Addis Ababa University
College of Natural and Computational Sciences
Department of Statistics

Determinants of Severity of Illness in Patients with COVID-19: The case of Eka Kottebe General Hospital, Addis Ababa, Ethiopia

By: Mamo Worku

Advisor: Dejen Tesfaw (PhD)

June, 2021

Addis Ababa, Ethiopia

Addis Ababa University
College of Natural and Computational Sciences
Department of Statistics

Determinants of Severity of Illness in Patients with COVID-19: The case of Eka Kottebe General Hospital, Addis Ababa, Ethiopia

Mamo Worku

A Thesis Submitted to the School of Graduate Studies of Addis Ababa University Department of Statistics in Partial Fulfillment of the Requirement for Degree of Masters of Science in Statistics (Biostatistics)

Advisor: Dejen Tesfaw (PhD)

June, 2021
Addis Ababa, Ethiopia

Addis Ababa University

School of Graduate Studies

This is to certify that the thesis prepared by **Mamo Worku**, entitled: **Determinants of Severity of Illness in Patients with COVID-19: The case of Eka Kottebe General Hospital, Addis Ababa, Ethiopia** and submitted in partial fulfillment of the requirements for the Degree Master of Science in Statistics (Biostatistics) compiles with the regulations of the university and meets the accepted standards with respect to originality and quality.

Signed by the examining committee:

Examiner_____signature_____Date_____

Examiner_____signature_____Date_____

Advisor_____signature_____Date_____

_____

Chair of Graduate Program Coordinator

Declaration

I, the undersigned, declare that the thesis is my original work, has not been presented for degrees in any other university and all sources of materials used for the thesis have been duly acknowledged.

Name: Mamo Worku

Signature: _____

Place of submission: Department of Statistics, College of Natural and Computational Sciences, Addis Ababa University.
Date of submission: June, 2021

This thesis has been submitted for examination with my approval as a university advisor.

Dejen Tesfaw (PhD) _____
Advisor's Name **Signature**

Acknowledgment

First of all I would like to thank the Almighty God for his love, grace and protection throughout the course and preparing this thesis of my graduate study at department of statistics, Addis Ababa University.

Secondly, Special thanks to my thesis advisor Dejen Tesfaw (PhD) for spending his valuable, constructive comments and direct me throughout my study. He always gave me a clear direction to follow and taught me with his generous patience.

Thirdly, my deepest thanks go to all staff members of Statistics Department at Addis Ababa University especially, my instructors Ass. Prof. Mekonnen Tadesse, Prof. Eshetu Wencheko, Dr. Bedilu Alamirie, Dr. Dejen Tesfaw, Dr. Merga Belina, Prof. M. K. Sharma and all individuals in the department of statistics for their moral support and positive attitude when I needed them during my course work and research period.

In addition, I would like to thank the Ethical Review Committee Board of the Eka Koetebe General Hospital for giving me the necessary support to get data for my thesis work. Especial tanks for Mr. Tadiyos, Dr. Ferehiwot (MD, MPH) and Ms. Kongit for their positive support.

Finally I am grateful to tanks all my families and friends who had the source of my strength to accomplish the study.

Abstract

Determinants of Severity of Illness in Patients with COVID-19: The case of Eka Kottebe General Hospital, Addis Ababa, Ethiopia

Mamo Worku

Department of Statistics, Addis Ababa University, 2021

Corona virus is a zoonotic disease caused by severe acute respiratory syndrome corona virus 2. It believed that the virus jumped the species barrier to humans from another intermediate animal host. An outbreak of COVID-19 emerged in Wuhan, Hubei Province; China on 31 December 2019. The aim of this study was to identify determinant factors of illness severity in patients with COVID-19 in Eka Kotebe general hospital COVID-19 treatment center Addis Ababa, Ethiopia. To analysis the data considered in this study descriptive statistics, chi-square test and stereotype ordinal logistic regression model used for data analysis. The descriptive statistics is used to summarize the data using numbers and percent and the chi-square test is used to test the association between the dependent and independent variables. The stereotype ordinal logistic regression model is also used to identify the determinant of illness severity in patients with COVID-19. Out of the total patients considered in the study (N=454), about 240 (52.86%), 60 (13.22%), 106 (23.35%) and 48 (10.57%) patients were at mild, moderate, sever and critical stage, respectively. Moreover, using chi-square test the independent variables like sex, presence of symptom, fever, headache, chest pain, shortness of breath, general weakness, cough, co-morbid, diabetes mullets and hypertension had a significant association with severity of illness at 5% level of significant. The proportional odds model were considered as a candidate model to fit the data but as the assumption of parallelism was violated, the model was not adequate to fit the data. Thus, stereotype ordinal logistic regression model were fitted as the constraints ensure the outcome variable is ordinal. The stereotype ordinal logistic regression result shows that sex, age, presence of symptom, shortness of breath, general weakness and having co-morbidity had statistically significant effect on the severity of illness in COVID-19 patients.

Keywords*: COVID-19,* Stereotype, Severity of illness

List of Acronyms

COPD	Chronic obstructive pulmonary disease
CVD	Cardiovascular disease
Df	Degree of freedom
EPHI	Ethiopian public health institute
FMOH	Federal ministry of health
IC	Information criterion
ICU	Intensive care unit
MLE	Maximum likelihood estimation
OR	Odds ratio
PO	Proportional odds
SL	Stereotype logistic
SLRM	Stereotype ordinal logistic regression model
UN	United Nation
UNICEF	United nation international children emergency fund
WHO	World health organization

Contents

List of Tables

CHAPTER ONE

INTRODUCTION

1.1 Background

Corona virus is a zoonotic (a virus that is regularly transmitted from animal to human) disease caused by severe acute respiratory syndrome corona virus 2 (SARS-CoV-2). It is believed that the virus jumped the species barrier to humans from another intermediate animal host. This intermediate animal host could be a domestic food animal, a wild animal, or a domesticated wild animal which has not yet been identified. It usually causes a respiratory infection ranging from the common cold to more severe diseases such as middle-east respiratory syndrome and severe acute respiratory syndrome (Murphy and Bell, 2020).

An outbreak of COVID-19 caused by the 2019 novel corona virus began in Wuhan, Hubei Province; China on 31 December 2019. Although it is still unknown exactly where the outbreak started, several early infected people had visited hunan seafood wholesale market, in Wuhan, Hubei, China and the first recorded case of COVID-19 outside of China was officials confirmed in Thailand 13 January 2020 (Sun et al., 2020).

WHO originally called this infectious disease Novel Corona virus-Infected Pneumonia (NCIP) and the virus had named 2019 novel corona virus (2019-nCoV). On 11th Feb 2020, WHO officially renamed the clinical condition COVID-19, the acronym derived from "corona virus disease 2019" (Murphy and Bell, 2020). The name chosen to avoid stigmatize the virus's origins in terms of populations, geography, or animal associations. The virus declared as a public health emergency of international concern in January 2020 and as a pandemic on 11 March 2020 (Sun et al., 2020).

Corona virus can affect different people in different ways. Most infected people will develop mild to moderate illness and recover without hospitalization. The virus is rapidly spread from one person to another through respiratory droplets produced during coughing and sneezing. It is most transmittable when people are symptomatic, although transmission may possible before symptoms show in patients. The Most common symptoms are fever, dry cough, tiredness and the

less common symptoms are aches and pains, sore throat, diarrhea, conjunctivitis, headache, loss of taste or smell, a rash on skin, or discoloration of fingers or toes (Murphy and Bell, 2020).

Currently two vaccines are authorized and recommended to prevent COVID-19: Pfizer-BioNTeche and Moderna. As of 28 December 2020, large-scale phase III clinical trials are in progress or being planned for three COVID-19 vaccines in the United States: AstraZeneca, Janssen and Novavax (Mishra, Mukesh, 2020). The best way to protect our self from the spread of COVID-19 is clean our hands often, use soap and water or an alcohol-based hand rub, keep up a safe distance from anyone who is coughing or sneezing, wear a mask, don't touch our eyes, nose or mouth, cover our nose and mouth with our bent elbow or a tissue when you cough or sneeze, stay home.

The COVID-19 pandemic was rapidly speared around the world at the end of 31 December 2020 there was 83,800,241 confirmed cases and 1,825,018 deaths reported (Corona virus updates, 2021). The pandemic created several economic impacts in different sectors, it negatively affect global trade, interest rates, financial market liquidity and creating demand and supply shocks (Roy, 2020). The pandemic also has severely affected the education sector; many countries have decided to close schools, colleges and universities. The closure of schools, has affected more than 1.576 billion students worldwide (Ngware, 2020).

The COVID-19 pandemic spread to Africa also, Africa's first confirmed case announced in Egypt on 14 February 2020. Within three months, the virus had spread throughout the continent, Lesotho is the last African sovereign state to have remained free of the virus, reported a case on 13 May (Maclean, 2020). At the end of December 2020, 2,773,166 and 65,566 cases and deaths had reported in Africa, respectively (Corona virus updates, 2021).

The federal ministry of health of Ethiopia has confirmed the first COVID-19 case on 13 March 2020 in Addis Ababa, Ethiopia (FMOH, 2020). The case spread rapidly and causes deaths in Ethiopia until 31 December 2020 there were 124,264 confirmed cases and 1,923 deaths (Corona virus updates, 2021). In order to make all preventive and treatment activities uniform in Ethiopia, the need for national COVID-19 prevention and treatment guideline prepared by FMOH and EPHI. A committee organized from consultants of different specialties and given the task of developing evidence based, cost-effective and applicable national guideline for prevention and

treatment of COVID-19 in Ethiopia. The guideline has taken in to consideration the culture, leaving condition and background of the people and as much as possible understandable and usable by most levels of health care professionals (FMOH, 2020).

The pandemic was affected the health system, economy activity, education system and political activity in Ethiopia. The impact of COVID-19 on economic activity is mainly on workers employed in micro, small and medium-size enterprises, informal sector (manufacturing, construction, trading, retail, hospitality) and tourism sector workers (ONE UN Assessment, 2020). In education sector schools, colleges and universities were closed and about 26,088,744 students were out of school (UNICEF Ethiopia, 2020). The teaching and assessments are moving online for higher education students. Whereas students in the kindergarten, primary school and secondary and preparatory school was pass to the next level by giving 45 day tutorial for the lost part excluding grade 8 and 12 in academic year 2019/2020. The political activity of the country was also affected: many conferences cancelled and/or minimized participants, extended and the national election planned also extended to the next year due to the pandemic of COVID-19 (ONE UN Assessment, 2020).

1.2 Statement of the Problem

Corona virus is a highly transmittable and pathogenic viral infection caused by SARS-CoV-2, which emerged in Wuhan, China and spread around the world. As of 31 December 2020 there were 83,800,241; 2,773,166 and 124,264 cases reported globally, Africa and in Ethiopia, respectively, and about (1,825,018); (65,566) and (1,923) deaths reported globally, Africa and in Ethiopia, respectively (Corona virus updates January 01, 2021).

Studies by Adrish et al., 2021; Rieg et al., 2020; Wang, et al., 2021 and mantovani et al., 2020 revealed that more severely ill COVID-19 patients were at risk of death. Other studies by Lapidus et al., 2020 and Rees et al., 2020 revealed that more severely ill patients were take long time in hospital to recover. Also Ross et al., 2020 shows that patient at mild stage have possibly recovered at home and patients at critical stage need oxygen to breathe.

In this study, we propose to identify factors that are associated with severity of illness by using data from Eka Kotebe general hospital COVID-19 treatment center.

1.3 Objective of the Study

General Objective

The main objective of this study was to identify the determinant factors of illness severity in patients in Eka Kotebe general hospital COVID-19 treatment center.

Specific Objective

Specific objectives of this study are

- To test the association between severity of illness and independent variables.
- To investigate the factors that lead to more severe cases of COVID-19.

1.4 Significance of the study

The findings from this research could be useful in many ways. The findings were believed to be useful for policy making, monitoring and evaluation activities of the government and different concerned agencies. Since the study is attempted to identify the determinant factors of illness severity in patients with COVID-19, the end-user governmental and non-governmental organizations could take intervention measures and set proper police and recommendation. Finally, the study may also serve as stepping stone for those who have interest to make further studies on the same area of research.

1.5 Limitation of the Study

In this study, socio-demographic variables like marital status, place of residence, religion, employment status and education level of patients were not recorded at the time of data collection.

CHAPTER TWO
LITERATURE REVIEW

Chang et al. (2020) study on the risk factors for disease progression on 211 adult who were asymptomatic or with mild presentations of COVID-19 patients from Korea worker's compensation & welfare service Daegu hospital (Daegu, South Korea) from Feb 28, 2020 to March 31, 2020. A multivariable logistic analysis showed that body temperature, chills, initial chest X-ray findings, and presence of diabetes were significantly associated with predicting the progression to severe stage of COVID-19.

Xu et al. (2020) identifies the determinants of illness severity of COVID-19 using ordinal logistic regression model. Retrospective cohort of COVID-19 patients on four hospitals, 598 patients included from 1 January to 8 March 2020. There finding conclude that patients in the age group of 70+ years, 40–69 years, myohaemoglobin>48.8 ng/ml, hypertension, ALT >50 μ/l, cTnI >0.04 ng/ml had greater risk of developing worse severity of illness.

Zhang, et al., (2020) conducted a study to identify the clinical characteristics of different subtypes and risk factors for severity of illness for early identification and prompt treatment in Zhejiang, China. Retrospective study conducted on 688 confirmed COVID-19 patients from 17 January to 12 February 2020. Univariable and multivariable ordinal logistic regression model is used and found that older age, male, cough, fever, hemoptysis, gastrointestinal symptoms and hypertension are the risk factors for severity of illness in patients with COVID-19.

Dong et al. (2020) conducted the study on identifying the epidemiological characteristics and transmission patterns on 2143 patients with COVID-19 reported to the Chinese center for disease control and prevention from January 16 to February 8, 2020 by using a log-normal distribution to fit data both onset and diagnosis dates. The researcher concluded that the clinical manifestations of children's COVID-19 cases were generally less severe than those of adults' patients.

Turcotte et al. (2020) studded on the risk factors for severe illness in hospitalized COVID-19 patients at a regional hospital, a retrospective chart review of 117 patients hospitalized for COVID-19 from March 1 to April 12, 2020 conducted. Multivariate regression analysis

identified sputum production, diabetes mellitus and chronic kidney disease as significant risk factors for critical illness.

Geng et al. (2021) investigated risk factors for developing severe COVID-19 in China by grouping mild and moderate cases together as non-severe cases and categorized severe and critical cases together as severe cases. The data for this study was disease surveillance data on symptomatic cases of COVID-19 reported from 30 provinces in China between January 19 and March 9, 2020. The risk factors for severe COVID-19 were being male, fever, cough, having fatigue, chronic kidney disease, hypertension and diabetes.

Gálvez-Barrón et al. (2021) conducted study on the prognostic factors of severe diseases patients with COVID-19. A cohort study on 464 patients hospitalized for COVID-19 in the hospitals of the Consorci Sanitari de l'Alt Penedes i Garraf. Multivariable regression analysis used to identify the risk factors, female sex, serum lactate dehydrogenase and no chronic disease had associated with severe disease.

Ann (2021) investigated the severity of COVID-19 and survival in patients with rheumatic and inflammatory diseases by using multivariable binary logistic regression models to indentify the most significant clinically relevant factors obtaining data from the French RMD COVID-19 cohort of 694 patients. The researcher identified that older age, male gender, obesity, and hypertension were found to be associated with severe COVID-19.

The study using meta-analysis conducted to identify factors associated with disease severity and mortality among patients with COVID-19 by reviewing 109 articles until May 8, 2020 total of 20296 participants was assessed (Chidambaram et al., 2020). The results showed that congestive heart failure, hilar lymphadenopathy, bilateral lung involvement and reticular pattern had higher odds of severe disease.

Degarege et al. (2020) conducted a systematic review of 71 studies that involved 216,843 patients on risk factors for severe illness and death in COVID-19 using meta-analysis. Severe illness was significantly associated with any co-morbidities, hypertension, cardiovascular diseases, diabetes, male sex and age at least 60 years.

Zheng et al. (2020) assesses the risk factors of critical and mortal COVID-19 cases on 3027 patient's data until March 20, 2020 using meta-analysis. Their findings show that male, older than 65, smoking, hypertension, diabetes, cardiovascular disease, respiratory disease, fever, shortness of breath or dyspnea were higher in critical patients compared to the non-critical patients.

The risk of developing dangerous symptoms of COVID-19 may increase in people who are older and in people of any age who have other serious health problems such as heart or lung conditions, weakened immune systems, severe obesity, diabetes, chronic kidney, liver disease, cystic fibrosis, pulmonary fibrosis, moderate to severe asthma, chronic obstructive pulmonary disease (COPD), cancer, certain blood disorders, chronic kidney disease and HIV/AIDS had at higher risk of developing more severe illness (Mayo Clinic, 2020).

Wu et al. (2020) conducted a systematic review of 55 studies that involved 1265 patients on clinical determinants of severity of corona virus disease 2019; meta-analysis has conducted to examine COVID-19 related death and risk factors for severity of COVID-19. The study investigated that male sex, older age, co-morbidities such as hypertension, diabetes, cardiovascular disease, respiratory disease and cerebrovascular disease could increase the risk of developing a severe case of COVID-19.

Meta-analysis on the clinical characteristics and risk factors of severe COVID-19 in China published before July 26, 2020 total of 30 papers and total of 1,457 severe COVID-19 patients included in this study. The analyses concluded that being male, elderly and obese patients and those with any co-morbidity, especially with hypertension, diabetes, and CVD were more likely to develop severe cases (Hu and Wang, 2021).

Leulseged et al. (2020) assessed the determinants of COVID-19 disease severity (mild, moderate and severe) among COVID-19 patients admitted to millennium COVID-19 care center in Ethiopia. A cross-sectional study was conducted from June to August 2020 among randomly selected 686 patients. Multivariable multinomial logistic regression models were used to analysis and conclude that being old, male sex, hypertension, diabetes mellitus, and having symptoms of fever and headache were determinants of developing a more severe COVID-19 disease category.

CHAPTER THREE
Data and Methodology

3.1 Study Area

The study was conducted at Eka Kotebe general hospital COVID-19 treatment center. The center was the first hospital designated to manage positive COVID-19 cases in Ethiopia. It has capacity of admitting 600 patients. The center was found in the capital city of Ethiopia, Addis Ababa, Yeka sub city

3.2 Source of Data

The data for this study was obtained from Eka Kotebe general hospital COVID-19 treatment center in Addis Ababa, Ethiopia for patients treated in the hospital from March 2020 to 31 December 2020 by reviewing patient's card.

3.3 Sample Size Determination

The dependent variable of interest is ordered scale in nature with the highest ordered categories indicates the critical stage whereas the lowest ordered categories is the mild illness. The required sample size formula for this study is used based on assumption of constant odds ratio Walters et al. (2001) and it is given by:

$$n = \frac{6(Z_{\frac{\alpha}{2}} + Z_{\beta})^2}{(\log OR)^2\{1 - \sum_{j=1}^{J} \bar{\pi}_j{}^3\}}$$

The total patient treated at the treatment center during the study period was 3081. The sample size required for this study was 532 patients (see detail sample size calculation in Appendix 1). Since the sample size n_0 exceeds 5% of the total patients, Cochran's (1977) correction formula was used to calculate the final sample size which is equal to 454 computed the by:

$$\mathrm{n} = \frac{\mathrm{n}_0}{1 + \frac{\mathrm{n}_0}{\mathrm{N}}}$$

Simple random sampling method was used to collect the recorded data at office level since the data is available at the same treatment center.

3.4 Study Variables

Dependent variable: The dependent variable for this study is severity of illness which is categorized as mild, moderate, sever and critical (WHO, 2020). The description of each classification for severity of illness in COVID-19 patients is given as follows.

1) **Mild illness:** Individuals with any of various signs and symptoms (fever, cough, sore throat, malaise, headache, muscle pain) without shortness of breath or abnormal imaging.
2) **Moderate illness:** Individuals with evidence of lower respiratory disease by clinical assessment or imaging and oxygen saturation (SaO2) over 93% on room air at sea level.
3) **Severe illness:** Individuals with a respiratory frequency of over 30 breaths per minute, SaO2 up to 93% on room air at sea level, ratio of arterial partial pressure of oxygen to fraction of inspired oxygen (PaO2/FiO2) below 300, or more than 50% lung infiltrates.
4) **Critical illness:** Individuals with respiratory failure, septic shock, and/or multiple organ dysfunctions.

The coding of severity illness is:

$$\text{Severity of Illness} = \begin{cases} 1 = \text{mild illness} \\ 2 = \text{moderate illness} \\ 3 = \text{sever illness} \\ 4 = \text{critical illness} \end{cases}$$

Independent variables: the description and coding for independent variables used for the study is given in the following table.

Variables	Categories
Age (continuous)	
Sex	0= Female 1= Male
Religion	0= Muslim 1= Christian
Presence of symptom	0= Asymptomatic 1= Symptomatic
Fever	0= no 1= yes
Headache	0= no 1= yes
Chest pain	0= no

Variables	Categories
Cough	0= no 1= yes
Diarrhea	0= no 1= yes
Co-morbidity	0= no 1= yes
Diabetes mellitus	0= no 1= yes
Asthma	0= no 1= yes
Hypertension	0= no 1= yes
Cardiovascular disease	0= no

	1= yes
Shortness of breath	0= no 1= yes
General weakness	0= no 1= yes

	1= yes
Chronic obstructive pulmonary disease	0= no 1= yes

3.5 Methods for Data Analysis

3.5.1 Introduction

In this chapter, both descriptive and inferential statistics is used. The descriptive statistics is used to describe the basic features of the data in the study. For inferential statistics the logistic regression model is used due to the assumptions of normality and constant variances of residuals are not satisfied and the dependent variable is categorical. It is one of the generalized linear model that can have three components link function, random component and systematic component (McCullagh and Nelder, 1989).

The effect of predictor variables are usually explained in terms of odds ratio and hence the name logistic regression, also called the log-odds function. This model applies maximum likelihood estimation after transforming the dependent into a logit variable. There are different types of logistic regression:- Binary logistic regression for dichotomy response variable, multinomial logistic regression for more than two category response variable and ordinal logistic regression for more than two ordered category of response variable.

3.5.2 Ordinal Logistic Regression Model

Ordinal logistic regression model is used to analyze for ordinal response variables, the response variable can be ordered by nature or by some ordered way. There are several ordinal logistic regression models: proportional odds model, partial proportional odds model, continuous ratio model and stereotype logistic model. The most frequently used ordinal logistic regression model in practice is proportional odds model (Peterson and Harrell, 1990).

Proportional Odds Model

Proportional odds model is used to model the ordinal dependent variable by defining the cumulative probabilities instead of considering the probability of an individual event. It

considers the probability of that event and all events that are ordered before it. Suppose the J categories of Y conditionally to the values of co-variables occur with probabilities $p_1, p_2, \ldots, p_J$ that is, $p_j = P(Y = j/\underline{X})$ for $j = 1, 2, \ldots J$. Cumulative probabilities are the probability that the response Y falls in category j or below, for each possible jth cumulative probability that is $p(Y \leq j) = p_1 + p_2 + \ldots + p_j$.

Let Y takes categorical response variable with J ordered categories and assume p(Y=1) is p_1, p(Y=2) is p_2, ..., p(Y=j) is p_j; for j = 1, ... , J. Cumulative probability reflect the ordering, with $p(Y \leq 1) \leq p(Y \leq 2) \leq \ldots \leq p(Y \leq J)$ and let the cumulative probability of the first $J-1$ of Y is $p(Y \leq j) = \pi_j$, $j = 1, \ldots, J-1$.

Then the odds of the first $J-1$ cumulative probabilities is

$$odds\big(\mathrm{p(Y \leq j)}\big) = \frac{\mathrm{p(Y \leq j)}}{\mathrm{p(Y>}j)} = \frac{\mathrm{p(Y \leq j)}}{1-\mathrm{p(Y \leq j)}} = \frac{\pi_\mathrm{j}}{1-\pi_\mathrm{j}}, \qquad j = 1, \ldots, J-1$$

The POM models the log odds of the first $j-1$ cumulative probabilities as:

$$logit[p(Y \leq j)] = \log\left[\frac{p(Y \leq j)}{1-p(Y \leq j)}\right] = \log\left[\frac{\pi_j}{1-\pi_j}\right]$$

And the relationship between the cumulative logit of Y is:

$$\log\left[\frac{\pi_j}{1-\pi_j}\right] = \log\left[\frac{\pi_j}{\pi_{j+1}+\cdots+\pi_J}\right] ; \; j = 1, \ldots, J-1$$

Consider a collection of p explanatory variables denoted by the vector $\underline{X}' = (X_1, X_2, \ldots, X_p)$. The relationship between the predictor and response variable is not a linear function in logistic regression; instead, the logistic regression function is used, which is the logit transformation of π.

$$\pi_j = \frac{\exp\{\alpha_j + \beta_1 X_1 + \ldots + \beta_\mathrm{p} X_\mathrm{p}\}}{1 + \exp\{\alpha_j + \beta_1 X_1 + \ldots + \beta_\mathrm{p} X_\mathrm{p}\}}$$

Then the logit or log-odds of having $p(Y \leq j) = \pi_j$ is modeled as a linear function of the explanatory variables as:

$$\log\left[\frac{p(Y \leq j)}{1-p(Y \leq j)}\right] = \log\left[\frac{\pi_j}{1-\pi_j}\right] = \alpha_j + \beta_1 X_1 + \ldots + \beta_\mathrm{p} X_\mathrm{p}$$

It is equivalent to

$$\log\left[\frac{\pi_j}{1-\pi_j}\right] = \alpha_j + \sum_{i=1}^{p}\beta_i X_i \quad , \qquad 0 \leq \pi_j \leq 1$$

Therefore, $logit[p(Y \leq j)] = \alpha_j + \sum_{i=1}^{p}\beta_i X_i \quad ; j = 1, \dots, J-1 \; and \; i = 1, \dots, p$

The model's intercept varies for each of the equations that satisfy the condition $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_{J-1}$ and their values do not depend on the values of the independent variable for a particular case. There are also *p* beta coefficients (β) whose elements correspond to the effects of the co-variables on the response variable.

Test of Parallel Lines

Proportional ordinal model assumes that the relationships between independent variables and the dependent variable are the same for all categories that is, the slope coefficients are the same across the categories of the response variable. That means ordinal logistic regression model, assume that all logit surfaces are parallel. Test of parallel lines helps to determine whether it is reasonable to assume that the values of the parameters are constant across categories of the response Brant test were used. The test of parallelism contains: -2LL for the reduced model, the model that assume the planes or surfaces are parallel across the category of the response variable and -2LL for the general model that assumes planes or surfaces are separate across the category. The chi-square statistic is the log-likelihood difference between the two models. If the lines or planes are parallel, the proportional odds model would be adequate. In other word if there is evidence to reject the null hypothesis that the logit surfaces are parallel, the slope coefficients are not the same for all categories of the response variable, thus PO model would not adequate.

If the proportional odds assumption is not met then we can use the other type of ordinal logistic regression like stereotype ordinal logistic regression model. If the assumption of stereotype ordinal logistic regression failed then we can also use multinomial logistic regression model.

Stereotype Ordinal Logistic Regression Model

Liu (2014) mentioned that because of the proportional odds assumption is often violated, instead of using partial proportional odds model, stereotype model is an alternative option for clinical

data analysis. Stereotype logistic regression model proposed by Anderson (1984), who states that medical diagnoses tend to be fixed and rigid, based on the classification of disease severity as mild, moderate, sever and critical. Later introduced by Greenland (1994), who showed that the progression of a disease through various stages. SLRM is an extension of PO model which estimates the ordinal response variable, does not assume the proportional odds assumption and allows the effect of each predictor to vary across the ordinal categories. Stereotype model aims to reduce the number of parameters by imposing constraints, without reducing the adequacy of the fit. Stereotype model compares each categories of the response variable with the reference category (either the first or the last category).

The stereotype ordinal logistic regression model can be written in the following form:

$$logit[\pi(j,J)] \;=\; \log\left\{\frac{\mathrm{p}(\mathrm{Y}=\mathrm{j}/\underline{X})}{\mathrm{p}(\mathrm{Y}=1/\underline{X})}\right\}$$

$$logit[\pi(j,J)] = \;\alpha_j - \phi_j(\beta_1 x_1 + \;\ldots + \;\beta_{\mathrm{p}} x_{\mathrm{p}}), \qquad j = 2,3,\ldots,J$$

where, $j = 1$ is the reference category, which is the first category (mild stage); Y is the ordinal response variable with categories from j to J; $\alpha_2,\ldots,\alpha_J$ are the intercepts; $\beta_1,\ldots,\beta_{\mathrm{p}}$ are the logit coefficients of the explanatory variables and $\phi_1,\phi_2,\ldots,\phi_J$ are the constraints which are used to ensure the outcome variable is ordinal if $1 = \phi_1 > \phi_2 > \cdots > \phi_J = 0$ is satisfied. ϕ_1= 1 in order to uniquely identify the parameters when using estimated scores. If any two pairs of the constraints are the same, these two categories are indistinguishable, thus can be collapsed into one. By exponentiating $(-\phi_j\beta_i)$, we get the odds of being in a category j versus the baseline category j *=1* for a unit change in a predictor x_{i}.

Odds Ratio

Suppose the response variable Y has J ordered categories ($Y_j\ with\ j\ =\ 1,2,\ldots,J$) and two categories (A and B) of explanatory need to be compared. For category j, the odds ratio is given by:

$$OR_j = \frac{\dfrac{p(Y \le Y_j)|x_A}{1-p(Y \le Y_j)|x_A}}{\dfrac{p(Y \le Y_j)|x_B}{1-p(Y \le Y_j)|x_B}} = \frac{\dfrac{p(Y \le Y_j)|x_A}{p(Y > Y_j)|x_A}}{\dfrac{p(Y \le Y_j)|x_B}{p(Y > Y_j)|x_B}} = \frac{odds(A)}{odds(B)}$$

According to the usual definition, OR is the ratio between two odds, but now odds has defined in terms of cumulative probabilities. Thus the cumulative odds ratio is proportional to the distance between X_A and X_B which call the cumulative logit is a proportional odds model.

Methods of Parameter Estimation

The coefficients of ordinal logistic model are estimated by using maximum likelihood method. The method of maximum likelihood produces values of the unknown parameters that best match the predicted and observed probability values. The parameters are estimated by maximizing the likelihood or by maximizing the logarithm of likelihood. The likelihood function is given by the equation:

$$L=\prod_{i=1}^{n}\left[\prod_{j=1}^{J}\pi_j\,(X_i)^{y_{ij}}\right]=\prod_{i=1}^{n}\left[\prod_{j=1}^{J}(P(Y\le j/X_i)-P\,(Y\le j-1/X_i))^{y_{ij}}\right]$$

$$=\prod_{i=1}^{n}\left[\prod_{j=1}^{J}\left\{\frac{\exp\,(\alpha_j+\beta'X_i)}{1+\exp\,(\alpha_j+\beta'X_i)}-\frac{\exp\,(\alpha_{j-1}+\beta'X_i)}{1+\exp\,(\alpha_{j-1}+\beta'X_i)}\right\}^{y_{ij}}\right]$$

$$L(\beta^*)=\prod_{i=1}^{n}\left[\pi_1(X_i)^{y_{i1}}\pi_2(X_i)^{y_{i2}}*\ldots*\pi_J(X_i)^{y_{iJ}}\right]$$

Where β^* denote both the slope coefficients and intercept coefficients.

The log–likelihood function is:

$$l(\beta^*)=\sum_{i=1}^{n}y_{i1}\ln\left[\pi_1(x_i)\right]+y_{i2}\ln\left[\pi_2(x_i)\right]+\ldots+y_{iJ}\ln\left[\pi_J(x_i)\right]$$

The maximum possible value of the likelihood for a given data set occurs if the model fits the data exactly. This occurs if observed values are close with the predicted value.

3.6 Model Selection

In ordinal logistic regression methods, it is better to compare models based on Akaike information criterion (AIC) and Bayesian information criterion (BIC) results (Agresti, 2003). AIC judges a model by how close its fitted values tend to the true values, in terms of a certain expected value. Thus, the optimal model is the one that tends to have fit closest to reality in other word the one with smallest AIC is the preferred model.

3.7 Assessment of Model Adequacy

Assessing the appropriateness, adequacy and usefulness of the model is appropriate when fitting ordinal logistic regression model. In this part of the study, the overall goodness of fit of the model tested and importance of each of explanatory variables also assessed by carrying out statistical tests of significance of the coefficients (Agresti, 2003).

Test of Overall Model Fit

Before proceeding to examine the individual coefficients, we should look at an overall test of the null hypothesis that the location coefficients for all of the variables in the model are zero (fit the intercept only model). It can base this on the change in – 2 log-likelihood when the variables are added to a model that contains only the intercept. The change in likelihood function has a chi-square distribution. This value provides a measure of how well the model fits the data.

Goodness of model fit

Goodness of model fit means that how well the model describes the response variable. Assessing goodness of fit involves investigating how close values are predicted by the model with that of observed values (Bewick et al., 2005).

❖ **Deviance and Pearson Chi-Square Goodness-of-Fit Test**

Deviance and Pearson chi-square statistics are used to determine goodness of fit of the model by stating the null hypotheses the model fit the data well. Deviance and Pearson chi-squared statistics can compute on the contingency table formed from tabulating covariate patterns with response categories. Observed and estimated frequencies for each group in each response category can be tabulated in a g × J contingency table.

The deviance goodness of fit statistic for ordinal logistic regression has the form:

$$D = 2\sum_{k=1}^{g}\sum_{j=1}^{J} O_{kj}\, log\left(\frac{O_{kj}}{\widehat{\mathrm{E}}_{\mathrm{kj}}}\right)$$

The Pearson chi square statistic for ordinal logistic regression by:

$$\chi^2 = \sum_{k=1}^{g}\sum_{j=1}^{J}\frac{\left(O_{kj} - \widehat{E}_{kj}\right)^2}{\widehat{E}_{kj}}$$

Where O_{ij} and E_{ij} denote the sums of the observed and estimated frequencies in each group for each response category, respectively, $O_{kj} = \sum_{I\,\epsilon\,\Omega k}\tilde{y}_{Ij}$ and $\widehat{E}_{ij} = \sum_{I\,\epsilon\,\Omega k}\hat{\pi}_{Ij}$; Ω_k denotes the set of indices of the n/g observations in group k; $\tilde{y}_{ij} is$ a binary indicator variable with $\tilde{y}_{ij} = 1$ when y_i = j and $\tilde{y}_{ij} = 0$ otherwise; $\hat{\pi}_{Ij}\, is$ the estimated probabilities of $\pi_{ij} = P(Y = j|x_i)$, $j = 1,\ldots,J$ that is the conditional probability of a response equal to category j for observation i given the explanatory variables x_i.

The reference distribution for both Pearson chi square statistic and deviance statistic is the chi-squared distribution with $(2g-1)(J-1) - p_{cat} - 1$ degrees of freedom, where p_{cat} denotes the number of dichotomous variables needed to model all the categorical covariates. Fall to reject the null hypothesis states that the model fit the data well. As usual large χ^2 and *D* value provides the evidence of lack of fit.

❖ Likelihood-Ratio Test

Likelihood ratio test (LRT) uses the ratio of the maximized value of the likelihood function for the full model over the maximized value of the likelihood function for the null model. The hypothesis that all logistic regressions coefficients are zero except the constant is to be tested.

$$\text{LRT} = -2ln\left[\frac{L_0}{L_f}\right] = -2\{\ln\ L_0 - \ln\ L_f\}$$

Where L_0 is the likelihood function of the null model and L_f is the likelihood function of the full model evaluated at the MLEs. This natural log transformation of the likelihood functions yields an asymptotically chi-squared statistic with degree of freedom equal to the difference between the numbers of parameters estimated in the two models (Menard, 2002).

❖ Pseudo-R ^{2}

To evaluate the goodness-of-fit of logistic models, there are several pseudo R-squares statistics that will used to measure the strength of the association between the dependent variable and the

predictor variables. Pseudo R-squares could range from 0 to 1 (though some pseudo R-squares never achieve 0 or 1). Let L_f be likelihood of the model with predictors and L_0 is likelihood of model with only intercept (null model),

1) Cox and Snell R^2 is computed by

$$R^2 = 1 - \left[\frac{L_0}{L_f}\right]^{2/n}$$

The ratio of the likelihoods reflects the improvement of the full model over the intercept model. Cox & Snell's pseudo R-squared does not attain the value one even if the full model predicts the outcome perfectly. With higher values indicating better model fit.

2) Nagelkerke R^2 is computed by

$$R^2 = \frac{1 - \left[\frac{L_0}{L_f}\right]^{2/n}}{1 - L_0^{2/n}}$$

This is an adjusted version of the Cox & Snell *R*-square that adjusts the scale of the statistic to cover the full range from 0 to 1. With higher values indicating better model fit.

3) McFadden's R^2 is computed by

$$R^2 = 1 - \frac{L_f}{L_0}$$

It can range from 0 to 1, but will never reach or exceed 1, ranging from 0.2 to 0.4 represent good fit model.

CHAPTER FOUR
Data Analysis

4.1 Descriptive Result

Table 4.1 shows the number and percentage distribution of severity of illness by sex, religion, presence of symptom, fever, headache, chest pain, shortness of breath, general weakness, cough, diarrhea, co-morbid, diabetes mullets, asthma, hypertension, cardiovascular disease (CVD) and chronic obstructive pulmonary disease (COPD). Moreover, chi-square test was conducted to test the association between severity of illness and each independent variable. The total patients considered for the study were 454. Out of the total patients, about 240 (52.86%) were at mild stage, 60 (13.22%) were at moderate stage, 106 (23.35%) were at sever stage and 48 (10.57%) were at critical stage.

As shown in Table 4.1, the male and female patients were 300 (66.1%) and 154 (33.9%), respectively. Out of the total male patients, the proportion of severity of illness at mild, moderate, sever and critical stage were 145 (48.3%), 45 (15.0%), 72 (24.0%) and 38 (12.7%), respectively, while the proportion of female patients were 95 (61.7%), 15 (9.7%), 34 (22.1%) and 10 (6.5%) at mild, moderate, sever and critical stage, respectively.

Table 4.1 also shows that about 251 (55.3%) patients had a symptom of COVID-19 whereas 203 (44.7%) patients didn't show any symptom of COVID-19. With regard to COVID-19 symptomatic cases, about 79 (31.5%), 41 (16.3%), 98 (39.0%) and 33 (13.1%) were at mild, moderate, sever and critical stage, respectively, while with regard to asymptomatic patients 161 (79.3%), 19 (9.4%), 8 (3.9%) and 15 (7.4%) were at mild, moderate, sever and critical stage, respectively.

As shown in Table 4.1, out of 177 (39.0%) patients who had co-morbid, about 20.9%, 14.1%, 44.1% and 20.9% patients were at mild, moderate, sever and critical stage, respectively, whereas the proportion of patients who had no any co-morbid (277 (61.0%)) at mild, moderate, sever and critical stage were 73.3%, 12.6%, 10.1% and 4.0%, respectively.

Table 4.1: Descriptive result of severity of illness by independent variables

Variables	Category	Severity of illness					Chi-square	Df	p-value
		Mild	Moderate	Sever	Critical	Total			
		Number (%)	Number (%)	Number (%)	Number (%)	Number (%)			
Total		240(52.86)	60(13.22)	106(23.35)	48(10.57)	454			
Sex	Female	95 (61.7)	15 (9.7)	34 (22.1)	10 (6.5)	154 (33.9)	9.392	3	.025*
	Male	145 (48.3)	45 (15.0)	72 (24.0)	38 (12.7)	300 (66.1)			
Religion	Muslim	37 (52.9)	10 (14.3)	16 (22.9)	7 (10.0)	70 (15.4)	0.11	3	.991
	Christian	203 (52.9)	50 (13.0)	90 (23.4)	41 (10.7)	384 (84.6)			
Symptom presence	Asymptomatic	161 (79.3)	19 (9.4)	8 (3.9)	15 (7.4)	203 (44.7)	115.46	3	.000*
	Symptomatic	79 (31.5)	41 (16.3)	98 (39.0)	33 (13.1)	251 (55.3)			
Fever	No	226 (58.5)	44 (11.4)	77 (19.9)	39 (10.1)	386 (85.0)	35.49	3	.000*
	Yes	14 (20.6)	16 (23.5)	29 (42.6)	9 (13.2)	68 (15.0)			
Headache	No	212 (56.4)	41 (10.9)	82 (21.8)	41 (10.9)	376 (82.8)	16.43	3	.001*
	Yes	28 (35.9)	19 (24.4)	24 (30.8)	7 (9.0)	78 (17.2)			
Chest pain	No	232 (55.2)	55 (13.1)	89 (21.2)	44 (10.5)	420 (92.5)	17.27	3	.001*
	Yes	8 (23.5)	5 (14.7)	17 (50.0)	4 (11.8)	34 (7.5)			
Shortness of breath	No	227 (64.7)	47 (13.4)	53 (15.1)	24 (6.8)	351 (77.3)	106.35	3	.000*
	Yes	13 (12.6)	13 (12.6)	53 (51.5)	24 (23.3)	103 (22.7)			
General weakness	No	218 (60.7)	45 (12.5)	63 (17.5)	33 (9.2)	359 (79.1)	48.46	3	.000*
	Yes	22 (23.2)	15 (15.8)	43 (45.3)	15 (15.8)	95 (20.9)			
Cough	No	207 (62.9)	37 (11.2)	56 (17.0)	29 (8.8)	329 (72.5)	50.34	3	.000*
	Yes	33 (26.4)	23 (18.4)	50 (40.0)	19 (15.2)	125 (27.5)			
Diarrhea	No	233 (53.8)	59 (13.6)	96 (22.2)	45 (10.4)	433(94.4)			
	Yes	7 (33.3)	1 (4.8)	10 (47.6)	3 (14.3)	21 (5.6)			
Co-morbid	No	203 (73.3)	35 (12.6)	28 (10.1)	11 (4.0)	277 (61.0)	138.86	3	.000*
	Yes	37 (20.9)	25 (14.1)	78 (44.1)	37 (20.9)	177 (39.0)			
Diabetes mullets	No	232 (59.0)	52 (13.2)	73 (18.6)	36 (9.2)	393 (86.6)	55.12	3	.000*
	Yes	8 (13.1)	8 (13.1)	33 (54.1)	12 (19.7)	61 (13.4)			

CVD	No	237 (53.5)	59 (13.3)	101 (22.8)	46 (10.4)	443 (97.6)			
	Yes	3 (27.3)	1 (9.1)	5 (45.5)	2 (18.2)	11 (2.4)			
COPD	No	238 (53.7)	58 (13.1)	103 (23.3)	44 (9.9)	443 (97.6)			
	Yes	2 (18.2)	2 (18.2)	3 (27.3)	4 (36.4)	11 (2.4)			
Asthma	No	233 (54.3)	56 (13.1)	94 (21.9)	46 (10.7)	429 (94.5)			
	Yes	7 (28.0)	4 (16.0)	12 (48.0)	2 (8.0)	25 (5.5)			
Hypertension	No	226 (62.1)	45 (12.4)	66 (18.1)	27 (7.4)	364 (80.2)	69.25	3	.000*
	Yes	14 (15.6)	15 (16.7)	40 (44.4)	21 (23.3)	90 (19.8)			

* P-value less than 0.05

The result of chi-square test on Table 4.1 shows that there was a significant association between severity of illness and each independent variable like sex, presence of symptom, fever, headache, chest pain, shortness of breath, general weakness, cough, co-morbid, diabetes mullets and hypertension at 5% level of significance.

Table 4.2 shows the summary result of severity of illness by age of patients. The minimum age of patients at mild, sever and critical stage were 1, 2 and 22 years, respectively, and for patients at moderate stage were six month. The maximum age of patients at mild, moderate, sever and critical stage were 85, 91, 86 and 99 years, respectively. The median age of the patients were 29, 37.5, 58 and 60.5 years at mild, moderate, sever and critical stage, respectively.

Table 4.2: Summary result of severity of illness by age of patient

		Severity of Illness			
		Mild	Moderate	Sever	Critical
Age (years)	Minimum	1.0	0.50	2.0	22.0
	Maximum	85.0	91.0	86.0	99.0
	Mean	33.5	41.7	55.9	57.6
	Median	29.0	37.5	58.0	60.5

4.2 Ordinal Logistic Regression Analysis

4.2.1 Proportional Odds Model

Univariable Proportional Odds Analysis

The univariable ordinal logistic regression analysis was conducted in order to incorporate the variables into the final model at 25% level of significance. Table 4.3 in the appendix, results show that independent variables like sex, age, presence of symptom, fever, headache, chest pain, shortness of breath, general weakness, cough, diarrhea, co-morbid, diabetes mullets, asthma, hypertension, cardiovascular disease and chronic obstructive pulmonary disease had a significant effect on illness severity of COVID-19 patients.

Multivariable Proportional Odds Analysis

Model Adequacy

Table 4.4, shows the result of model adequacy when fitting the proportional odds model. From the table, we observed that the difference between the log-likelihoods of the final and intercept only model (chi-square = 279.578 with degree of freedom = 6; p-value < 0.001). We conclude that the final model is a better fit the data than the model that contains only the intercept.

The result of deviance and Pearson chi-square test for checking the goodness of a model fit given in Table 4.4 results that the null hypothesis that the model fits the data well is not rejected for deviance test but for Pearson chi-square test it was rejected. So we conclude that the model did not fit the data well.

Table 4.4 also shows that the values of pseudo R-square measures; Cox and Snell, Nagelkerke and McFadden were 46.0%, 50.8% and 26.1%, respectively. The pseudo R^2 values (Nagelkerke= 50.8%) indicates that the variables in the model predicted 50.8% of the total variation in the severity of illness. Thus, we conclude that the model fit the data well.

Test of Parallel Line Assumption

The result of test of parallelism shown in Table 4.4 rejects the null hypothesis that the slope coefficients are the same across response categories. This indicates that the assumption of parallelism was violated. Thus we conclude that the slope coefficients are not the same across the response categories.

Table 4.4: Model fitting information, goodness-of-fit, pseudo R-square, test of parallelism and IC

Model Fitting Information				
Model	-2 Log Likelihood	Chi-square	Df	p-value
Intercept Only	1009.682	279.578	6	0.000
Final	730.104			
Goodness-of-Fit				
	Chi-Square	Df	p-value	
Pearson	1195.737	942	0.000	
Deviance	682.358	942	0.998	
Pseudo R-Square				
Cox and Snell	0.460			
Nagelkerke	0.508			
McFadden	0.261			
Test of Parallel Lines				
Model	-2 Log Likelihood	Chi-Square	Df	p-value
Null Hypothesis	730.104	597.632	12	0.000
General	132.472			

Akaike's information criterion and Bayesian information criterion					
Model	ll(null)	ll(model)	Df	AIC	BIC
	-536.4599	-396.4909	9	810.98	848.04

(AIC and BIC were used as the base for future model comparisons)

The results of proportional odds model analysis were given in Table 4.5 in the appendix. Using stepwise variable selection, variables like sex, age, presence of symptom, shortness of breath, general weakness and co-morbid had statistically significant with severity of illness. Using Brant test of parallel line assumption, for variable presence of symptom the assumption is violated.

Since the assumption of parallelism is violated, proportional odds model is not valid. Thus, stereotype ordinal logistic regression model was fitted. Such analysis was required to assess the correct functional form of the covariates to build models with adequate goodness-of-fit.

4.2.2 Stereotype Ordinal Logistic Regression Model

Univariable Stereotype Logistic Regression Analysis

Table 4.6 in the appendix, presents the output for the univariable stereotype ordinal logistic regression analysis. The table show that variables like sex, age, presence of symptom, fever, shortness of breath, general weakness, cough, co-morbid, diabetes mullets, hypertension and chronic obstructive pulmonary disease had a significant effect on severity of illness at 25% level of significance.

Multivariable Stereotype Logistic Regression Analysis

As shown in Table 4.7, using stepwise variable selection, the variables like sex, age, presence of symptom, shortness of breath, general weakness and co-morbid had a significant effect on severity of illness. The table also shows a significant chi-square value of 89.58 with 7 degree of freedom (p-value < 0.001) indicating that the full model provides a better fit the data than the null model. And also as the constraints were satisfying the condition $1 = \phi_1 > \phi_2 > \phi_3 > \phi_4 = 0$, it ensures that the outcome variable is ordinal. Moreover, ϕ_j give information about the distance between categories. Thus, the difference between mild stage and moderate stage is 0.495, between moderate and sever is 0.313 and between sever and critical stage is 0.192.

Comparing the PO model with the SL model using AIC were 810.98 and 714.19, respectively, and BIC were 848.04 and 763.60, respectively. Therefore, the SL model had a better fit, which indicated that the SL model was a better choice when the proportional odds assumption violated in the PO model.

Table 4.7: Parameter estimates of stereotype ordinal logistic regression analysis

Variables		Estimate	SE	Z	Sig.	95% CI of estimate	
Sex	Female (ref.)						
	Male	-0.977	0.344	-2.839	0.005	-1.651	-0.303

Age		-0.054	0.010	-5.318	0.000	- 0.074	-0.034
Presence of symptom	Asymptomatic (ref.)						
	Symptomatic	-1.891	0.383	-4.942	0.000	-2.641	-1.141
Shortness of breath	No (ref.)						
	Yes	-1.615	0.437	-3.692	0.000	-2.473	-0.758
General weakness	No (ref.)						
	Yes	-1.001	0.404	-2.475	0.013	-1.794	-0.208
Co-morbid	No (ref.)						
	Yes	-2.174	0.369	-5.883	0.000	-1.449	-2.898
/phi1		1.000					
/phi2		0.808	0.050	16.016	0.000	0.709	0.907
/phi3		0.495	0.056	8.828	0.000	0.385	0.605
/phi4		0.000					
/theta1		-3.869	1.074	-3.602	0.000	-5.975	-1.764
/theta2		-1.828	0.911	-2.007	0.045	-3.613	-0.042
/theta3		-1.313	0.527	-2.492	0.013	-2.346	-0.281
/theta4		0.000					

Mild is the reference category ref. = reference category of the variable

The Wald chi-square test = 89.58, df = 7, p-value < .001, AIC= 714.19 and BIC= 763.60.

The odds of male patient being at critical stage is exp{-(-0.977*1)} = 2.656 times more likely than being at mild stage; the odds of male patient being at sever stage is exp{-(-0.977*0.808)} =2.202 times more likely than being at mild stage and the odds of male patient being at moderate stage is exp{-(-0.977*0.495)} = 1.622 times more likely than being at mild stage, holding the effect of other variable constant.

The odds of female patient being at critical stage is exp{-(0.977*1)} = 0.376 times more likely than being at mild stage or 62.4% less likely than being at mild stage; the odds of female patient being at sever stage is exp{-(0.977*0.808)} = 0.454 times more likely than being at mild stage or 54.6% less likely than being at mild stage and the odds of female patient being at moderate stage is exp{-0.977*0.495 } = 0.616 times more likely than being at mild stage or 38.4% less likely than being at mild stage, holding other variable constant.

For one year increase in age, the odds of patient being at critical stage is 1.055 times more likely than being at mild stage. The odds of patient being at severe stage is 1.045 times more likely than

being at mild stage for one year increase in age. For one year increase in age, the odds of patient being at moderate stage is 1.027 times more likely than being at mild stage, holding other variable constant.

The odds of patient, who have a co-morbid being at moderate, sever and critical stage is 2.93, 5.79 and 8.79, respectively, times more likely than being at mild stage, holding other variable constant. The odds of patient, who have no any co-morbid being at moderate, sever and critical stage is 65.9%, 82.7% and 88.6%, respectively, less likely than being at mild stage, holding other variable constant.

4.3 Discussion of the Result

The results of stereotype ordinal logistic regression model shows that an increasing in age and being male sex had greater risk of developing worse severity of illness. This finding is in line with the findings obtained by Zhang, et al., 2020; Ann, 2020; Zheng et al., 2020; Wu et al., 2020 and Leulseged et al., 2020. On the other hand different finding were obtained from the study done by Gálvez-Barrón et al., 2021 being female was associated with more severe disease.

The presence of symptom was the determinants of developing a more severe COVID-19 illness. This finding is also in line with Leulseged et al., 2020 who had conducted a research on the determinants of COVID-19 disease severity among COVID-19 patients admitted to millennium COVID-19 care center in Ethiopia.

The study findings indicates shortness of breath had greater risk of at worse stage severity of illness and this findings also in line with the finding of Zheng et al., 2020 who conducted research on assessing the risk factors of critical and mortal COVID-19. Having any co-morbidity is also identified as significant factor for developing more severs COVID-19 illness severity. This finding is significant with the findings of Degarege et al., 2020; Wu et al., 2020 and Hu and Wang, 2021.

This study also identified that general weakness of the patient were the risk factors of developing worse illness severity. This finding is consistent with the finding of Mayo Clinic, 2020 who conducted research on who's at higher risk of serious symptoms.

CHAPTER FIVE
Conclusions and Recommendations

5.1 Conclusions

The result of the study using stereotype ordinal logistic regression model showed that variables like sex, age, presence of symptom, shortness of breath, general weakness and co-morbidity had statistically significant effect on the severity of illness in COVID-19 patients.

COVID-19 patients who had any co-morbidity were at greater risk of developing more severe case compared to being at mild stage. Male patients were at risk of developing more severe COVID-19 case whereas female patients were at risk of being at mild stage than being at more severe stage. A similar result has also obtained for patients who had a symptom of COVID-19, having shortness of breath, having general weakness and increasing in age.

5.2 Recommendations

Based on the findings of the study, we forward the following recommendations:

- As being symptomatic is at risk of more sever stage, it is better to test the community before they show the symptom.
- As having any co-morbidity is at risk of more sever stage, it is better to aware them about illness severity.

Reference

- Adrish, M., Chilimuri, S., Mantri, N., Sun, H., Zahid, M., Gongati, S., Jog, A.P., Purmessur, P. and Singhal, R., 2020. Association of smoking status with outcomes in hospitalized patients with COVID-19. BMJ open respiratory research, 7(1), p.e000716.
- Agresti, A., 2002. Categorical data analysis. (2nd Ed). John Wiley & Sons, New York.
- Anderson, J.A., 1984. Regression and ordered categorical variables. Journal of the royal statistical society: Series B (Methodological), 46(1), pp.1-22.
- Ann Rheum Dis., 2021. Severity of COVID-19 and survival in patients with rheumatic and inflammatory diseases. Ann Rheum Dis; 80: 527–538.
- Bewick, V., Cheek, L., & Ball, J., 2005. Statistics review 14: Logistic regression. Critical care (London, England), 9(1), 112-118.
- CDC report: Assessing risk factors for severe COVID-19 illness. Updated Nov. 30, 2020.
- Chang, M.C., Park, Y.K., Kim, B.O. and Park, D., 2020. Risk factors for disease progression in COVID-19 patients. BMC infectious diseases, 20(1), pp.1-6.
- Chidambaram, V., Tun, N.L., Haque, W.Z., Majella, M.G., Sivakumar, R.K., Ayeh, S.K. and Salia, E.L., 2020. Factors associated with disease severity and mortality among patients with COVID-19: A systematic review and meta-analysis. PloS one, 15(11), p.e0241541.
- Cochran's, 1977. Sampling technique, (3rd Ed). John Wiley & Sons Inc., New York.
- Corona virus updates: Accessed: January 01, 2021, https://www.worldometers.info/coronavirus/.
- Degarege, A., Naveed, Z., Kabayundo, J. and Brett-Major, D., 2020. Risk factors for severe illness and death in COVID-19: a systematic review and meta-analysis. medRxiv.
- Dong, Y., Mo, X.I., Hu, Y., Jiang, Z. and Tong, S., 2020. Epidemiological characteristics of 2143 pediatric patients with COVID-19 in China. Pediatrics, 145(6), p.e20200702.
- Ethiopia UNICEF, 2020. Socio-economic impacts of COVID-19 7th April 2020. Addis Ababa, Ethiopia.
- FMOH, Ethiopia, 2020. National comprehensive COVID-19 management handbook. (1st Ed). Addis Ababa, Ethiopia.

- Gálvez-Barrón, C., Arroyo-Huidobro, M., Miñarro, A., Añaños, G., Chamero, A., Capielo, A., Ventosa, E. and Tremosa, G., 2020. COVID-19: Clinical presentation and prognostic factors of severe disease and mortality in the oldest-old population. A cohort study.
- Geng, M.J., Wang, L.P., Ren, X., Yu, J.X., Chang, Z.R., Zheng, C.J., Li, Y., Yang, X.K., Zhao, H.T. and Li, Z.J., 2021. Risk factors for developing severe COVID-19 in China: an analysis of disease surveillance data. Infectious diseases of poverty, 10(1), pp.1-10.
- Greenland, S., 1994. Alternative models for ordinal logistic regression. Statistics in medicine, 13(16), pp.1665-1677.
- Hu, J. and Wang, Y., 2021. The clinical characteristics and risk factors of severe COVID-19. Gerontology, 67(3), pp.255-266.
- Jennifer M., 2020. Summary of COVID-19 long-term health effects: emerging evidence and ongoing investigation. Health Sci. J, 14, p.706.
- Lapidus, N., Zhou, X., Carrat, F., Riou, B., Zhao, Y. and Hejblum, G., 2020. Biased and unbiased estimation of the average length of stay in intensive care units in the COVID-19 pandemic. Annals of intensive care, 10(1), pp.1-9.
- Leulseged, T.W., Abebe, K.G., Hassen, I.S., Maru, E.H., Zewde, W.C., Siyoum, D.F., Edo, M.G. and Mesfin, E.G., 2020. COVID-19 Disease severity and determinants among Ethiopian patients: A study of the millennium COVID-19 care center. medRxiv.
- Liu Xing, 2014. Fitting stereotype logistic regression models for ordinal response variables in educational research. Journal of modern applied statistical methods: Iss. 2, Article 31.
- Maclean, R., 2020. Africa braces for coronavirus, but slowly. The New York Times.
- MAYO Clinic: COVID-19: Who's at higher risk of serious symptoms? Archived: https://www.mayoclinic.org/coronavirus-who-is-at-risk/art-20483301.
- McCullagh and Nelder, 1989. Generalized linear models. (2nd Ed). Chapman and Hall, London New York.
- Menard, S., 2002. Longitudinal research, series: quantitative applications in the social sciences, Publication # 76, (2nd Ed), Sage, Thousand Oaks.
- Mishra, Mukesh Kumar, 2020. The world after COVID-19 and its impact on global economy, ZBW – leibniz information centre for economics, Kiel, Hamburg.
- Ngware, M., 2020. Delivering education online: corona virus underscores what's missing in Africa. Africa population and hearth research center, a-13491, 4.

- ONE UN Assessment: Socio-economic impact of COVID-19 in Ethiopia. May 2020, Addis Ababa, Ethiopia.
- Peterson, B., and F. E. Harrell, Jr. Partial proportional odds models for ordinal response variables. Appl. Statistics 1990, 39: 205-217.
- Rees, E.M., Nightingale, E.S., Jafari, Y., Waterlow, N.R., Clifford, S., Pearson, C.A., Jombart, T., Knight, G.M. and CMMID Working Group, 2020. COVID-19 length of hospital stay: a systematic review and data synthesis. BMC medicine, 18(1), pp.1-22.
- Rieg, S., von Cube, M., Kalbhenn, J., Utzolino, S., Pernice, K., Bechet, L., Baur, J., Lang, C.N., and Kern, W.V., 2020. COVID-19 in-hospital mortality and mode of death in a dynamic and non-restricted tertiary care model in Germany. PloS one, 15(11), p.e0242127.
- Roy, S., 2020. Economic impact of COVID-19 pandemic. Purdue University, 7:26
- Sun, J., He, W.T., Wang, L., Lai, A., Ji, X., Zhai, X., Li, G., Suchard, M.A., Tian, J., Zhou, J. and Veit, M., 2020. COVID-19: epidemiology, evolution, and cross-disciplinary perspectives. Trends in molecular medicine, 26(5), pp.483-495.
- Turcotte, J.J., Meisenberg, B.R., and MacDonald, E.B., 2020. Risk factors for severe illness in hospitalized COVID-19 patients at a regional hospital. PloS one, 15(8), p.e0237558.
- Walters SJ. Campbell M.J. Lall R., 2001. Design and analysis of trials with quality of life as an outcome: a practical guide. Journal of biopharmaceutical statistics, 11(3), pp.155-176.
- WHO, Diagnostics for Corona virus Disease 2019 (COVID-19) patients updated: Jun 16, 2020. Available from https://emedicine.medscape.com/article/2500119-overview#a10.
- Wu, Y., Li, H. and Li, S., 2020. Clinical determinants of the severity of coronavirus disease 2019 (COVID-19): A systematic review and meta-analysis.
- Xu K, Zhou M, Yang D, Ling Y, Liu K, Bai T, Cheng Z, Li J (2020). Application of ordinal logistic regression analysis to identify the determinants of illness severity of COVID-19 in China. Epidemiology and Infection 148, e146, 1–11.
- Zhang, S.Y., Lian, J.S., Hu, J.H., Zhang, X.L., Lu, Y.F. and Jia, H.Y., 2020. Clinical characteristics of different subtypes and risk factors for the severity of illness in patients with COVID-19 in Zhejiang, China. Infectious diseases of poverty, 9(1), pp.1-10.

Appendix 1

Sample size calculation

By using Zhang, et al. (2020) as reference, we compute the required sample size as follow:

$$n = \frac{6(Z_{\frac{\alpha}{2}} + Z_{\beta})^2}{(logOR)^2 * \{1 - \sum_{j=1}^{4} \bar{\pi}_j{}^3\}}$$

The proportions in group 1 are p11, p12, p13 and p14 (where p11+ p12+ p13 + p14 = 1) with similar notation for group 2.

Here C11, C12, C13 and C14 be the cumulative probabilities of the first category, so C11= p11, C12= p11+ p12, etc.

The proportions expected for one group may already be known through a pilot study or from previous research.

$\bar{\pi}_1 = \frac{\text{p11+ p21}}{2}$ $\bar{\pi}_2 = \frac{\text{p12+ p22}}{2}$ $\bar{\pi}_3 = \frac{\text{p13+ p23}}{2}$ And $\bar{\pi}_4 = \frac{\text{p14+ p24}}{2}$

The odds ratio is the chance of a subject being in a given category or lower in one group compared with the other group.

$OR_1 = \frac{\text{CA1}}{\text{l−CA1}} / \frac{\text{CB1}}{\text{l−CB1}}$ And like.

		Mild	Moderate	Severe	critical
Co-morbid	Yes	10 (0.046 =p11) C11=0.046	161 (0.739 =p12) C1=0.785	31 (0.142=p13) C13= 0.927	16 (0.073=p14) C14= 1
	No	42 (0.074 =p21) C21=0.074	497 (0.872=p22) C22= 0.946	30 (0.053 =p23) C23= 0.998	1 (0.002 =p24) C24= 1
		$\bar{\pi}_1$= 0.060	$\bar{\pi}_2$= 0.805	$\bar{\pi}_3$=0.098	$\bar{\pi}_3$= 0.038
		OR1=(c11/1-c11)/ (c21/1-c21) = 0.603	OR2=0.208	OR3= 0.025	AvgOR= 0.2786667

n = 6*(1.96+1.644854)^2/{(log0.2786667)^2*(1-(0.060^3+0.805^3+0.098^3+0.038^3))}

= 77.96983/(0.30794*0.4771378)

= 531.6887 ≈ 532

Appendix 2

Table 4.3: Univariable ordinal logistic regression analysis

Variables		Estimate	SE	Z	Sig.	95% CI of estimate	
Sex	Female(ref.)						
	Male	0.513	0.194	2.649	0.01*	0.133	0.892
Age		0.061	0.006	10.839	0.00*	0.050	0.072
Religion	Muslim (ref.)						
	Christian	0.024	0.246	0.099	0.92	-0.457	0.506
Presence of	Asymptomatic(ref.)						
symptom	Symptomatic	2.024	0.213	9.484	0.00*	1.606	2.442
Fever	no (ref.)						
	yes	1.108	0.230	4.811	0.00*	0.657	1.560
Headache	no (ref.)						
	yes	0.495	0.221	2.245	0.02*	0.063	0.928
Chest pain	no (ref.)						
	yes	1.005	0.308	3.264	0.00*	0.401	1.608
Shortness of	no (ref.)						
breath	yes	2.125	0.222	9.569	0.00*	1.690	2.561
General	no (ref.)						
weakness	yes	1.326	0.212	6.267	0.00*	0.911	1.741
Cough	no (ref.)						
	yes	1.268	0.196	6.469	0.00*	0.884	1.652
Diarrhea	no (ref.)						
	yes	0.864	0.400	2.159	0.03*	0.080	1.648
Co-morbid	no (ref.)						
	yes	2.319	0.207	11.210	0.00*	1.914	2.725
Diabetes	no (ref.)						
mullets	yes	1.648	0.249	6.631	0.00*	1.161	2.135
Asthma	no (ref.)						
	yes	0.766	0.353	2.173	0.03*	0.075	1.458
Hypertension	no (ref.)						
	yes	1.770	0.223	7.954	0.00*	1.334	2.207
Cardiovascular	no (ref.)						
disease	yes	1.023	0.539	1.899	0.06*	-0.033	2.079
COPD	no (ref.)						
	yes	1.515	0.563	2.692	0.01*	0.412	2.617

* P-value less than 0.25 ref. = reference category of the variable

Table 4.5: Parameter estimates of ordinal logistic regression

Variables		Estimate	SE	Sig.	OR	95% CI of estimate		PL test Sig.
Sex	Female (ref.)							
	Male	0.694	0.223	0.002	2.002	0.258	1.131	0.070*
Age		0.036	0.006	0.000	1.037	0.023	0.049	0.278*
Symptom	Asymptomatic (ref.)							
	Symptomatic	1.165	0.267	0.000	3.205	0.641	1.688	0.000
Shortness of	No (ref.)							
breath	Yes	1.035	0.254	0.000	2.815	0.538	1.532	0.928*
General	No (ref.)							
weakness	Yes	0.578	0.249	0.020	1.782	0.090	1.066	0.585*
Co-morbid	No (ref.)							
	Yes	1.420	0.238	0.000	4.135	0.953	1.886	0.828*
cut1		3.655	0.369	-	-	2.932	4.378	-
cut2		4.711	0.405	-	-	3.918	5.504	-
cut3		6.951	0.479	-	-	6.012	7.891	-

* P-value greater than 0.05 ref. = reference category of the variable

Table 4.6: Univariable stereotype ordinal logistic regression analysis

Variables		Estimate	SE	Z	Sig.	95% CI of estimate	
Sex	Female(ref)						
	Male	0.912	0.379	2.606	0.016*	0.169	1.655
Age		0.083	0.011	7.850	0.000*	0.063	0.104
Religion	Muslim (ref.)						
	Christian	0.065	0.446	0.146	0.884	-0.809	0.940
Presence of symptom	Asymptomatic(ref.)						
	Symptomatic	1.500	0.340	4.408	0.000*	0.833	2.167
Fever	no (ref.)						
	yes	1.315	0.461	2.852	0.004*	0.411	2.219
Headache	no (ref.)						
	yes	-0.0002	0.002	-0.133	0.895	-0.004	0.003
Chest pain	no (ref.)						
	yes	0.969	0.634	1.526	0.126	-0.273	2.212

Shortness of breath	no (ref.)						
	yes	2.860	0.406	7.048	0.000*	2.065	3.655
General weakness	no (ref.)						
	yes	1.505	0.383	3.925	0.000*	0.754	2.256
Cough	no (ref.)						
	yes	1.413	0.350	4.042	0.000*	0.728	2.099
Diarrhea	no (ref.)						
	yes	0.797	0.709	1.124	0.261	-0.593	2.187
Co-morbid	no (ref.)						
	yes	2.915	0. 387	7.530	0.000*	2.157	3.674
Diabetes mullets	no (ref.)						
	yes	2.269	0.490	4.627	0.000*	1.308	3.230
Asthma	no (ref.)						
	yes	-0.0005	0.0004	-0.121	0.903	-0.008	0.007
Hypertension	no (ref.)						
	yes	2.530	0.401	6.315	0.000*	1.745	3.315
Cardiovascular disease	no (ref.)						
	yes	1.234	0.927	1.331	0.183	-0.583	3.051
COPD	no (ref.)						
	yes	2.381	0.881	2.702	0.007*	0.654	4.109

* P-value less than 0.25 ref. = reference category of the variable