MODELLING CORONAVIRUS INFECTIONS PROGRESSION IN SOUTH AFRICA

BY

Emeritus Professor Eric O Udjo (PhD, London)
Department of Economics, University of South Africa
Independent Consultant
Eric Udjo Consulting
www.ericudjoconsulting.co.za
Email: info@ericudjoconsulting.co.za

ABSTRACT

Although some people infected with COVID-19 have recovered, it remains a deadly disease. The long-term effect of COVID-19 on national mortalities and the economy is unknown. The first case of COVID-19 in South Africa was confirmed on 5 March 2020. Since then the Government has attempted to track down and test all those the first case may have been in contact with and if necessary, isolate/quarantine those testing positive. The number of infections from COVID-19 has continued to increase and it seems the national numbers of infections officially reported by the Ministry of Health do not reflect the actual number of infections in the country. This poses a challenge for the effective management of COVID-19. To mitigate the effect of COVID-19, the Government like some other countries announced full national lock-down on 23 March 2020. The effect of lock-down on the progression of COVID-19 is unknown. Fitting mathematical/demographic models to the figures on COVID-19 progression released by the Ministry of Health, this study assessed the plausibility of the progression numbers, the effect of lock-down, and projected case fatality rates from COVID-19. At the time of this study, less than 1% of the total population of South Africa had been tested for COVID-19. The results of this study suggest that official figures of COVID-19 under-state the number of COVID-19 infections in the general population. Although the prevalence rates of infections among those tested ranged between 27 and 32 per thousand persons tested during the lock-down period, the number of infections continued to increase, and the daily growth rates showed an upward trend during the lock-down period. It is reasonable and logical to argue that the full lock-down in South Africa may have lowered the slope of the gradient in the daily number of infections but the empirical evidence to prove that is lacking at this point. There is no evidence from this study that the national full lock-down flattened the curve of COVID-19 progression in South Africa.

INTRODUCTION

Background:

According to the World Health Organisation (WHO, 2019) Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus. There are several myths about the virus circulating in different forms that are not scientifically proven. The WHO has attempted to debunk some of these myths using its website. According to the WHO (2019), currently, the source of SARS-CoV-2, the coronavirus (CoV) causing COVID-19 is unknown but they note that all available evidence suggests that SARS-CoV-2 has a natural animal origin and is not a constructed virus. The virus that causes COVID-19 is mainly transmitted through droplets generated when an infected person coughs, sneezes, or speaks (WHO, 2019).

The first human cases of COVID-19 were identified in Wuhan City, China in December 2019. At this stage, it is not possible to determine precisely how humans in China were initially infected with SARS-CoV-2 (WHO, 2019). According to the WHO (2019), SARS-CoV, the virus which caused the SARS outbreak in 2003, jumped from an animal reservoir (civet cats) to humans and then spread between humans. The WHO (2019) further notes that in a similar way, it is thought that SARS-CoV-2 the coronavirus jumped the species barrier and initially infected humans, but more likely through an intermediate host, that is another animal species more likely to be handled by humans - this could be a domestic animal, a wild animal, or a domesticated wild animal and, as of yet, has not been identified. The origin of the virus is not the focus of this study but on the numbers relating to the virus.

On 11 March 2020, the WHO declared COVID-19 a pandemic. The WHO website states that since the first human cases of the COVID-19 were identified in Wuhan City, China in December 2019, the virus has spread to 215 countries/territories as at 12 May 2020, 58 of these countries/territories are in Africa. According to WHO (2020) website, as at 12 May 2020, globally, there were 4,088,848 confirmed cases of COVID-19 with 46,839 (about 1.1%) of the confirmed cases located in Africa. Furthermore, according to WHO (2020) website, there were 283,153 confirmed COVID-19 deaths globally as at 12 May 2020. Coverage and efficiency in screening and testing for COVID-19 as well as in the determination of cause of death would vary from country to country, thus global figures on confirmed COVID-19 cases as well as the number of COVID-19 fatalities are likely to be under-stated at any point in time. Different countries have introduced different measures ranging from advocating social/physical distancing, frequently sanitising and washing of hands to partial or total lock-down to contain the spread of the virus to mitigate the demographic and socio-economic impact of the virus. Whether these measures are effective remain to be seen and require long term evaluation.

Statement of the Research Problem

Although many people infected with COVID-19 have recovered, it remains a deadly virus among all age groups in the population particularly the elderly and those who have a pre-existing condition (WHO: 2019) because there is currently no cure for it, and currently no vaccine has been developed to protect against the virus. The measures that different countries have taken to contain the virus have had devasting effect on the economy nationally and locally ranging from the airline industry and tourism in general, to small and medium enterprises as well as the self-employed. It is unknown the long-term effect of COVID-19 on national mortalities and the economy because it is unknown when COVID-19 would run its course, or an effective vaccine would be developed for massive roll out in different countries.

In the history of any epidemic/pandemic, the course of the epidemic/pandemic in human populations resembles a stretched S-curve (or sigmoid/logistic curve): prior to the epidemic, the prevalence is zero infections but once it enters a population, the prevalence rises steeply or exponentially and then plateau before the prevalence starts to decline and then eventually take a bell-shape. Countries are at different stages of the COVID-19 curve and the emphasis in different countries in response to COVID-19 is to flatten the curve i.e. reach the plateau of the curve as soon as possible. With the exception of China it is unknown when countries will reach the plateau of the curve, thus the number of infections taking into consideration current trends, would probably continue to rise exponentially resulting in increased fatality from the virus depending on the effectiveness of the health system in each country to manage those infected.

The first case of COVID-19 in South Africa was confirmed on 5 March 2020, a 38-year-old man who had visited Italy on holiday with eight friends (news24: 08/03/2020), thus the first case of COVID-19 in South Africa was imported. Although the Government has attempted to track down and test all those the first case may have been in contact with and if necessary, isolate/quarantine those testing positive, it is not practically possible that all the possible contacts from this first positive case can be identified. To illustrate the problem, the first positive case aside the eight friends, may have been in contact with other people whom he may not have been aware of from the time he arrived from Italy before he became symptomatic. Such people therefore cannot be identified and tested for the virus.

The same argument holds for each of the eight friends some of whom eventually tested positive. If it had been possible to identify all the people the first case had contact with. Similarly for all subsequent imported cases if it had been possible to identify and tested all the contacts, one would not expect an exponential rise in the number of COVID-19 infections in South Africa. That there has been an exponential increase up to a point in the number of infections since the first case was confirmed suggests that not all COVID-19 infections arising from international and local transmissions were identified. It would appear therefore, that the current total national numbers of infections officially reported by the Ministry of Health probably do not reflect the actual number of infections in the country. This poses a challenge for effective management of COVID-19. A better reflection of the actual number of infections at a given point in time in the country would be through screening of the entire population of South Africa but this would pose a logistic nightmare and a huge strain on resources. As a means of mitigating the effect of the virus, the Government of South Africa announced on 23 March 2020 a national 21-day full lock-down (later classified as level 5 lock-down) with effect from 26 March 2020 and lasting up to 15 April 2020. The lock-down has now been extended. It is not possible for security forces to police every part of the country to enforce the lock-down. Some people may not comply with the lock-down and this can potentially spread the virus. The probable effect of the lock-down on the progression of COVID-19 in South Africa is unknown.

Objectives

In view of the above, the objectives of this study were to model the progression of the COVID-19 in South Africa. Specifically, the study:

1. Assessed the plausibility of the official numbers of COVID-19 infections in South Africa.
2. Assessed the effect of the full national lock-down on progression of COVID-19 infections including prevalence rates in South Africa.
3. Computed and projected fatality from COVID-19 in South Africa.

DATA AND LIMITATIONS

The basic information for the modelling consisted of the daily numbers of confirmed cases of COVID-19, deaths and total number of persons tested (when available) as announced by the Minister of Health in Media Release statements. These numbers were captured in EXCEL spreadsheet by the author and transformed in a manner amenable to analysis. For example, figures for Day 5 from the start of the pandemic has a reference date of 9 March 2020 being the fifth day from the date of the first confirmed case (5 March 2020) of COVID-19.

Two limitations of the data need mentioning. Firstly, is coverage. Coverage of testing for COVID-19 is still extremely low. As of day, 69 from the start of the pandemic (i.e. 12 May 2020), the reported number of persons tested by the Ministry of Health was 369,697 nationally. This constituted only about 0.64% of the estimated mid-2020 total population of South Africa (see Udjo, 2020 for the population estimates). Secondly, the 369,697 persons tested may not be a representative sample of South Africa's population. These were largely people who showed symptoms or some of whom had been in close contact with persons who showed symptoms of COVID-19. It is arguable therefore that these were probably high-risk persons and therefore could bias estimates upward. These limitations should be borne in mind when interpreting the results presented in this study.

METHODS

Assessing the plausibility of the Official Numbers of COVID-19 Infections

A scatter plot was made of the official number of positive COVID-19 cases (Y-axis) released by the Ministry of Health for each day since the start of the pandemic in South Africa (X-axis). The scatter plot revealed an exponential curve. An exponential mathematical curve was then fitted to the scatter plot. The equation defining the mathematical curve may be expressed as:

$Y = \alpha * e^{(\beta x)}$ ……………….(1)

Where Y is the model estimate of the number of COVID-19 infections, α and β are respectively the intercept and slope of the curve and x is the number of days since the start of the pandemic in South Africa. Using this equation, model numbers of the number of COVID-19 infections for each of the days since the start of the epidemic were estimated and plotted in the same scatter plot for comparison with the official numbers. Note that equation (1) implies a constant exponential rate of growth (the β parameter). However, epidemics and pandemics do not increase at a constant rate exponentially forever. Thus, if equation (1) were applied to estimate the number of COVID-19 infections, it would generate exaggerated numbers of COVID-19 infections the further away from the start of the pandemic. Though an epidemic curve may be exponential at the beginning, it eventually manifests as a stretched S as earlier indicate for the half-life of the epidemic. This means that at the early stage of an epidemic, the growth rates of the number of persons infected are high but as the epidemic matures, the growth rates get lower and stabilise and eventually reduce as the epidemic runs its course with or without intervention. In view of this, the best fit to the official figures in the lower part of the curve using equation (1) was examined. These were the points corresponding to days 17-19 since the start of the pandemic in South Africa.

To avoid generating exaggerated model numbers of COVID-19 infections in the population, it was assumed: (A) that the persons tested for COVID-19 in the population is a random sample of the population. This assumption is not necessarily true and as earlier noted, probably a higher risk group than in the general population. (B) The daily rates of growth in the number of positive cases since the start of the pandemic are a better reflection than a single constant exponential growth implied in equation (1) and therefore a more reasonable estimate of the rate of progression of positive cases of COVID-19 in the general population. Based on the above assumptions and starting from day 17 since the start of the pandemic, revised model numbers of COVID-19 infections were estimated using the following equations:

$$r = LN((COVID19^{o}_{dn}/COVID19^{o}_{dn-1})/(d_{n}-d_{n-1})) \quad \text{....................(2)}$$

where r is the rate of growth in the number of observed positive cases, $COVID19^{o}$ among those tested for *COVID-19* on a specified day, *dn, LN* is natural logarithm.

$$COVID19^{m}_{dn} = COVID19^{m}d_{n-1} * e^{rt} \quad \text{..(3)}$$

Where $COVID19^{m}$ is the model estimate of the number of COVID-19 infections on a specified day, *dn*, r is as defined above and varies rather than a constant, $t = d_{n}- d_{n-1}$ is as defined above and e is exponential function.

Estimating Probable Number of COVID-19 cases

The probable number of COVID-19 cases in South Africa at a specified date can be estimated theoretically from the prevalence rate of the virus in the general population. The estimation equation for computing prevalence (total number of cases existing at a point in time) rate of COVID-19 (or any disease) may be expressed as:

$$PrevCOVID19^{t} = (COVID19^{t}_{c}/Pop^{t}) * k \quad \text{.......................(4)}$$

Where $PrevCOVID19^t$ is the prevalence rate at a specified point in time, $COVID19^t{}_c$ is the total number of COVID-19 cases (old plus new cases) existing at a specified point in time, Pop^t is the mid-year population at the specified date (the total number of person-years of exposure to the risk of an event in a period is usually estimated as the mid-year population) , *k* is a constant which could be 100, 1,000 etc arbitrarily chosen (depending on the rarity of the specific disease) for the purpose of comparison between different population. Incidence rate (i.e. new infection rate) among persons tested was computed as

$$INCOVID19^t = (NEWCOVID19^t{}_c/TPop^t) \;\; * k \qquad \ldots\ldots\ldots\ldots\ldots(5)$$

Where $INCOVID19^t$ is the incidence rate among persons tested at a specified time, $NEWCOVID19^t{}_c$ is the number of new positive cases of COVID-19 at a specified time and $TPop^t$ is the total number of persons tested at a specified time.

Estimates of the denominator in the right side of equation (4) are available, see for example Udjo (2020) but the numerator is unknown. Given the assumptions above one may use the number of persons that have been tested and those testing positive as a best guess of prevalence. This may then be used to estimate the probable number of COVID-19 cases to be expected in the population at a specified date. Thus, equation (4) was modified as:

$$TPrevCOVID19^t = (TCOVID19^t{}_c/TPop^t) \;\; * k \qquad \ldots\ldots\ldots\ldots\ldots(6)$$

Where $TPrevCOVID19^t$, $TCOVID19^t{}_c$, $TPop^t$ respectively refer to the prevalence among those tested, those testing positive, and the total number of persons tested at a specified time. The results from equation (5) were examined to determine the time from which the prevalence started to stabilise. The average of the prevalence rates from this point onwards was chosen as the best estimate of prevalence in the general population. Since the group tested for COVID-19 could be a higher risk group, the estimate of the probable number of COVID-19 cases in the general population derived from the equation (6) would probably be the upper limit of the number of positive cases of COVID-19 to be expected in the general population at a specified date. This number was derived using the following equation:

$$ProbCovid\text{-}19 = Pop^t \;\; * \;\; (TPrevCOVID19^t/k) \qquad \ldots\ldots\ldots\ldots\ldots(7)$$

Where *ProbCovid-19* is the estimated probable number of COVID-19 cases expected in the general population at a specified date. Udjo (2020) provides estimates of mid Pop^t for 2020 at national, provincial, district and local municipalities levels. It is the estimate at national level that is relevant to this study. The estimate of *ProbCovid-19* is therefore for the South African population as at the end of 2020.

The 95% confidence interval of the estimate of *ProbCovid-19* was based on the standard error of $TPrevCOVID19^t$. Treating prevalence as a proportion, the standard error was estimated as:

$$SE(TPrevCOVID19^t) = \text{Square root } TPrevCOVID19^t ((1- TPrevCOVID19^t)/n^t)\ \ldots\ldots\ (8)$$

Where *n* is the total number of persons tested at a specified time. Consequently, the confidence interval of *ProbCovid-19* was estimated as

$$IntProbCovid\text{-}19 = TPrevCOVID19^t \pm \{1.96 * (SE(TPrevCOVID19^t))\} * Pop^t \quad \ldots(9)$$

Where *IntProbCovid-19* is the confidence interval of *ProbCovid-19, 1.96* is the 95% confidence.

Regarding Pop^t, Udjo (2017) showed that Statistics South Africa's official figures from the 2011 census overstated the population of South Africa arising mainly from the over estimation of Gauteng's population in the official figures. Consequently, subsequent mid-year estimates from the organisation using the 2011 census figure as a base overstate the size of South Africa's population as well as that of Gauteng province. This has implication for modelling of COVID19-19 in South Africa. It is the contention of this author that any model that uses Pop^t as a parameter in estimating absolute numbers, for example, the expected number of infections at a specified time, would probably overstate such numbers.

Effect of Lock-down on Progression of COVID-19

The effect of lock-down on the progression of COVID-19 in South Africa was done in two ways: Firstly, the daily progression of the growth rates of the number of persons testing positive before and during the lock-down period was examined bearing in mind that in any epidemic, the growth rate of infections of the disease in the population initially increases exponentially but eventually stabilises.

The formula for computing the growth rate was given in equation (2) above. Secondly, the progression in the number of infections before and during the lock-down period was examined. Projecting the number of COVID-19 infections was given in equation (3) above. The specific periods of interest in assessing the effect of lock-down are described as follows. Full lock-down (later classified as level 5 lock-down) was announced by South Africa's President on 23 March 2020 to take effect on 26 March 2020 i.e. a-3-day lag following the announcement before the lock-down came into effect. Given a 14-day incubation period of the virus, COVID-19 infections corresponding to 23-26 March should show up on 5 April – 8 April 2020. This study explored whether there was a spike in the growth in the number of infections on 5 – 8 April 2020 due to crowding arising from panic buying (unintended consequence if any). Similarly, the study explored the progression in growth rates in the number of infections during the period of the full lock-down itself. As already noted, the initial end of the full lock-down was 15 April 20 but subsequently extended to 30 April 2020. Given a 14-day incubation period of the virus, the period that one should examine in the growth rates in the number of infections during the full lock-down is 8 April – 14 May 2020 i.e. days 35-71 of the virus in South Africa. Full lock-down (i.e. level 5 lock-down) was downgraded nationally to level 4 lock-down (announced by the President on 23 April 2020) to take effect from 1 May 2020. Level 4 lock-down is still a full lock-down except some selected businesses allowed to operate under strict conditions. It was made mandatory that all persons should wear face mask (irrespective of the type of mask) outside home from 1 May 2020 during level 4 lock-down. Post-full lock-down COVID-19 infections period would be 28 May 2020 onwards i.e. day 85 onwards. The outcome of all the measures announced by the Government should be evident in the progression of the virus pre- and post-full lock-down. Infections during days 35-71 and day 85 onwards are therefore critical in assessing the effect of lock-down in the progression of the virus in South Africa. For this initial study, the focus was specifically on days 35-71 of the pandemic in South Africa.

Computing, Projecting Case Fatality Rates and Number of Deaths from COVID-19

Projecting the number of fatalities from COVID-19 involved the following steps. Firstly, the reported number of deaths from COVID-19 in South Africa were cumulated from the first case of reported death (which was on day 22 of the pandemic) upwards. Next the cumulated crude fatality rate was computed as

$$CCFatRateCOVID19^t = (CDeathCOVID19^t / CTCOVID19^t) * k \quad \ldots\ldots\ldots\ldots\ldots(10)$$

Where $CCFatRateCOVID19^t$ is the cumulative crude case fatality rate from COVID-19 at a specified time, $CDeathCOVID19^t$ is the cumulated number of COVID-19 deaths at a specified time, $CTCOVID19^t$ is the total number of persons tested as at the specified in time, k is a constant set at 1,000.

An examination of the values of $CCFatRateCOVID19^t$ showed that the rates stabilised from day 43 onwards ranging from values of about 16.7 and 19.72 i.e. a crude fatality rate of between 16.7 deaths and 20.01 per thousand confirmed cases. Next the crude cumulative death rates from day 43 onwards were plotted on a graph. The intercept, a, and slope, b of these values were then estimated by fitting a straight line by least squares to the values. Using these constants, the cumulative crude fatality rates were projected to mid-2020. Based on earlier assumptions, the number of COVID-19 deaths in the general population was then estimated as

$$EDeathCOVID19^t = (PCCFatRateCOVID19^t/1000) * ProbCovid\text{-}19^t \quad \ldots\ldots\ldots \quad (11)$$

Where $EDeathCOVID19^t$ is the estimated number of COVID-19 deaths in the general population at a specified time, $PCCTFatRateCOVID19^t$ is the projected cumulative crude fatality rate due to COVID-19 at mid-year, $ProbCovid\text{-}19^t$ is as earlier defined.

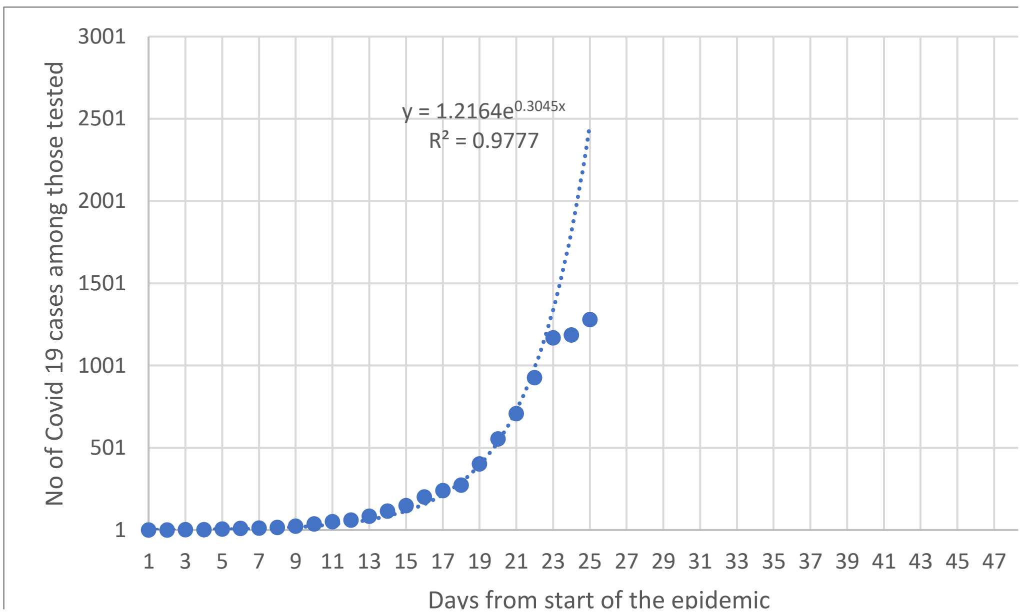
RESULTS

Plausibility of the Official Numbers of COVID-19 Infections in South Africa

Figure 1 is a graphical illustration of the official numbers of COVID-19 infections by days since the start of the pandemic in South Africa up to day 25. The blue dots are the official numbers reported while the broken line is an exponential curve fitted to the official numbers. Firstly, the graph shows that at the initial stage, COVID-19 is time dependent exponential progression hence the fitted curve shows a r^2 of 0.98 resulting in a correlation coefficient r, of 0.98. There is nothing new about this. It only confirms that COVID-19 is behaved like any other historical epidemic at the initial stage.

Figure 1 was deliberately truncated at day 25 of the epidemic because of a puzzling pattern. It can be seen from the graph that there is a sudden drop in the gradient of the curve starting from day 24 (28 March 2020) from the start of the pandemic in South Africa i.e. two days after full national lock-down commenced. Full national lock-down commenced on 26 March 2020 i.e. day 22 from the start of the epidemic in South Africa. Because of the time lag in the incubation period of COVID-19, the sudden drop in the gradient of the curve starting on day 24 cannot possibly be the sudden effect of lock-down. The official figure of manifested number of COVID-19 infections (1,187) on day 28 March 2020 (day 24) among persons tested up to 28 March 2020 were infections that already happened 2-14 days prior to March 28, 2020. If lock-down had any effect, one would only expect to see a decline in the gradient in the number of infections from 8 April i.e. from day 35 of the pandemic irrespective of the number of persons tested during the lock-down period. This raises several questions. What was responsible for the sudden change (two days into full national lock-down) in the gradient of the progression COVID-19 in South Africa? Has it got to do with differences in the characteristics of persons who tested from March 14, 2020 onwards (given a 14-day incubation period) compared with persons who tested prior to March 14 2020? Or, is the sudden drop in the gradient an artefact of the data? Some scientists have claimed on national Television that the curve of COVID-19 in South Africa is different from the experience of other countries. They need to provide scientific proof concerning the characteristics of the persons tested before and after March 14, 2020 why this is so resulting in a deviation from what is generally known about an epidemic curve at the early stage of an epidemic.

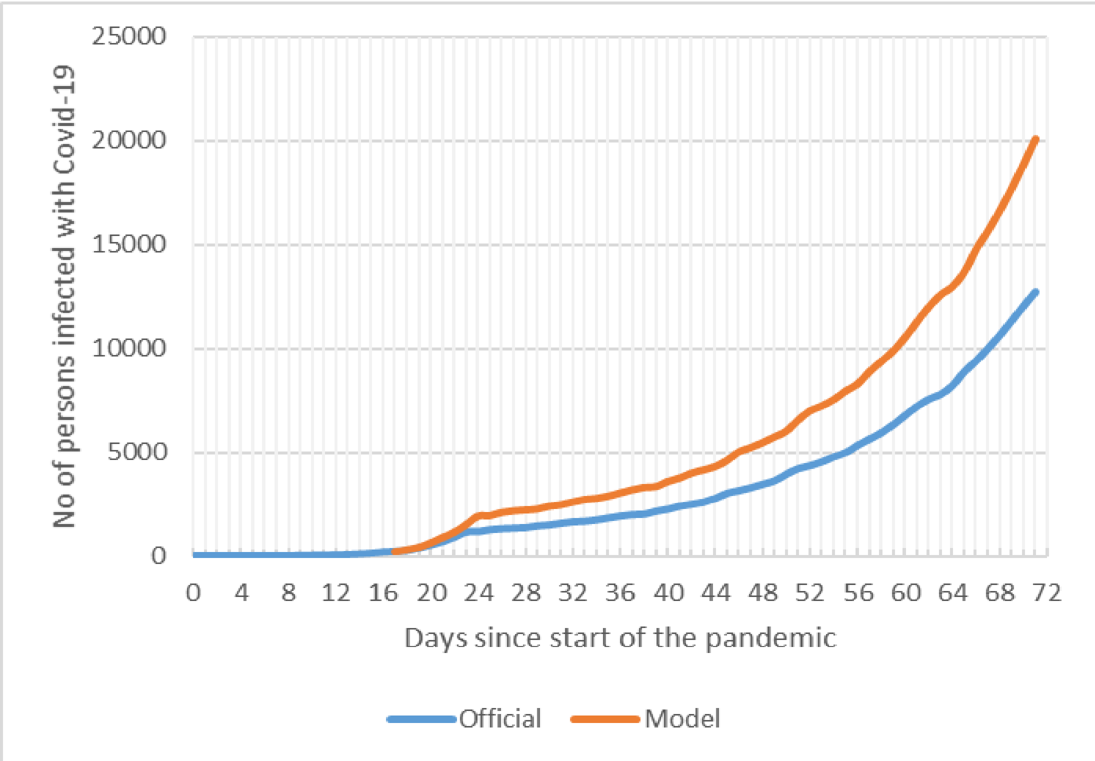
Figure 1: Number of COVID-19 Infections Among Those Tested in South Africa



Source: Author’s graph from official numbers

Figure 2 compares a model estimates of the number of COVID-19 infection with the official numbers using the methods described above. The graph suggests that the official numbers of COVID-19 infections do not reflect the number of infections in the general population. The model suggests that official figures probably underrepresent the number of COVID-19 infections in South Africa. This is understandable because coverage of testing is still exceptionally low and some people who are infected may be in the asymptomatic and therefore not captured by the Government's current efforts. For example, whereas on day 65 of the pandemic i.e. 8 May 2020, the official number of COVID-19 positive cases reported in the media release was 8,895 but the model suggests there were probably about 13,700 COVID-19 cases in the country as of 8 May 2020, a difference of 4,805 between the official and model figures.

Figure 2: Model Estimates of COVID-19 Infections Compared with Official Figures in South Africa



Source: Author's model from official numbers

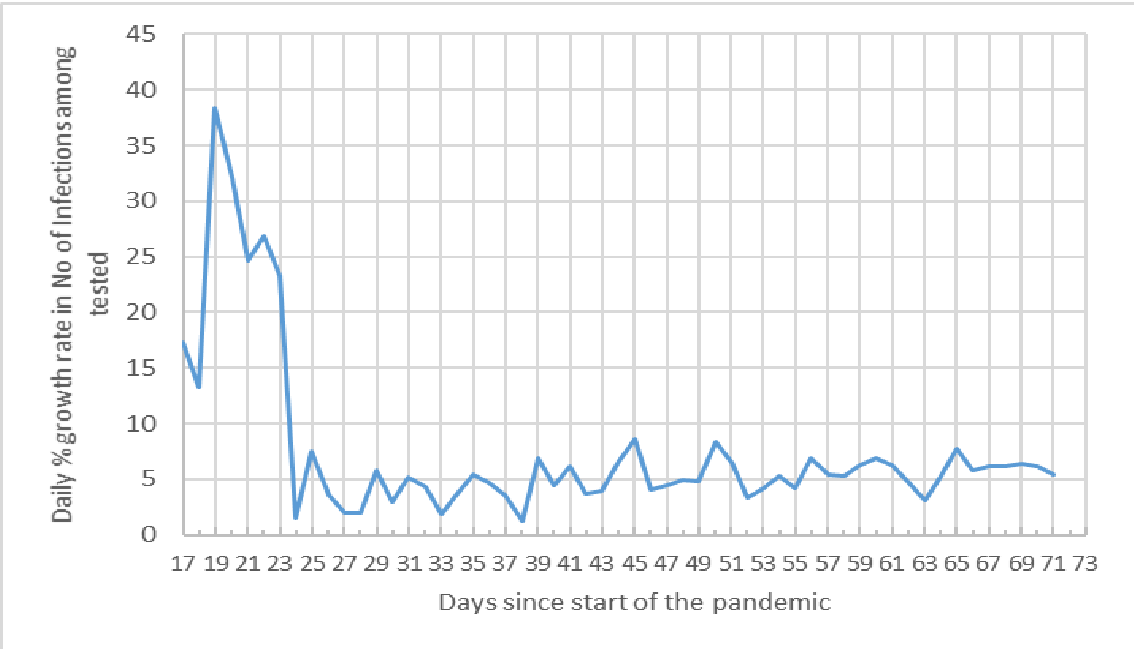
Effect of Lock-down on Progression of COVID-19 in South Africa

Another important point to note in Figure 2 is the gradient in the official numbers for days 35 -71 when the effect of lock-down should have been evident if any. There is no decline in the gradient in the official number of infections during days 35-71. Another way of looking at the effect of lock-down is to examine the percent daily growth rates in the number of infections among those tested. The percent growth rates are derived from equation (2) but multiplying the values of r in the equation by

100 to convert to percentages. Firstly, the focus was on the 3-day lag between 23 March 2020 (when the lock-down was announced) and 26 March 2020 (when the lock-down came into effect). As already noted, the COVID-19 positive cases corresponding to this would be for days 32-35 from the start of the pandemic in South Africa. The second focus was the growth rates from days 35-71 from the start of the pandemic in South Africa which correspond to 8 April 2020 – 14 May 2020. This period provides an assessment of the general effect of the full national lock-down on the progression of COVID-19.

Figure 3 highlights the percent daily growth rates from day 17 up to day 71. As seen in the graph, the daily growth rate of COVID-19 infections apparently dropped sharply from about 25% on day 22 to less than 5% on day 24. This is related to the odd sudden change in the gradient of the progression of the virus noted earlier. Turning to days 32-35, there was a marginal increase in the growth rate between day 34 and day 35. This corresponds to the 3-day lag between 23-26 March 2020 in which there was panic buying in those three days before the lock-down came into effect.

Figure 3: Percent Daily Growth Rates in the Number of COVID-19 before and During full National Lock-down in South Africa

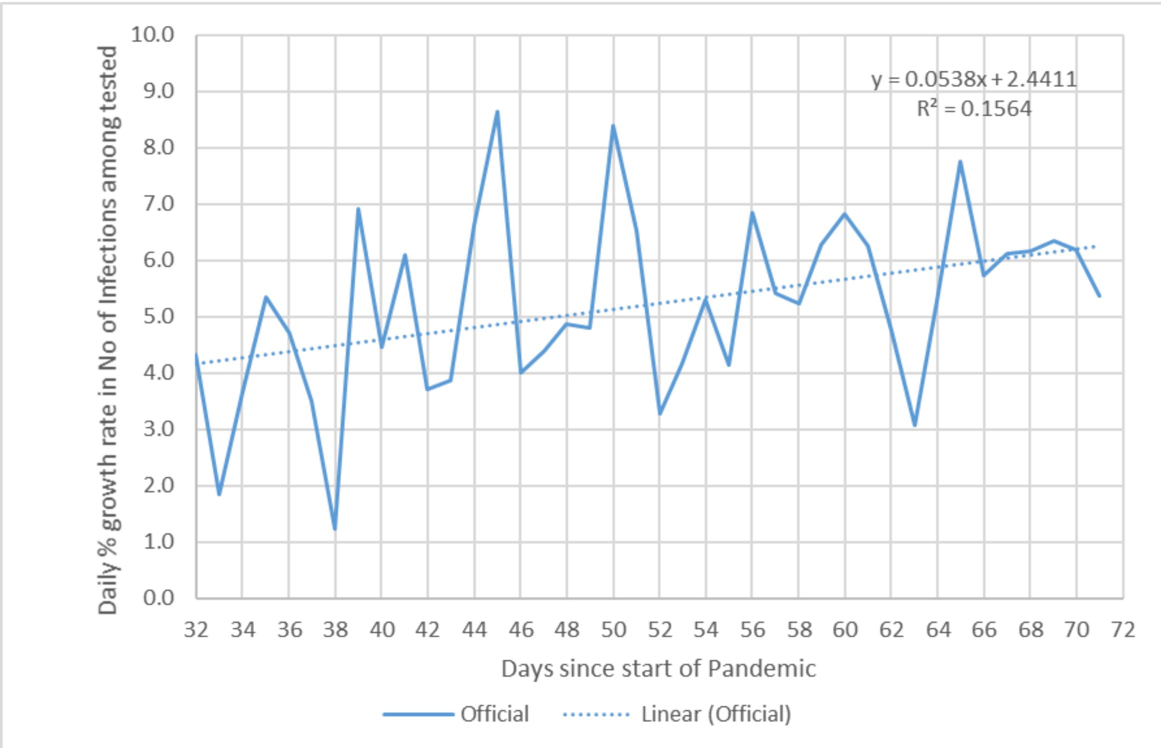


Source: Author's model from official numbers

Focusing on the entire period 35-71, the daily growth rates are erratic probably partly due to time lag in confirmation of some of the positive cases. For example, the daily growth rates declined marginally between days 35 and 38 then increased marginally between days 38 and 39. The graph suggests that the daily growth rate increased by about 4 percentage points between days 43 and 44. What is most evident from Figure 3 is that in the early days of the epidemic, daily growth rates of the infections were high but apparently dropped sharply after day 22. As previously argued, the sharp drop cannot be attributed to lock-down because the transition period was only two days after the full national lock-down commenced.

Figure 4 amplifies the daily growth rates by focusing on the period that infections occurred during the lock-down period assuming a 14-day incubation period in COVID-19. The erratic pattern described earlier is evident. Fitting a straight line to the trend by least squares suggests a slightly upward trend in the daily growth rates in the number of infections that occurred during the lock-down period. However, the correlation coefficient r, is 0.40 implying weak effect of lock-down period on the daily growth rates of infections among those tested.

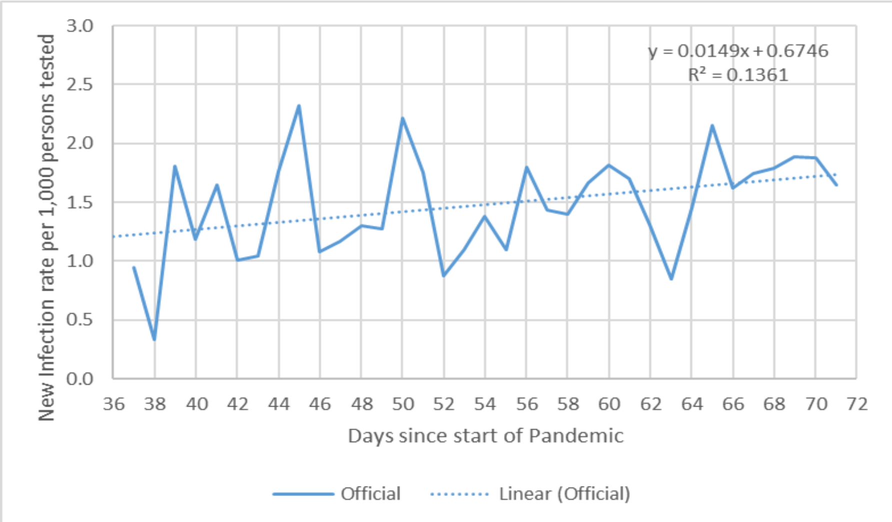
Figure 4: Percent Daily Growth Rates in the Number of COVID-19 Infections that Occurred during Lock-down among those Tested in South Africa



Source: Author's model from official numbers

The effect of the full lock-down on the progression of COVID-19 was also examined by looking at the trend in the incidence rates focusing on days 35-71 from the start of the pandemic. The rates for this period correspond to the new infections that occurred during the full national lock-down period. The rates are illustrated in Figure 5. As seen in the graph, the rates are erratic. There were two pronounced peaks in new infection rates on days 45 and 50. The erratic pattern could partly be due to time lag in confirmation of test results in new cases. Fitting a straight line to the trend by least squares suggests a slightly upward trend. We see from these analyses that there is no evidence from the data that the full national (level 5) lock-down reduced the daily growth or the incidence rates in COVID-19 infections in South Africa.

Figure 5: Trend in Incidence Rates in COVID19 During Full Lock-down



Source: Author’s model from official numbers

Speaking at a session with doctors from the South African Medical Association in Pretoria, the Health Minister announced that COVID-19 will affect 60-70% of South Africa’s population (Mlambo 2020). When this percentage is translated into numbers and using Udjo’s (2020) population estimates, it would mean that COVID-19 would affect between 34,502,235 and 40,252,609 South Africans. Using equations 7 and 8 above, a 95% confidence interval estimates, suggest that between 2.79% and 2.91% of South Africa’s population may have verified COVID-19 infection. This implies that the number of verified COVID-19 infections in South Africa could reach 1,614,383 (95% confidence interval: 1,611,604 – 1,676,901) by the end of 2020 if present trends continue, or higher if prevalence rates in the general population for a sustained period within the year exceeds 2.91% substantially. It should be noted that prevalence rates among those tested have been steadily increasing since after 71 days from the start of the pandemic in South Africa. For example, it was 3.52% as at 19 March 2020 i.e. by day 76 from the start of the pandemic in South Africa. At the time of adding this statement to this report, one is not sure whether this latest steady increase would go on for a sustained period.

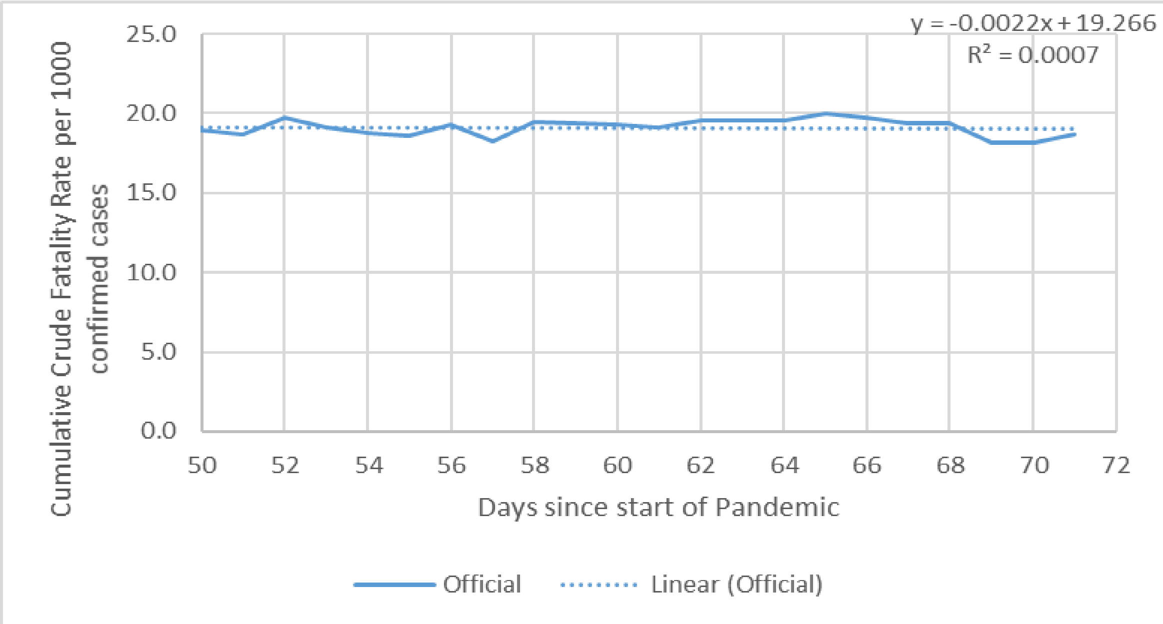
Projections of Case Fatality in COVID-19 in South Africa

The official figures of the number of fatalities due to COVID-19 (case fatality) probably understate the magnitude of COVID-19 case fatality in the population for at least two reasons as follows. As seen from Figure 1 above, there is disjunction between the official number of COVID-19 infections and what the model suggests. This disjuncture would also manifest in case fatalities. If the number of COVID-19 infections in the population are understated for the reasons mentioned earlier, the number of COVID-19 fatalities would also be understated. Additionally, several COVID-19 infections in the population would be asymptomatic and some of the cases that are symptomatic may not been captured in the testing efforts of the Government's programme. Therefore, in the event of death of some of these cases, they would not be captured in the official numbers of COVID-19 fatality and in the absence of autopsy the cause of death of some of these deaths would likely be mis-classified as death due to some other causes especially if there was co-morbidity. However, it is also possible that some non- COVID-19 deaths may be classified as COVID-19, but this is likely to be a rarer phenomenon than a COVID-19 death being classified as non-COVID-19. Let us now examine the projection results of COVID-19 case fatality.

Figure 6 shows the cumulative crude case fatality rates computed from the official numbers of COVID-19 deaths from day 50 to day 71 onwards as well as a linear fit (the dotted line) to the official cumulative figures. The starting point of day 50 was the point at which the cumulative rates had begun to stabilise and ranging between 17.97 and 20.1 per 1,000 confirmed cases. One of the advantages of cumulation is to remove erratic fluctuations (Brass 1971) due to small numbers and in this case, as well as time lags in confirmation COVID-19 fatalities while retaining biases that may be present in the data.

The linear fit projected suggests that fatalities due to COVID-19 could reach about 32,400 (95% confidence interval: 32,355 – 33,666) deaths in South Africa by the end of 2020 if present trends continue.

Figure 6: Cumulative Crude Case Fatality Rates from COVID-19 in South Africa



Source: Author's model from official numbers

DISCUSSION AND CONCLUSION

To mitigate the effect of COVID-19, the Government of South Africa like some other countries have done, announced a full national lock-down on 23 March 2020 with effect on 26 March 2020 later categorised as level 5 lock-down. This was downgraded to level 4 lock-down on 23 April 2020 with effect from 1 May 2020. Level 4 lock-down with some minor changes is like level 5 lock-down. The effect of lock-down on the progression of COVID-19 is unknown. Fitting mathematical/demographic models to the figures on COVID-19 progression released by the Ministry of Health, this study assessed the plausibility of the progression of the numbers of COVID-19 infections, the effect of lock-down and projected case fatality rates from COVID-19 in South Africa

The results suggest that official figures of COVID-19 under-state the number of COVID-19 infections in the general population. This study suggests that as at 71 days into the pandemic in South Africa i.e. as at 14 May 2020, there were probably 20,094 COVID-19 infections in the general population – 7,355 higher than the official figures (12,739) from the Ministry of Health for 14 May 2020. Quoting an example, official media releases usually read: "As of today, the total number of confirmed COVID-19 cases in South Africa is 11,350" (Ministry of Health: 12 May 2020). Since the figure reflects confirmed COVID-19 cases among those tested and not for South Africa as a whole, the statement should be qualified to read: As of today, the total number of confirmed COVID-19 cases *among those tested (emphasis mine)* in South Africa is 11,350.

It has been said in the public domain that the increasing number of infections during the full lock-down period is due to the increasing number of testing. A close look at equation (4) above for computing prevalence rate and substituting the terms: Prevalence rate among those tested = total number of COVID-19 cases among those tested/total number of persons tested for COVID-19 multiplied by a constant at any specified point in time in equation (4) reveals that this argument is illogical. The computation of the rate controls for variation in the denominator i.e. the total number of persons tested for COVID-19. Furthermore, if indeed the lock-down reduced infection rate, then the number of infections should stabilise resulting in decrease in the prevalence rates even with increasing number of testing. However, the empirical analysis by this author from the official figures reveals that during the full lock-down period, the number of infections did not stabilise but continued to increase, the prevalence rates of infections that occurred during the full lock-down period among those tested did not decline but rather ranged between 27 and 32 per thousand persons tested, and the daily growth rates of infections among those tested showed an upward trend. There was also a slightly upward trend in the rates of new infections that occurred during the full lock-down period.

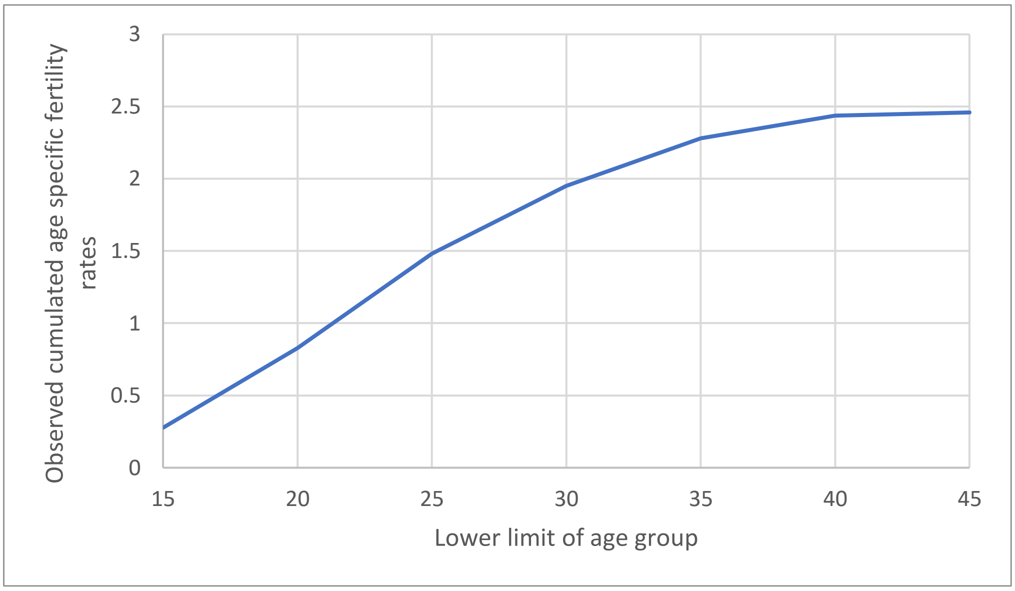
The models in this study were based entirely on official figures from the Ministry of Health. No assumptions from the experiences of other countries were incorporated into the models as countries differ in their population characteristics. Aside this, some of the across countries comparisons being made by some researchers and others reporting on the progression of COVID-19 are inappropriate, for example, the absolute numbers of COVID-19 confirmed cases and absolute numbers of fatalities from the virus in the United States, Russia, United Kingdom, Italy, etc compared with South Africa or other countries at different points in time. If the population size of all countries were the same at different points in time, then such comparisons (of absolute numbers) would be meaningful. But that is not the case. For example, the population of the United States is more than five times larger than the population of South Africa.

Screening of the population is not necessarily the same as testing of the population for COVID-19. That a substantial number of persons or proportion of the population has been screened according to official figures does not necessarily mean that a substantial number of persons or proportion of the population has been tested for COVID-19. Among the issues with screening are the following. It is highly likely that the numbers reported to have been screened contain double or multiple screening of the same persons either in a health facility or some other location. If for example, a person went to a clinic twice to see a GP on two different days, say within a week, this person would be part of the numbers collated each day that were sent to the Department of health as the reported numbers screened for COVID-19. Unless there is special effort in the data base used by the Department of Health to identify double of multiple screening of the same persons, the total national numbers of persons reported screened would be biased upwards. Additionally, a person may be infected with COVID-19 but pass the screening test because at the time of screening, the person was asymptomatic. From the above illustrations, while screening for COVID-19 is necessary, it is critically important that a large percentage of the population is tested for COVID-19. It should be noted therefore that as a limitation in this study, at the time of this study, the total number of persons tested for COVID-19 in South Africa was less than 1% of the estimated total population of South Africa. If these persons are characteristically different from the general population, this would constitute biases in the estimates provided. Furthermore, the results in absolute numbers projected assume that there would be no effective vaccine to protect against the virus in South Africa before the end of 2020. Should there be wide scale use of an effective vaccine against the virus in South Africa before the end of 2020, the numbers presented would be biased upwards.

Some researchers have purportedly predicted when the peak of COVID-19 would happen. This author is sceptical about such predictions. The upper asymptote of the COVID-19 may be assumed but it is another ball game to put a date when that upper asymptote would be reached in South Africa. One would only probably know when it has happened, not before. It is reasonable and logical to argue that the full lock-down in South Africa lock-down may have lowered the slope of the gradient in the daily numbers and growth of infections but convincing empirical evidence to prove this is lacking at this point. On the other hand, there is no evidence from this study that the national full lock-down lowered the slope of the gradient of COVID-19 progression in South Africa.

Flattening the curve have become buzz words in the era of COVID-19 but do the public and specialists have a common understanding of what flattening the curve means? I doubt it. Figure 7 provides an illustration. It shows the observed (i.e. unadjusted) cumulated age specific fertility rates from the 2016 South Africa Demographic Health Survey (SADHS). As can be seen from the graph, primarily because of biological factors, the curve is flat from age 40 onwards: fertility is exceptionally low after this age and rare after age 50 years.

Figure 7: Observed Cumulated Age Specific Fertility Rates 2016 SADHS: an illustration of time dependent flattening of a curve



Source: Author’s computation from the 2016 SADHS

Any epidemic will eventually run its course hence the first half of an epidemic curve resembles a sigmoid or logistic curve. When the epidemic reaches maturity, the curve becomes flat from the apex until the epidemic starts waning. These two examples are this author’s understanding of a curve becoming flat. However, this author has seen in the context of COVID-19 modelling where a report presented two overlapping separate curves in which the left curve has a narrower base and a higher peak (without containment policy) and the right curve has a broader base and a lower peak (with containment policy). From this perspective, the second curve is flattened by containment policy. From this author’s understanding of an epidemic curve flattening, there is no evidence from this study that the national full lock-down flattened the curve of COVID-19 progression in South Africa.

In the course of this study, two questions kept cropping up in my mind: (1) the focus of this study was the first 71 days of the pandemic in South Africa which included the full national lock-down period, what is going to happen to the progression of COVID-19 in South Africa as the country moves from one lock-down level to the other? (2) From a demographic perspective, what will be the long-term demographic effects of COVID-19 globally and in individual countries especially regarding population age structures, growth rates and net reproduction rates, and impact on the natural environment? Further research is required to address these and other questions.

ACKNOWLEDGMENTS

I wish to acknowledge the critical comments of the reviewers of this study: Dr Sandile Simelane (United Nations Population Fund, UNFPA); Dr Jabu Mkhize (Chair, Department of Economics, University of South Africa), and Professor Zurika Robinson (Department of Economics, University of South Africa). My profound gratitude to them for taking time to review the study. The views expressed in this paper however, are those of the author and not necessarily the views of UNFPA or the University of South Africa (UNISA), or the College of Economics and Management Sciences, UNISA or the Department of Economics, UNISA.

REFERENCES

Brass, W, 1971. Methods for estimating fertility and mortality from limited and defective data. Chapel Hill: The University of North Caroline, p. 5.

Mlambo S. 2020. Coronavirus will affect 60-70% of South Africans, warns Mkhize. IOL.

Ministry Health Republic of South Africa, 2020. Media Release, 12 May 2020.

News24, 2020, Coronavirus: All the fuss is irritating, says a doctor who diagnosed first SA case.

Udjo, EO, 2017, Can estimating completeness of death registration be used as evidence of inaccuracy of population size estimates from a census? The case of the 2011 South African population census. African Population Studies, 31 (1): 3133-3143.

Udjo, EO, 2020, South Africa's projected population size 2019-2020 by province, district and local municipalities disaggregated by age and population group, 2016 municipal boundaries.

WHO, 2019, Q & A on coronaviruses (Convid-19). Available on https://www.who.int/news-room/q-a-detail/q-a-coronaviruses.

WHO, 2020, Coronavirus disease (COVID-19) outbreak situation. Available on https://www.who.int/emergencies/diseases/novel-coronavirus-2019

Udjo, EO, 2017, Can estimating completeness of death registration be used as evidence of inaccuracy of population size estimates from a census? The case of the 2011 South African population census. African Population Studies, 31 (1): 3133-3143.