$ ; it is often referred to as the cumulative survival rate. Since the probabilities of sur-vival are

assumed to be independent across the intervals, this probability is computed by multiplying out the probabilities of survival across all previous intervals,

$$P_t = P_{(t-1)} p_t (P_0 = 1 \text{ and } P_1 = p_1).$$

This cumulative proportion is used to estimate the *survival function*  $S(t)$ .

9. *Standard Error Of Cumulative Proportion Surviving*. This is a standard error of cumulative proportion surviving at the end of interval. It is denoted as  $SE(P_t)$  and computed by :

$$SE(P_t) = P_t \sqrt{\sum_{j=1}^t \left[ \frac{q_j}{r_j p_j} \right]}$$

10. *Probability Density*. This is defined as the probability of failure in the respective interval, computed per unit time, that is :

$$f_t = \frac{d_t}{l O_t}$$

( $l$  is interval width)

11. Standard deviation of probability density. This denoted as  $SE(f_t)$  and computed as :

$$SE(f_t) = \frac{p_t q_t}{l} \sqrt{\left[ \sum_{j=1}^{t-1} \frac{q_j}{r_j p_j} \right] + \frac{p_t}{r_t q_t}}$$

For first interval,

$$SE(f_1) = \frac{p_1 q_1}{l} \sqrt{\frac{p_1}{r_1 q_1}}$$

12. *Hazard Rate*. The hazard rate (denoted as  $h_t$ ) is defined as the probability per time unit that a case that has survived to the beginning of the respective interval will fail in that interval. Specifically, it is computed as the number of failures per time units in the respective interval, divided by the average number of surviving cases at the mid-point of the interval, mathematically, it can be expressed as :

$$h_t = \frac{2 q_t}{l(1+p_t)}$$

13. *Standard Deviation of Hazard Rate* is denoted as  $SE(h_t)$  and formulated as :

$$SE(h_t) = \frac{h_t^2}{\sqrt{r_t q_t}} \left\{ 1 - \left( \frac{h_t l}{2} \right)^2 \right\}$$

If  $q_t = 0$  then  $SE(f_t)$  and  $SE(h_t)$  is assumed to has value equal to 0.

## DATA AND METHODS

### Data

Data used in this paper are obtained from 93 hepatitis A and 15 hepatitis B patients at M.Djamil Hospital, Padang, which were cured between Januari – December 2008. Patient is said having an event when he is allowed to be treated at home after being hospitalized. The survival time is defined as the recovery time, i.e the duration (in days) from a patient being hospitalized until the event happened.

The case failure in this research is a positive event. In negative sense, we can view the event as the death of hepatitis virus and the survival time as the time that the hepatitis virus has survived in patients body over some follow-up period. A patient is said being censored if he is out of the hospital by his own decision, before properly cured.

### Methods

The method used for this research is Life Table Survival Analysis. The steps in using this method are as follows :

1. Constructing Life Table, for hepatitis A and hepatitis B patients.
  - a. Determining the *interval start time*

- b. Computing the number of patient at interval  $t$  ( $O_t$ ) at column 2.
  - c. Computing the number censored patients at interval  $t$  ( $W_t$ ), at column 3
  - d. Computing the number of patient that is allowed to be treated at home ( $d_t$ ) at column 4
  - e. Computing the number of patients exposed to risk at interval  $t$  ( $r_t$ ) at column 5
  - f. Computing proportion of patients considered had properly cured ( $q_t$ ) at column 6
  - g. Computing proportion of patients that still continued the treatment at interval  $t$  ( $p_t$ ) at column 7.
  - h. Computing the cumulative proportion of patients that still continued the treatment at the end of interval  $t$  ( $P_t$ ) and its standard error at column 8 and 9.
  - i. Computing density probability ( $f_t$ ) and its standard error at column 10 and 11.
  - j. Computing hazard rate ( $h_t$ ) and its standard error at column 12 and 13.
- 2.. Constructing, interpreting and comparing resulted survivor and hazard function.
  3. Computing the median survival time hepatitis A and hepatitis B patients. In this research, we considered the median survival time as the median of treatment duration of patients.

## RESULT AND DISCUSSION

In general, most of hepatitis A patients that are cured in M.Djamil Hospital during Januari – December 2008 are male (57%) and most of hepatitis B patients are female (66.7%). Most of that patients are about 15 – 49 years old.

The life table of Hepatitis A is constructed based on hepatitis A patients' data. The interval width is 3 days, so, we have 15 interval for 45 days of observations. There are 93 patients observed. The life table is shown in Table 1.

Table 1 Life Table of Hepatitis A

$T$	Interval Start Time	$O_t$	$W_t$	$d_t$	$r_t$	$q_t$	$p_t$	$P_t$	SE ( $P_t$ )	$f_t$	SE ( $f_t$ )	$h_t$	SE ( $h_t$ )
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1	0	93	3	1	91.5	0.011	0.989	0.989	0.980	0.004	0.004	0.004	0.000
2	3	89	1	9	88.5	0.102	0.898	0.888	0.930	0.034	0.035	0.036	0.010
3	6	79	2	32	78	0.410	0.590	0.524	0.550	0.122	0.128	0.172	0.030
4	9	45	0	21	45	0.467	0.533	0.279	0.300	0.082	0.087	0.203	0.040
5	12	24	1	13	23.5	0.553	0.447	0.125	0.130	0.052	0.056	0.255	0.070
6	15	10	0	4	10	0.400	0.600	0.075	0.090	0.017	0.019	0.167	0.080
7	18	6	0	3	6	0.500	0.500	0.037	0.050	0.012	0.015	0.222	0.120
8	21	3	0	1	3	0.333	0.667	0.025	0.040	0.004	0.006	0.133	0.130
9	24	2	0	0	2	0.000	1.000	0.025	0.040	0.000	0.000	0.000	0.000
10	27	2	0	1	2	0.500	0.500	0.012	0.020	0.004	0.008	0.222	0.210
11	30	1	0	0	1	0.000	1.000	0.012	0.020	0.000	0.000	0.000	0.000
12	33	1	0	0	1	0.000	1.000	0.012	0.030	0.000	0.000	0.000	0.000
13	36	1	0	0	1	0.000	1.000	0.012	0.030	0.000	0.000	0.000	0.000
14	39	1	0	0	1	0.000	1.000	0.012	0.030	0.000	0.000	0.000	0.000
15	42	1	0	0	1	0.000	1.000	0.012	0.030	0.000	0.000	0.000	0.000

The Table 2 shows the life table for hepatitis B patients, where we use 11 3-days intervals, smaller than the number of intervals used to construct hepatitis A patient life table. Initially, there are 15 patients observed.

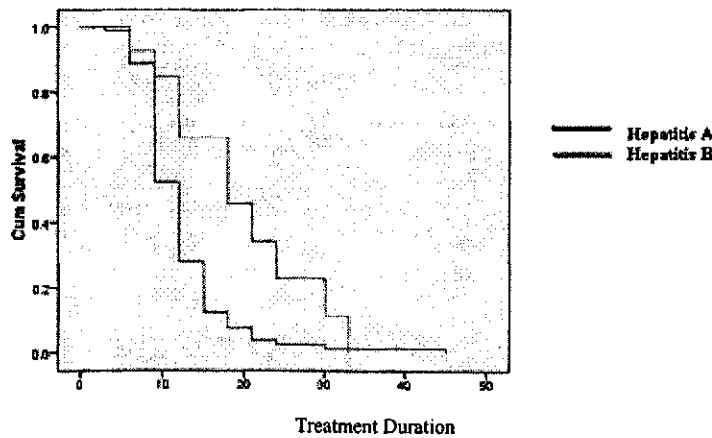
**Table 2. Life Table of Hepatitis B**

$t$	Interval Start Time	$O_t$	$W_t$	$d_t$	$r_t$	$q_t$	$p_t$	$P_t$	SE ( $P_t$ )	$f_t$	SE ( $f_t$ )	$h_t$	SE ( $h_t$ )
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1	0	15	1	14.5	0	0.000	1.000	1.000	1.000	0.000	0.000	0.000	0.000
2	3	14	0	14	1	0.071	0.929	0.929	1.290	0.024	0.033	0.025	0.020
3	6	13	3	11.5	1	0.087	0.913	0.848	1.430	0.027	0.045	0.030	0.030
4	9	9	0	9	2	0.222	0.778	0.659	1.190	0.063	0.113	0.083	0.060
5	12	7	0	7	0	0.000	1.000	0.659	1.250	0.000	0.000	0.000	0.000
6	15	7	1	6.5	2	0.308	0.692	0.457	0.910	0.068	0.135	0.121	0.080
7	18	4	0	4	1	0.250	0.750	0.342	0.740	0.038	0.083	0.095	0.090
8	21	3	0	3	1	0.333	0.667	0.228	0.530	0.038	0.088	0.133	0.130
9	24	2	0	2	0	0.000	1.000	0.228	0.560	0.000	0.000	0.000	0.000
10	27	2	0	2	1	0.500	0.500	0.114	0.290	0.038	0.097	0.222	0.210
11	30	1	0	1	1	1.000	0.000	0.000	0.000	0.038	0.097	0.667	0.000

Now, we will re-write the cumulative proportion of patients still continue the treatment (column 8) for Table 1 and Table 2. This is the estimate of survivor function for both group of patients.

**Table 3. Comparison of Survivor Functions**

Interval	Cumulative Proportion of Patient Still Continue the Treatment (S(t))	
	Hepatitis A	Hepatitis B
1	0.989	1.000
2	0.888	0.929
3	0.524	0.848
4	0.279	0.659
5	0.125	0.659
6	0.075	0.457
7	0.037	0.342
8	0.025	0.228
9	0.025	0.228
10	0.012	0.114
11	0.012	
12	0.012	
13	0.012	
14	0.012	
15	0.012	



**Figure 2. Survival Curves of Hepatitis Patients**

The Table 3 shows that, in general, the survival probability surviving of hepatitis B virus is greater then the survival probability of hepatitis A virus at all interval. On patients point of views, it means that the recovery time of hepatitis B patients is longer then the recovery time of hepatitis A patients. These survival functions of hepatitis A and B patients are plotted in the Figure 2.

As from the Table 3, this Figure 2 gives the same interpretation. The survival curves clearly show that the survival probability of hepatitis B virus is greater than of hepatitis virus. The hepatitis B virus had better survival experience than hepatitis A virus, therefore, hepatitis A patients will recover sooner then hepatitis B patients. For example, 50% of hepatitis A viruses survived least 10 days, while about 82% of viruses B patients survived that long. It also means that 50% of hepatitis A patients will recover more then 10 days while about 82% of hepatitis B will recover more then 10 days. The survival curve also shows that 20% of hepatitis B patients still continued the treatment for more than 30 days, while only 2% of hepatitis A patient still continued the treatment that long.

In this paper, we refer the median of survival time as the median of viruses survival time. On the patients point of view, we also refer the median of survival time as the median of recovery time of hepatitis patients. This median can be estimated by linear interpolation. This following table shows the median of survival time.

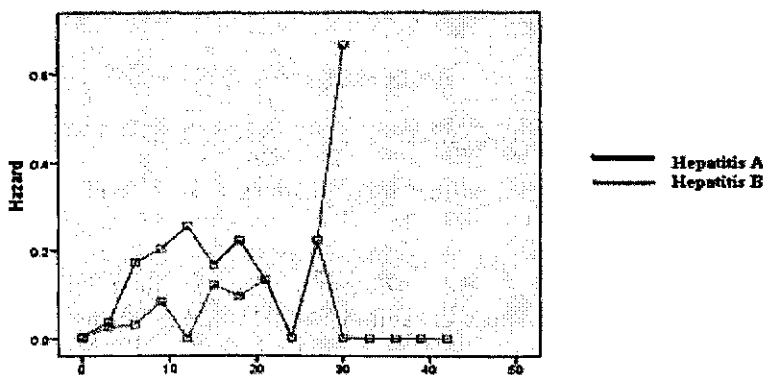
**Table 4. The Median of Survival Time**

Type of Hepatitis	Median Survival Time (days)
A	9.29
B	17.36

It can be seen that the median survival time for hepatitis A is 9.29 days. It means that 50% of hepatitis A patient followed the treatment for more than 9 days. For hepatitis B, the median survival time is 17.36 days. It means that 50% of hepatitis B patients followed the treatment for more than 17 days.

The hazard rate is the opposite of the survival rate. The higher the probability surviving, the lower the hazard rate. The hazard rate is defined as the probability per unit of time that viruses had survived to the beginning of the respective interval of days will fail in that interval. On the patients point of view, the hazard rate is defined as the probability that a patient will recover at interval  $t$ , given that he is still being hospitalized in the beginning of respective interval. To get clear interpretation, let us refer hazard rate as conditional death rate (from virus view of point) or conditional recovery rate (from the patients view of view).

The estimated hazard functions of hepatitis patients (or viruses) is described in following figure. The hazard function are plotted at the midpoint of interval.



**Figure 4. Hazard Curve of Hepatitis Patients**

The estimated hazard function for hepatitis A shows that the death rate of hepatitis A viruses (or the recovery rate of hepatitis A patients) is low in the beginning of observation and increases up to 5<sup>th</sup> interval. From the beginning of 5<sup>th</sup> interval, this rate tends to decrease until the end of study.

The estimated hazard function for hepatitis B shows that the death rate of hepatitis A viruses (or the recovery rate of hepatitis B patients) is also low in the beginning of observation and tends to increase until the end of study.

## CONCLUSION

Finally, we can conclude that, in general, the hepatitis B viruses had better survival experience than the hepatitis A viruses. It also means that treatment duration of hepatitis B patients is longer than hepatitis A patients. The median survival time of hepatitis A is 9.29 days, while the median survival time of hepatitis B patients is 17.36 days. It means that 50% of hepatitis A patients followed the treatment more than 9 days while 50% of hepatitis B patients followed the treatment for more than 17 days.

## REFERENCE

- Colton, T. 1974. *Statistika Kedokteran Terj. Sanusi, R.* Gadjah Mada University Press, Yogyakarta
- Dalimartha, S. 2002. *Seri Agrisehat : Ramuan Tradisional Untuk Pengobatan Hepatitis.* Penebar Swadaya, Jakarta
- Gehan, E. A. 1975. *Statistical methods for survival time studies In: Cancer Therapy: Prognostic Factors and Criteria, M. J. Staquet, eds.* Raven Press (Menu Help SPSS 16.0), New York
- Kleinbaum, D.G and M.Klein. 2005. *Survival Analysis : A Self-Learning Text 2<sup>nd</sup> edition.* Springer Science+Business Media, Inc., New York
- Lee, E.T. 1992. *Statistical Methods for Survival Data Analysis 2<sup>nd</sup> edition.* Jhon Wiley & Son, New York
- Menggang, Y. 2004. Introduction to survival analysis. <http://www.biostat.iupui.edu/ShortCourse/survival.pdf>, 23 Mei 2007
- Noor, N. N. 1997. *Pengantar Epidemiologi Penyakit Menular.* Rineka Cipta, Jakarta
- Notoatmodjo, S. 2003. *Prinsip – Prinsip Dasar Ilmu Kesehatan Masyarakat.* Rineka Cipta, Jakarta
- Palilingan, R. N. 2006. *Aplikasi Analisis Kesintasan (Survival Analysis) Menggunakan Tabel Kehidupan Program Doktor Studi Ilmu Kedokteran.* Universitas Udayana, Bali
- Rizam, M., Abu Bakar dan E. D. Supandi. *Pendugaan Selang Tahan Hidup dari Data Uji Hidup Berdistribusi Pareto di Bawah Sensor Tipe II dengan Metode Bayesian.* *Jurnal Statistika : Forum Teori dan Aplikasi Statistika.* 3 : 146 -147
- Sastroasmoro. S. dan S. Ismael. 2002. *Dasar-Dasar Metodologi Penelitian Klinis.* CV Sagung Seto, Jakarta
- Sulaiman, A., Daidiyono, N. Akbar and A.A. Rani. 1990. *Gastroenterologi Hepatologi.* CV. Sagung Seto, Jakarta
- Yasril and H. S. Kasjono. 2009. *Analisis Multivariat Untuk Penelitian Kesehatan.* Mitra Cendikia, Yogyakarta