

PAPER • OPEN ACCESS

Survival analysis with the Cox Proportional Hazard Method to determine the factors that affect how long the Large-Scale Social Distancing (LSSD) will applied in various areas affected by the covid-19 pandemic

To cite this article: A A Fauziah *et al* 2021 *J. Phys.: Conf. Ser.* **1751** 012004

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

Survival analysis with the Cox Proportional Hazard Method to determine the factors that affect how long the Large-Scale Social Distancing (LSSD) will applied in various areas affected by the covid-19 pandemic

A A Fauziah¹, Budiman², D L Safitri³, A Meiza⁴

^{1,2,3,4}Department of Mathematics, Faculty of Science and Technology, Universitas Islam Negeri Sunan Gunung Djati, Bandung, Indonesia

email: ajeng.08maret@gmail.com¹, budimanbudi126@gmail.com², dianlinda.safitri88@gmail.com³, asti.meiza@uinsgd.ac.id⁴

Abstract. This study will determine the length of time for large-scale social restrictions, in Indonesia is known as the Large-Scale Social Distancing (LSSD), which requires in a region to reduce the number of different covid-19 cases. Associated with the implementation of the LSSD, it turns out that many things can influence the LSSD to be able to run effectively in the community, from internal and external factors. This research was conducted for 15 days from 1 June 2020 to 15 June 2020, to find out the factors that had a significant influence on the effectiveness of LSSD using Cox Proportional Hazard Regression. The dependent variable used in the study is the length of time the LSSD system (Y). From the results can concluded that the socialization variable (X_1) has a significant effect to the effectiveness of LSSD.

Keywords: LSSD, cox proportional hazard, covid-19, survival analysis.

1. Introduction

During the Covid-19 pandemic, there were various methods from several regions that implemented a regulation to inhibit and suppress the growth of the Covid-19 virus, such as requiring washing hands before entering community service sites, requiring wearing masks, physical distancing, and LSSD. LSSD rules are actually not much different from social distancing, which distinguishes only the movement of residents outside their homes is limited so that the number of positive cases of corona virus decreases. LSSD only gives affirmations, such as the closure of places of worship, restricted religious activities, restricted public transportation operations, schools and workplaces are closed [8].

In PMK Regulation Number 9 of 2020 article 2, that to be stipulated as a LSSD, a province / district / city region must meet two criteria. First, namely the number of cases or deaths due to disease increased and spread significantly quickly to several regions. While the second criterion is that the area with the disease also has epidemiological links with similar events in other regions or countries. From these two criteria, the Minister of Health can determine whether the area or region is suitable for LSSD implementation or not [3].



Related to the implementation of LSSD, it turns out there are many things that can affect the duration of the implementation of LSSD both internal and external factors. Based on the description above, we are interested to know the significant influence of these factors in the effective LSSD in areas affected by the corona virus.

2. Survival Analysis

Survival analysis has become an important tool for analyzing time to event data or analyzing time-related data, from time origin to the occurrence of a specific event. Special events (failure events) can be in the form of failure, death, recurrence of a disease, the response of an experiment, or other events chosen in accordance with the interests of researchers. These special events can be positive events such as birth, school graduation, recovery from an illness (Kleinbaum & Klein, 2005: 4)[6].

2.1. Data survival

Survival data is data about the observation period of time from the beginning of the observation to the occurrence of an event. Survival time can be defined as the time from the start of observation to the occurrence of a failed event, it can be in days, months or years. The time of origin (time origin or start-point) is the time at the time of the initial occurrence, such as when a person is sentenced to suffer cancer, the time of treatment and others. Failure time (failure time or end-point) is the time at the time of the final event such as death, events and others (Collett, 2003: 1)[1].

2.2. Types of Censorship

According to Klein & Moeschberger (2003: 64-70) in survival analysis there are four types of censorship[7].

2.2.1. Right censoring

Censorship occurs if the observed object or individual is still alive at a specified time. In other words the individual has not experienced events until the end of the observation period, while the initial time of the object of observation can be observed in full. For example, a cancer patient observed from the beginning of the treatment to the end of the treatment turns out that the patient is still alive. Then the patient continued treatment abroad so that it could not be observed again (lost to follow up). This patient has a survival time of at least some time. So that the individual observation time is said to be right censorship.

2.2.2. Left censoring

Left censorship occurs when all the desired information known from an individual has been obtained at the beginning of the observation. In other words at the time of initial observation the individual is not observed at the beginning of the observation while the incident can be observed in full before the research ends. For example, in a study to determine the distribution of cannabis users among boys in a school. By asking the question "when did you first use marijuana?". It turned out that there were some children who answered "I have never used it, but I don't know exactly when I first used it", in this case the child experienced left censorship.

2.2.3. Interval censoring

Censoring interval occurs if the required information can already be known on the events in the observation interval or censorship whose endurance time is within a certain interval. For example, some mice are given carcinogens in their food, a 10-month study of 10 mice and research is conducted at the end of the year, if 2 out of 8 mice die of cancer in the 5th and 7th months, then the two mice undergo censorship hose.

2.2.4. Random censoring

Random censorship occurs if the individual observed dies or experiences an event due to another reason, not due to the main purpose of the study. For example, 10 mice were given carcinogens in their food.

At the time of observation there was 1 in 10 of these mice died due to being pinched (died not because of the main research) not due to cancer, so the mice experienced random censorship.

2.3. Survival Functions

The density function of opportunity from endurance time T is defined as the opportunity of an individual who fails at intervals of t to Δt notified by f(t). This function is formulated as follows [5]:

$$\begin{aligned}
 f(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(\text{the individual fails in interval } t + \Delta t)}{\Delta t} \\
 f(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T(t + \Delta t))}{\Delta t} \\
 f(t) &= \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \tag{1}
 \end{aligned}$$

T is a non negative random variable in the interval [0,∞), F(t) is a cumulative distribution function of T so that this function can be defined as an opportunity for an individual to experience events up to time t:

$$\begin{aligned}
 F(t) &= P(T \leq t) \\
 F(t) &= \int_0^t f(x)dx \tag{2}
 \end{aligned}$$

From equation (2) by decreasing dt on both sides obtained:

$$F(t) = D_t \left(\int_0^t f(x)dx \right) = f(t) \tag{3}$$

2.3.1. Survival Function

The Survival function explains the size of time an event occurred. This endurance function can be defined as the opportunity for students to survive in continuing their studies for the period t stated in the form [9]:

$$S(t) = P(\text{An individual survives } > t = P(T \geq T) = \int_t^\infty f(u)du \tag{4}$$

2.3.2. Hazard Function

Hazard function h(t) from the time endurance T is referred to as conditional failure rate which is defined as the chance of failure that the individual has survived for a time t. Hazard function stated as follows [9]:

$$\begin{aligned}
 h(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(\text{the individual fails at intervals}(t, t + \Delta t))}{\Delta t} \\
 h(t) &= \lim_{\Delta t \rightarrow 0} \frac{P\{(t \leq T < t + \Delta t | T > t)\}}{\Delta t} \\
 h(t) &= \frac{f(t)}{S(t)} \tag{5}
 \end{aligned}$$

Relation between S(t) and h(t) is if value S(t) raise up then the value h(t) reduce, and vice versa if value S(t) reduce then the value of h(t) raise.

2.4. Cox Proportional Hazard Regression Model

Cox Regression Model introduced by D.R Cox in the year 1972 and first applied to survival data. Cox Proportional Hazard Regression Model used to determine the relationship between the dependent variable with the independent variable, where the data used on Cox Proportional Hazards Regression in the form of survival data from an individual. According to Collett (2004), Cox Proportional Hazards Regression Model is as follows [2]:

$$h(t, X) = h_0(t). (\exp (\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)) \tag{6}$$

with

h(t, X) : the length of time the LSSD system

h₀(t) : Basic hazard function or hazard function at the time t = 0 does not depend on

characteristics
 X : predictor or explanatory variable
 p : (p_1, p_2, \dots, p_n) the sum of the explanatory variables X
 B : $(\beta_1, \beta_2, \dots, \beta_n)$ regression coefficient vector or parameter vector
 Hazard function modeling in Cox Regression analysis is as follows:

$$h(t) = h_0(t) \exp(y) \tag{7}$$

2.5. Kaplan-Meier Estimation

The method used to describe the survival of a random sample t_0, \dots, t_n namely by depicting graphs of survival functions or empirical distribution functions by Kaplan-Meier estimation. In addition it also provides estimates of non-parametric distribution.

Be given $(t'_i, \delta_i), i = 1, \dots, n$ which states a censored random sample, with $\delta_i=1$ is observable data and $\delta_i = 0$ is censored data. Suppose there is $k(k \leq n)$ with a different time $t_1 < t_2 < \dots < t_k$ which states how much data was observed. The likelihood that one or more of the events will be observed will be denoted as $d_j = \sum I(t'_i = t_j, \delta_i = 1)$ or state the number of events observed at the time t_j . Estimate of $\hat{S}(t)$ can be defined as follows [5]:

$$\hat{S}(t) = \prod_{t_j < t} \frac{n_j - d_j}{n_j} \tag{8}$$

With $n_j = \sum I, t_i \geq t_j$ is the number of individuals who are at risk at the time t_j in other words the number of individuals who have not experienced an event or event and were not censored before at the time t_j .

2.6. Assumption Testing of Proportional Hazard

How to visually examine Proportional Hazard assumptions by looking at graphics from the Log plot $\{-\log[S(t,x)]\}$ towards survival time. Each plot between categories in one explanatory variable looks parallel or not intersect, so the proportional hazard assumption is fulfilled [7].

3. Research Method

3.1. Time and Location

This research was conducted on 1-15 June 2020. The research was administered with online media, namely filling out questionnaires via Google form. The place of research was centered on areas that implemented of the LSSD that were Bogor, Karawang, Tasikmalaya, Depok, Bandung, Sumedang, Jakarta Selatan, Bekasi, Garut, Sukabumi, Indramayu, and Tangerang.

3.2. Data Source

The data used in this study are primary data. Primary data obtained from the results of filling out the questionnaire using Google Form with a total of 34 subjects.

3.3. Research Subjects

Research subjects is the community whose area is affected by the implementation of the LSSD which is elderly who has a high risk.

3.4. Research Variables

The dependent variable used in the study is the length of time the LSSD system (Y). While the independent variables that are thought to be factors that influence the effectiveness of the LSSD program that is implemented in the area of each subject include [8]:

- Socialization (X_1)
- Material assistance from the government (X_2)
- Intensity to go out (X_3)
- LSSD knowledge (X_4)
- Go out (X_5)

Obedience (X_6)

4. Results and Discussion

4.1. Data Distribution

Table 1. Number of surveys of LSSD affected communities

Status	Frequency	Percentage
<i>Event (ineffective)</i>	7	20,58%
Sensor (effective)	27	79,42%
Total	34	100%

In Table 1 it can be seen that 20.58% or as many as 7 respondents experienced the event or the length of the LSSD system in their area due to the ineffectiveness of the LSSD system. While as many as 27 respondents or 79.42% experienced censorship or the rapid entry into force of the LSSD system in their area due to the effectiveness of the LSSD system.

Table 2. The results of a survey of people affected by LSSD based on independent variables by category

Variable	Categoric	Total
Socialization (X_1)	0. There is no socialization	8
	1. there is socialization	26
Material assistance from the government (X_2)	0. There is no help	16
	1. There is help	18
Intensity to go out (X_3)	1. 1 – 5 times	22
	2. 6 – 10 times	3
	3. > 10 times	2
	4. Every day	7
LSSD knowledge (X_4)	1. Enough to know	9
	2. Know	21
	3. very knowledgeable	4
Go out (X_5)	1. Rarely	23
	2. Sometimes	3
	3. Often	2
	4. Every day	6
Obedience (X_6)	1. Yes, always obedient	20
	2. Sometimes	7
	3. Rarely	7
The length of the LSSD time (Y)		

4.2. Data Analysis

4.2.1. Parameter Significance Test

Significance test conducted to determine the significance of the model by looking at it overall with hypothesis as follows:

$H_0 : \beta_0 = 0$ there is no influence of the independent variable on the dependent variable.

$H_1 : \beta_0 \neq 0$ there is influence of the independent variable on the dependent variable

H_0 will be rejected if and only if the value, $p_{value} \leq \alpha$, we can use $\alpha = 0,05$.

From the data analysis using data processing software, the results of the Cox Regression analysis with the Enter method are obtained as shown in Table 3.

Table 3. Omnibus Tests of Model Coefficients

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block
	Chi-square	Df	Sig.	Chi-square	df	Sig.	Chi-square
99.436	12.273	7	.092	11.327	7	.125	11.327

In testing Cox Regression using the Enter method. This model has a significance of 0.092 where it can be concluded that the model is not acceptable because of p_{value} greater than 0,05. Because the model cannot be accepted, we must therefore look for a suitable model using the Backward LR method to obtain the following results:

Table 4. Omnibus Tests of Model Coefficients

Step	-2 Log Likelihood	Overall (score)			Change From Previous Step		
		Chi-square	Df	Sig.	Chi-square	df	Sig.
1 ^a	99.436	12.273	7	.092	11.327	7	.125
2 ^b	99.738	11.866	6	.065	.302	1	.583
3 ^c	101.407	10.890	4	.028	1.669	2	.434
4 ^d	102.704	9.066	3	.028	1.297	1	.255
5 ^e	105.413	4.649	1	.031	2.708	2	.258

In Table 4 the model can be accepted in step 3 by eliminating the help variables and the LSSD knowledge gained $p_{value}=0,028$. Can be concluded that H_0 rejected, which means there are independent variables that affect the dependent variable.

4.2.2. Cox Proportional Hazard Regression Model

Based on the data analysis that has been done with the variable coefficient values are obtained as follows:

Table 5. Variables in the equation

	B	SE	Wald	Df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
Socialization	-1.199	.642	3.484	1	.049	.302		.086
Intensity to go out			3.709	2 ^a	.157			
Intensity to go out (1)	-2.041	1.772	1.327	1	.249	.130		.004
Intensity to go out (2)	.939	1.185	.629	1	.428	2.558		.251
Go out			1.443	1 ^a	.230			
Go out (1)	1.701	1.416	1.443	1	.230	5.477		.342

It can be seen in Table 5. The significance value or p-value of each variable less than 0.05 is the socialization variable with a significance value or p-value of 0.049. This means that H_0 is accepted so that the cox proportional hazard regression model is as follows:

$$H(t, X) = h_0(t)e^\alpha \tag{9}$$

with

$$\alpha = -1,199X_1 - 2,041 X_{3(1)} + 0,939 X_{3(2)} + 1,701 X_{5(1)} \tag{10}$$

$$H(t, X) = h_0(t) \cdot (\exp (-1,199X_1 - 2,041 X_{3(1)} + 0,939 X_{3(2)} + 1,701 X_{5(1)})) \tag{11}$$

4.2.3. Survival Function and Hazard Function Graphics

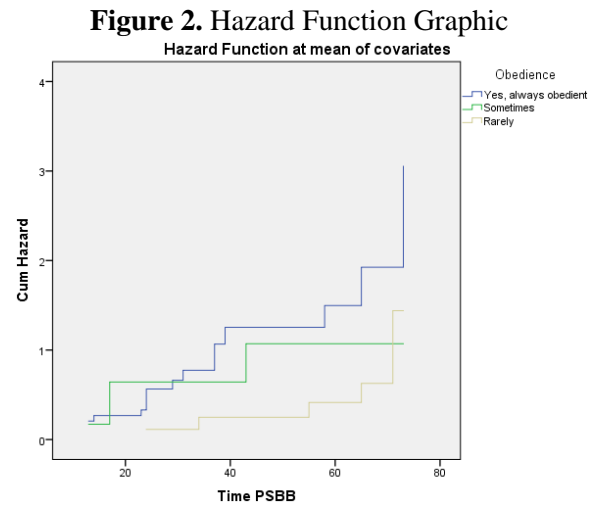
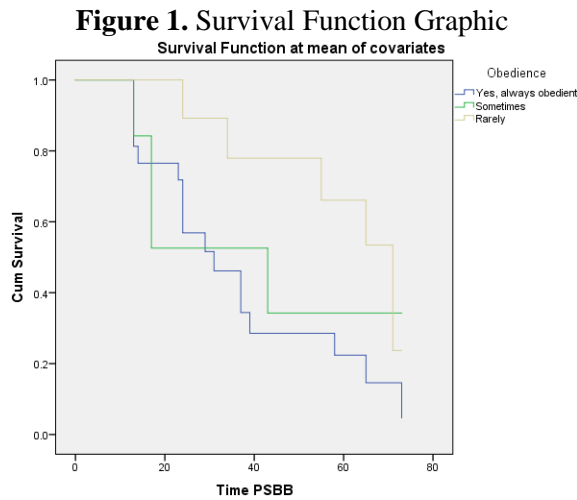


Figure 1. It is known that the obedient stratum variable with the category "Yes, always obedient" has the least chance of an event occurring. This means that the more obedient a person is, the more effective the LSSD policy will be, marked by the minimum time for implementing the LSSD.

Figure 2. It is known that the obedient stratum variable with the "Rarely" category has a lower failure rate than the other categories. This means that if a person rarely complies with the LSSD policy, the opportunity for the event to occur is even greater.

5. Conclusion

Based on the results of the analysis and discussion, a factor influencing the effectiveness of the implementation of the LSSD is a socialization where the significance value of the socialization is only 0.49, which is 0.49 smaller than the standard standard of 0.05. By using Cox Regression analysis with the Enter method the significance value of 0.092 is obtained so that the model cannot be accepted. Because it was replaced with the Backward: LR method which produced a significance level of less than 0.05, which means the model was acceptable. So a better method to use is Backward: LR. Cox Proportional Hazard Model is $h(t, X) = h_0(t) \cdot (\exp(-1,199X_1 - 2,041 X_{3(1)} + 0,939 X_{3(2)} + 1,701 X_{5(1)}))$.

References

- [1] Collett D 2003 *Modelling Survival Data in Medical Research* (US: Chapman & Hall)
- [2] Collett D 2004 *Modelling Survival Data in Medical Research* (London: CRC Press)
- [3] Damasha Kayla 2020 *Apa itu LSSD dan Apa Fungsinya Cek aja Internet* (accessed on 14 June 2020 in <https://www.cekaja.com/info/mulai-diterapkan-di-indonesia-apa-itu-LSSD-dan-apa-fungsinya/>)
- [4] Iskandar B M 2015 *Model Cox Proportional Hazard pada Kejadian Bersama* Skripsi. (Yogyakarta : Universitas Negeri Yogyakarta)
- [5] Hari N C, Hanny K, Langi Y A R, 2018 *Analisis Survival dalam Menentukan Faktor-Faktor yang Mempengaruhi Lama Studi Mahasiswa Matematika Di Jurusan Matematika FMIPA Universitas Sam Ratulangi Manado* (Manado : Universitas Sam Ratulangi)
- [6] Klein J P & Moeschberger M L 2003 *Survival Analysis: Techiques for Censored and Truncated Data Second Edition* (New York: Springer-Verlag)
- [7] Kleinbaum D G & Klein M 2005 *Survival Analysis: A Self-Learning Text Second Edition* (USA: Springer Science and Business Media, Inc)
- [8] Setyvani, Putri Gloria (2020, 23 April) *Faktor yang Mempengaruhi Kepatuhan Masyarakat Terhadap Upaya LSSD Internet* (accessed on 14 June 2020 in

- <https://www.kompas.com/sains/read/2020/04/23/163100023/9-faktor-yang-memengaruhi-kepatuhan-masyarakat-terhadap-upaya-LSSD?page=all>
- [9] Siregar, Hasanal Bolkia (2017). *Analisis Survival Dengan Metode Regresi Cox Pada Penderita Tb Paru Di Rsud Gunung Tua Kabupaten Padang Lawas Utara Tahun 2017. Skripsi* (Padang : Universitas Sumatera Utara)