**Survival Analysis of Novel Corona Virus (2019-Ncov) Using Nelson Aalen Survival Estimate**

**Abere, Omotayo Johncally**

**Department of Actuarial Science & Insurance,**

**University of Lagos, Nigeria**

**johncally68@yahoo.com**

***Abstract***

***Introduction****: Coronavirus contains virus whose genome consists of a single strand of ribonucleic acid. Novel Coronavirus is a new kind of coronavirus which originated from a Chinese city, Wuhan, in December 2019. W.H.O has officially tagged the disease caused by this coronavirus as COVID-19 because of how it spreads. It can be spread from animal to mankind or from human to human. This disease outbreak has recorded tens of thousand of infected people while over thousand people lost their lives within short period of time. Therefore, there is a need to investigate the survival pattern of this disease in order to have ideal knowledge about it and to combat its spread.*

***Objective****: The specific objective of this research is to analytically estimate the survival chances of the infected persons within specified time frame. The paper also recommended ways to prevent the disease from spreading*

***Data and Methods****: In the course of this study, different methods and strategies of investigations were conceived to gather data/materials related to the study in order to have comprehensive knowledge on the subject matter based on the flexibility of approach. The major data were got from the internet based on the surveys/reports of health or health related organizations. The two statistical tools used in this research are descriptive and inferential. Nelson Aalen survival model was used to carried out the survival estimate based on the computation arrived at using Microsoft excel package. This study analysed novel coronavirus cases from January 22 to February 10, 2020.*

***Setting****: The geographical location of this research is not limited to China where the disease broke out but explored situations in the world as a whole, although special attention was paid to cases in China.*

***Result****: From the Nelson Aalen estimate analysis carried out, the chances of survival was higher within the short time of contacting the virus. The chances reduce as time passes. The closer the estimate obtained is closer to value* ***1*** *from* ***0.5*** *upward, the better the chance of survival from the virus.*

***Conclusion****: In order to minimize the rate at which the disease spreads, this paper recommended WUHAN prevention concept. The survival chances will be maximized if responses that are aimed at mitigating exposure to the risk factors and adequate access to health services are effectively and efficiently employed. Also, this research recommended vigorous health programs to encourage lifestyles preventing this disease and more budgetary spending in affected areas should be the focus of public health actions in order to provide effective treatments*

***KEYWORDS****: Novel Coronavirus, survival estimate, WUHAN, cases, censoring, deaths.*

**1.0 BACKGROUND OF THE STUDY.**

Coronavirus contains virus whose genome consists of a single strand of ribonucleic acid. According to New York Times (2020), medical experts are concerned about cornavirus which infected/killed thousands of people. This disease belongs to the family of pathogen that does cause severe acute respiratory syndrome (SARS) which can be contacted easily by another person. This virus originated from a Chinese city, Wuhan, in December 2019. The Chinese authority gave travel ban in and out of the affected Chinese cities to manage the disease outbreak. This lockdown has been criticized as a bad move by the affected communities owing to psychological effects emanating from uneasy access to medical aid, supplies or materials(Nature Research,2020). It can spread or be contacted from animal to person or from person to person. As a result of this, the Chief Scientist at W.H.O in Switzerland, Soumya Swaminathan, has officially tagged the disease caused by this coronavirus as COVID-19 (that is, coronavirus disease 2019). This is to distinguish this disease from other coronavirus outbreak in the past or future. On February 7, 2020, the researchers at the South China Agricultural University Guangzhou, suggested in a press conference that an animal species (pangolin) is the source of the disease outbreak. This was as a result of genetic comparison of the animal and the infected person. Pangolins are animals often used in local Chinese medicine. On 30th January 2020, the W.H.O Director-General, Tedros G. Adhanom, declared this disease outbreak a global health emergency due to the fact that its spread to countries with porous healthcare facilities could be epidemic and disastrous (W.H.O, 2020a). The global health emergency declaration will gear experts up to probe the infection biology, develop drug/vaccine and weigh up worse case scenario. A study at Schenzhen in China showed that some infected persons with the virus were not exhibiting any symptoms of the disease. This makes this virus deadly as it can be contacted without knowing by someone carrying the disease without showing signs. Within a very short time of novel coronavirus (nCov) emergence, many papers/articles had been written on it. However, these were based on survey reports rather than empirical studies, and the role of explanatory factors was not adequately investigated. This study will provide answers to questions such as: What are the survival chances of the infected people? What is the time frame of the survival probability? How can one survive the virus infestation in any areas, especially the affected areas? Therefore, the specific objective of this research, which must be attained at the end, is to analytically estimate the survival chances of the infected persons within specified time frame. The geographical location of this research is global (worldwide). That is, it is not limited to China where the disease emerged, although concise attention will be paid to cases in China. The survival analysis can be described as the analysis of survival data. The key feature of survival data is censoring. Without loss of generality, this research shall employ the concept of data censoring. Data are censored if the exact values of cases observed are fickling or not known but the information concerning each observed value as it relates to one another is ascertained. Censoring can be right, left, interval, random, informative, non-informative or truncated. Right censoring mechanism is adopted when the ongoing observation is cut short. That is, some observed cases being investigated are still alive when the analysis ended. When data are left censored, it means we cannot ascertain when entry into a state (such as infected or dead) took place during the observation/investigation. Both left and right censoring are considered special cases of interval censoring. Non-informative censoring gives no or low information about the actual lifetimes of observed cases. Censoring can also be of type 1 or type 2. Type 1 censoring is when the censoring time is known in advance but in Type 2, the observation continues until the expected or targeted number of deaths has occurred. This study will provide a basis for health policy making that ultimately leads to a decrease in unwarranted deaths and overall development in the quality of life.

1. **BRIEF LITERATURE REVIEW.**

In the word at the moment, according to the data compiled by Dadax (worldometer) and made available by the Chinese Health Commission of Hubei Province and the National Health Commission of the Peoples’ Republic of China, only four (4) continents had confirmed cases of the Novel Coronavirus disease. In all these continents, only twenty-eight (28) countries had recorded infected cases. 1110 people, 1 person and 1 person were reported dead in China, Philippines and Hongkon respectively in all these twenty-eight countries. Out of the twenty (20) cases in the North America, 65% of the cases were recorded in the U.S with the highest reported cases in California and followed by Illinois (CDC, 2020). Table 1 depicts the countries/continent currently infected.

**Table 1: Infected Countries/Continents**

|  |  |  |  |
| --- | --- | --- | --- |
| Continents | Cases | Deaths | Countries |
| Asia | 44709 | 1112 | China, Japan, Hongkong, Singapore, Taiwan, South Korea, Malaysia, Thailand, Vietnam, Macao, U.A.E, Philippine, India, Sri Lanka, Cambodia and Nepal |
| Europe | 45 | 0 | Germany, France, U.K, Italy, Sweden, Russia, Spain, Belgium and Finland |
| North America | 20 | 0 | U.S and Canada |
| Australia | 15 | 0 | Oceania |

**Source:**Dadax ( February, 2020)

The table shows 44789 confirmed cases and 1112 deaths were recorded in the world with 99.82% (infected) and 100% (death) cases in Asian countries/communities. The total confirmed case figure represents the cumulative count which comprises the number infected, dead and recovered/discharged. Out of the 17000 recorded cases in china, 510 cases were critical, 15% cases were severe while about 13940 cases were mild cases (Worldometer, 2020).39148 people were considered infected out of the cases recorded worldwide. The remaining 5641 cases were cases with outcome (that is, death or recovered). 80% of these people had recovered while the remaining 1112 people had been reported dead. Table 2 presents the worldwide growth rate of daily cases recorded from January 23 to February 10, 2020. The daily growth rate of the confirmed cases was computed using the formula:

1+

When the growth rate exceeds one, it shows an increase in the number of new cases compared to the initially confirmed cases. Consequently, values between ***0.1*** and ***1*** depict decline in the daily cases recorded. Also in the same manner, Table 3 indicates the world daily death growth factor. When a growth factor value goes below ***1*** or the line trend downward, it indicates a positive sign that lower deaths are being recorded in comparison with the previous day record of deaths. A value of ***1*** means constant number of deaths. That is, neither increase nor decrease in the number of infected people dying. The incubation period for this virus is 2(minimum) to 14(maximum)

days. Incubation period is the time it takes for the virus signs and symptoms to begin to show. This may vary depending on the patients’ medical history. In some patients, the incubation period can extend to twenty four (24) days, although this is an outlier. The explanation of the incubation period among groups of people depends upon the strength of the association between the external and internal risk/health factors. The incubation period can change as more data/situations become readily available due to further medical observation/investigation. In a BBC report, the transmission rate from infected person to uninfected person is ***1:4***. This means that an average of four (4) new cases can be generated from just one infected person. According to W.H.O (2020b), the virus can infect any persons irrespective of their age but the older people and those with existing health issues/challenges (such as diabetes, heart disease, cardiovascular disease, e.t.c) are more vulnerable to contact the virus than any other group of individuals. Although the virus hits irrespective of gender, but more males have got more confirmed cases than females. A reasonable explanation for this is that most of the pre-existing health conditions are commonly associated with older males. In a statement released by the European Centre for Disease Prevention and Control (ECDC), Novel Coronavirus can be zoonotically transmitted. This means an animal disease which can be transmitted to humans (that is, spillover). Methods of transmission can be through droplet, contamination, sneezing or cough. It was observed that those who have high chances of contacting the virus are those who are regularly in contact with animals, either as live animal market workers or those involved in treating infected people or animals. The symptoms range from mild, severe to critical. Few symptoms are nausea, fever, dry cough, shortness of breath, pneumonia, fatigue, kidney failure and other breathing difficulties. Polymerase chain reaction (PCR) is one way of testing if someone is infected by the virus. It identifies the virus based on genetic fingerprint.

1. **METHODOLOGY.**
   1. **METHODS OF DATA COLLECTION.**

In the course of this study, different methods and strategies of investigations were conceived to gather data/materials related to the study in order to have comprehensive knowledge on the subject matter. Visits were made to hospitals, experts, friends and family members in the relevant organizations for personal interviews and/or corporate observations based on the flexibility of approach. The major secondary data were got from the internet, periodic reports, forums, newspapers, paper presentations, journals and other publications/surveys of health or health related organizations. In recent days, there have been inconsistencies and differences in the use and interpretation of available data on the number of cases and deaths but there have also been different attempts to maximize the consistencies and minimize the differences through the help of national and international organisations such as World Health Organisation (W.H.O.), United Nations, Centres for Disease Prevention and Control, the National Health Commission of the Peoples’ Republic of China, Chinese Health Commission Of Hubei Province and others. Most data are generally drawn from surveys. As a result of this, it must be noted that the reliability of the estimates from the surveys depends upon the overall quality of the sampling frames and methods used. Generally, the larger the sample frames, the better the quality and validity of the result. At present, most developing/underdeveloped countries lack adequate health information systems they need to accurately and effectively monitor health trends and make salient health policies/decisions. That is why W.H.O., U.N. and others are creating an entirely new process for strategic policy dialogue with these countries, using evidence and information strategically to drive change. This paper shall carry out survival analysis on 42,653 cases reported/observed within twenty (20) days. It must be noted that this figure represents those cases that were reported at the time the data was retrieved. This figure may be different later due to some unreported cases which might be added to the earlier reported figure.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Year**  **Table 4 : Reported Cases/Deaths of Novel Coronavirus in the World** | **Date(**tm**)** | **Day(**m) | **Reported cases** | **Reported deaths** |
| 2020 | 22-Jan | 0 | 133 | 0 |
| 23-Jan | 1 | 266 | 8 |
| 24-Jan | 2 | 472 | 16 |
| 25-Jan | 3 | 698 | 15 |
| 26-Jan | 4 | 785 | 24 |
| 27-Jan | 5 | 1781 | 26 |
| 28-Jan | 6 | 1477 | 26 |
| 29-Jan | 7 | 1755 | 38 |
| 30-Jan | 8 | 2008 | 43 |
| 31-Jan | 9 | 2127 | 46 |
| 1-Feb | 10 | 2604 | 45 |
| 2-Feb | 11 | 2837 | 58 |
| 3-Feb | 12 | 3239 | 64 |
| 4-Feb | 13 | 3925 | 66 |
| 5-Feb | 14 | 3723 | 73 |
| 6-Feb | 15 | 3163 | 73 |
| 7-Feb | 16 | 3437 | 86 |
| 8-Feb | 17 | 2676 | 89 |
| 9-Feb | 18 | 3001 | 97 |
| 10-Feb | 19 | 2546 | 108 |
|  |  | **total** | **42653** | **1001** |
| **source:** |  |  | [www.wordometers.info/coronavirus](http://www.wordometers.info/coronavirus) |  |
|  |  |  | accessed on Feb 10,2020 . 23:58GMT |  |
|  |  |  |  |  |

* 1. **STATISTICAL TOOLS.**

In statistical inference, censoring mechanisms play vital role. Interval censoring is when the actuarial investigation of event of interest falls within certain observational interval of time. In this research, we assume non-informative, Type 1 and interval censoring. The interval censoring is assumed because the observational design allowed us to observe event of interest within certain interval period of time. In this paper, observation/investigation of cases was carried out from January 22 to February 10, 2020.The two statistical tools to be used in this research are descriptive and inferential. According to Mojekwu (2012), the descriptive statistics is concerned largely with summary calculations and graphical displays of results/data to derive reasonable decisions. However, the displays and summary calculations tend to describe the general characteristics of the data without delving deep into those characteristics. The inferential statistics deals with how inferences are made from numerical data. It involves analysis and interpretation of the characteristics of the given data in order to arrive at realistic decisions. The Nelson Aalen survival model will be used to analyse the survival estimate of data displayed in Table 4 using Microsoft Excel Package for the computation. This estimate exists between values ***0*** and ***1***. Value ***0*** means no chance of survival at all while value ***1*** means a certain (100%) chance of survival.

**Nelson Aalen survival estimate formula/notations**

Nelson Aalen survival estimate (SNA(t))

Estimate of the cumulative hazard function

*m* day of observation

*tm* date/time death observed

*dm* number of deaths at tm

*nm* number of people available to die at tm

= hazard rate or probability that a life fails at tm

1. **ANALYSIS AND INTERPRETATION OF DATA.**

**4.1 ANALYSIS**

In Nelson Aalen survival model, Value of ***0.5*** indicates equal chance of survival and death. This survival model was specifically chosen to analyse the data because it is non-parametric approach whose estimator assumption is based on non-informative censoring. The closer the value is to ***1*** from ***0.5*** upward, the better the probability of survival and vice versa. Table 5 is the result of Nelson Aalen survival analysis of the data displayed In Table 4.

**Table 5: Computation of Nelson Aalen Survival Analysis**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **m** | **dm** | **nm** | µ**m** | |  | | --- | |  | |  |  |
|  |  |  |  |  |  |  |
| 0 | 0 | 42653 | 0 | 0 | 1 |  |
| 1 | 8 | 42645 | 0.000187595263219604 | 0.00018760 | 0.999812 |  |
| 2 | 16 | 42629 | 0.000375331347204954 | 0.00056293 | 0.999437 |  |
| 3 | 15 | 42614 | 0.000351996996292298 | 0.00091492 | 0.999085 |  |
| 4 | 24 | 42590 | 0.000563512561634186 | 0.00147844 | 0.998523 |  |
| 5 | 26 | 42564 | 0.000610844845409266 | 0.00208928 | 0.997913 |  |
| 6 | 26 | 42538 | 0.000611218204899149 | 0.00270050 | 0.997303 |  |
| 7 | 38 | 42500 | 0.000894117647058824 | 0.00359462 | 0.996412 |  |
| 8 | 43 | 42457 | 0.00101278941046235 | 0.00460741 | 0.995403 |  |
| 9 | 46 | 42411 | 0.0010846242720049 | 0.00569203 | 0.994324 |  |
| 10 | 45 | 42366 | 0.00106217249681348 | 0.00675420 | 0.993269 |  |
| 11 | 58 | 42308 | 0.00137089912073367 | 0.00812510 | 0.991908 |  |
| 12 | 64 | 42244 | 0.00151500804848026 | 0.00964011 | 0.990406 |  |
| 13 | 66 | 42178 | 0.00156479681350467 | 0.01120491 | 0.988858 |  |
| 14 | 73 | 42105 | 0.00173376083600523 | 0.01293867 | 0.987145 |  |
| 15 | 73 | 42032 | 0.00173677198325086 | 0.01467544 | 0.985432 |  |
| 16 | 86 | 41946 | 0.00205025508987746 | 0.01672569 | 0.983413 |  |
| 17 | 89 | 41857 | 0.00212628712043386 | 0.01885198 | 0.981325 |  |
| 18 | 97 | 41760 | 0.0023227969348659 | 0.02117478 | 0.979048 |  |
| 19 | 108 | 41652 | 0.00259291270527226 | 0.02376769 | 0.976513 |  |

**4.2 INTERPRETATION**

The Nelson Aalen estimate of S(t) is:

**SNA(t)**

The survival estimate and the range of time it applies have been clearly stated for better understanding. In the same vein, the value of ***m*** which applies has also been shown. The more the number of ***tm*** the more the estimate tends to zero. That is, if there are no lives remaining to be censored among the observed lives when the investigation ended, the last value of the estimate will be zero which signifies sure death at that time. From the analysis carried out in Table 5, at the initial day of the investigation, the estimate shows the value of one (*1*) which means that no live was lost due to the virus on the first day of investigation. This does not mean that Novel Coronavirus did not kill the infected people on that day. It simply means that no death was recorded among the cases of infected people being observed or investigated on that day. From the analysis, it is evident that as time passes by, the estimate figure begins to reduce, meaning that more deaths are being recorded. This indicates that there are chances of surviving the disease at the earliest period of the infection, provided the infected persons take adequate medical aid. Invariably, as shown in the table, the more the virus stays in the body, the lower the probability of survival or the chance of recovery from the disease. It will be noted that the survival estimate follows a particular pattern. The values of the estimate decrease as the values of ***m*** increase. This is because the number of cases and deaths used in the analysis also follow same pattern. Within the study period, both the number of cases and the number of deaths recorded were increasing because these days happened to be when the virus hit people hard. In practical application, using the result obtained in this research to forecast any future outbreak of this virus, an average number of eighty-eight (*88*) infected persons are expected to die in less than two days out of a million infected people since the survival chance of that period is ***0.999812***(from the computed estimate). This forecast will be accurate or valid if responses that are aimed at mitigating exposure to the risk factors and adequate access to health services are not effectively and efficiently employed. Life will be short when its quality is poor.

1. **SUMMARY,CONCLUSION AND RECOMMENDATION**

Human coronaviruses were first detected in the 1960s. Novel coronavirus *2019* is one of the coronaviruses. Other cornaviruses witnessed in the past are Severe Acute Respiratory Syndrome (SARS-Cov) and Middle East Respiratory Syndrome (MERS-Cov). Novel Coronavirus started in Wuhan, one of the largest Chinese cities with a population of over eleven (*11*) millions. This virus has never been identified in humans prior to this time. It has greatly created inflation in the affected areas due to shortage in the supplies of goods and services following the lockdown measure adopted by the Chinese government to curb the spread. In other words, products/services to market are difficult to get to where they are needed due to disruption in the supply and demand chains. On the basis of transmission rate of four (4) newly infected persons arriving from one case, within a few days, the number of reported cases of the Novel Coronavirus was on the high side. The number was in tens of thousand. This shows the virus is easily transmitted from one person to the other or from animal to person. The number may run to a million if adequate effort or measures are not timely intensified. The aiding factor to its spread is the fact that some infected people are asymptomatic. It has been established in this research that some people might have this virus without knowing because the infection symptoms are not showing. These people infect others without knowing. From the analysis carried out, people who have just contacted the disease have higher probability of survival if medical aid is given without stringent or tedious conditions attached. This research recommends that those who are not sure on whether they have contacted the disease due to absence of sickness symptoms should see health provider immediately for tests. Government of every nation should establish a centre for quick test for such confirmation at no cost. Making it free and accessible will encourage any unhealthy individuals to easily go for confirmation test. This will help to curb the spread of the disease. Since this virus broke out from Wuhan, this research recommends WUHAN prevention concept to combat the virus from spreading or infecting people.

**WUHAN Prevention Concept**

* **W**ash your hands/body regularly
* **U**se nose cover/mask
* **H**ave your hotness/coldness of your body checked
* **A**void unnecessary crowd
* **N**ever touch sensitive parts of your body with unclean hands/materials/equipment

The virus from the infected person should be isolated and shared with recommended research laboratories in order to develop accurate diagnostic vaccines/drugs. Vaccines are much easier to work on when one gets the sample of the virus. By this, tests to detect particular immune cells and antibodies will be conducted and achieved. Also, sharing the virus sample will enable a group of scientists to pool wisdom/resources together. Medical supplies to fight the virus epidemic should readily be available in not just the affected areas but also in some neighbouring areas especially those areas whose governments’ health plans are nothing to write home about. The international health organizations should liaise with global experts, government of each nation and partners to expand the scientific knowledge on the virus, advise on the protective measures, track and to prevent the spread of the disease outbreak. Coronavirus is an imminent threat to public health and a series of measures are seriously needed to prevent it from spreading. Special hospital should be established to attend to cases of coronavirus to combat its spread. Vigorous health programs to encourage lifestyles preventing this disease and more budgetary spending in such area should be the focus of public health actions in order to provide effective treatments. Research that helps to understand the specific organisms responsible for infection or deaths from this disease, as well as research on practices related to seeking healthcare, should be embarked on and implemented by the government. . In order to reduce preventable deaths, high quality data system should be taken seriously and improved upon. With right information and health care, diseases can be prevented, cured or treatable. Quality data are crucial for improving health decisions. Finally, there is still considerable room for further research on Novel Coronavirus. The data used for this study have been censored. Meaning that some lives observed were still alive at the time this study ended. This might slightly affect the result obtained. It will be fascinating to have a complete data by following up the observed lives until the outcome of the last live observed. Therefore, further work on this area will be certainly warranted. Nevertheless, this study has paved way for future research and provided basic knowledge on the survival of Novel Coronavirus.

**REFERENCES**

BBC (2020): *The Global Fight Against Coronavirus.* Accessed on Jan 28, 2020 from <www.bbc.com/future/article>.

CDC (2020): 2019 Novel Coronavirus (2019-nCov)*: Situation Summary in the U.S.* Retrieved from the U.S Centre for Disease Control and Prevention on February 10, 2020.

Dadax (2020): Countries Where the Coronavirus Has Spread. Accessed on February 11, 2020 23: 10 GMT from [www.worldometer/info](http://www.worldometer/info).

ECDC (2020): *Overview of Novel Coronavirus.* European Centre for Disease Prevention and Control. [www.ecdc.europa.eu/en/novel-coronavirus](http://www.ecdc.europa.eu/en/novel-coronavirus)

Mojekwu, J.N.(2012): *Business Statistics with Solved Examples* (3rd Ed. Chapter 1, Pg 1). Easy Prints Publication, Lagos, Nigeria.

Nature Research (2020): Coronavirus Latest: *W.H.O Officially Names Disease COVID-19*. Retrived from [www.nature.com/articles/coronavirus-latest](http://www.nature.com/articles/coronavirus-latest) on February 7, 2020.

New York Times (2020): Coronavirus Updates: *A Grim Landmark as Official Death Toll Tops in China.* [www.nytimes.com/2020/02/10/world/asia/coronavirus](http://www.nytimes.com/2020/02/10/world/asia/coronavirus)

W.H.O (2020a): Health Topic: *Novel Coronavirus 2019*. Accessed on January 31, 2020 from [www.who.int/health-topics/coronavirus](http://www.who.int/health-topics/coronavirus)

W.H.O (2020b): *Online Training as a Weapon to Fight the New Coronavirus.*Obtained on Feb 8,2020 from <https://openwho.org/courses/intoduction-to-nCov>

Worldometer (2020): Coronavirus: *Cases, Deaths and Mortality Rate.* Accessed on February 10, 2020 23:58 GMT from [www.worldometer/info/coronavirus](http://www.worldometer/info/coronavirus).