Prof T E Cloete (Vice Rector, Research, Innovation and Postgraduate studies, Stellenbosch University) and Prof Kanshu Rajaratnam (Director of the School for Data Science and Computational Thinking)

Data in the time of COVID-19

The solution to the coronavirus crisis is not merely medical but one that involves data. Data allows us to analyse the pandemic in different ways. This pandemic serves as an example of how data and data science can uncover trends and aid in making informed decisions. For example, Blue dot, a Canadian firm used data and artificial intelligence (AI) to identify a respiratory illness in Wuhan, China prior to World Health Organisation's global warning. Since its initial infection in Wuhan, the virus has spread globally, infecting over 6.3 million people and over 375,000 fatalities. The spread and subsequent activities have generated much Covid-19 related data. Methodological and technological advances mean that we have an opportunity to use this data to analyze the virus spread, to better understand its impact on health and economy, and most importantly, to prioritize policies and actions, from testing to lockdown strategies.

Cummulative total number of infections

The cumulative total number of positive cases is currently 34,357 in South Africa (1 June 2020). This number is the sum of daily active cases since the first infection and includes fatalities and those who have recovered. Apart from showing the geographical spread of the disease, this number is of limited value and could be alarmistic, as is the case currently, if viewed out of context.

A high variation in daily new cases has been observed, as a result of the variation in the number of daily tests conducted and the timing of the release of the test results. Furthermore, a large number of infected people are either asymptomatic or mildly symptomatic and hence, do not get tested. This implies an underestimation of the true rate (daily new numbers) of infection and the actual real total number of positive cases.

Testing efficiency

To determine testing efficiency, the relative daily confirmed cases (percentage of positive tests) as a percentage of total tests in a given period must be analysed. This number has averaged 5,8% for May 2020 nationally and remained relatively stable.

In April a community screening and testing (CST) program was implemented. CST was implemented differently in different provinces, varying from mass screening (including asymptomatic testing) to clusters of cases (hot-spots). This strategy could lead to misleading low percentage of positive cases, as was the case during the CST program (4,2% overall).

CST was discontinued on the 17th of May and a subsequent increase in the percentage of positive cases was noted because testing was more directed at people with Covid 19 Symptoms. An analysis of hot-spots (high infection rate) will inevitably show an increase in the percentage of positive cases. This has

been the case in general in the country and more specific in the Western Cape for example, where 1 out of every 10 (10%) people presenting symptoms tested positive, compared to 1 out of every 66 (1,5%) positive tests in Gauteng (Department of Health, 26 May). The implementation and phasing out of the CST program diferred from province to province. The testing efficiency is therefore merely a reflection of the testing program in a particular province and because these programs differ, makes a comparison amongst provinces impossible. Recently it was reported that the Head of Western Cape's Department of Health is implementing a new policy not to test Capetonians younger than 55 without other underlying health issues. This will change the database and skew the percentage of confirmed cases even more.

Fatalities as an indicator of the severity of the pandemic – age matters

Deaths are recorded accurately in South Africa and is a good indicator of the severity of the pandemic. Currently, total deaths attributed to COVID19 in the country is 755 (2 June 2020). The ratio of fatalities to total number of positive cases is 2.05%. Given the shortage of test kits and the backlog in reagents for testing and the resultant underestimation of the real total number of infections, the fatality rate in South Africa is probably lower than 2.05% of those that test positive. To date (2 June 2020) a total number of 761 534 tests were done and 755 deaths were reported representing a fatality rate of 0,09% of those tested (again a worse – case scenario).

Number of deaths per day
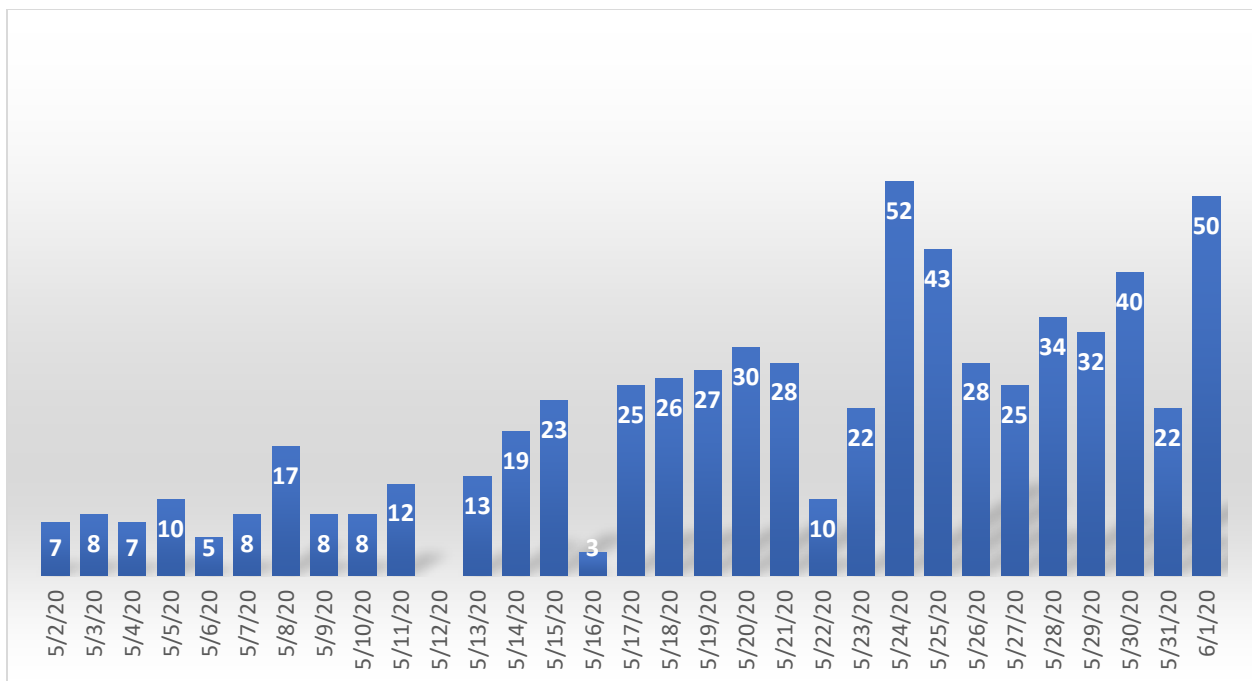
Mathematical models predict a cumulative number of deaths in South Africa of 40,000 by the end of November. This imples a sustained 241 deaths per day across the country over the next six months as compared to the average number 34 deaths per over the last seven days. Given the current 2.05% fatality rate, it would require approximately 11 500 new positive cases per day. Keeping an eye on the death rate is hence the most reliable indicator on how the pandemic will play out, especially in view of the limitations (backlogs, shortgae of reagents etc.) of testing programs.

Data internationally indicates low fatality in younger age groups, particularly for children. This is also the case in South Africa.

Total number of deaths per age group in South Africa (2 June 2020).

Age (years)	Total*	% of population	No of deaths	% of SA population
0 – 10	11 471 385	19,5	2	71,6
10 – 20	10 087 904	17,1	3	
20 -30	10 442 757	17,7	4	
30 - 40	10 109 138	17,1	42	
40 – 50	6 631 025	11,2	83	11,2
50 - 60	4 727 560	8,0	197	13,4
60 – 70	3 154 597	5,3	197	
70 – 80	1 547 686	2,6	143	3,6
80 – 90	602 696	1,0	66	
90 -100			18	
Total	58 774 749		755	

Data indicates fatalities for those under 20 years of age is extremely low (9 fatalities) in South Africa.

Given South Africa's young demographic advantage, we may experience lower hospitalization and fatality rates than countries with an older population. The country's young population (average age of 27 years and 71,6% of the population under the age of 40) is an advantage in this fight against the disease. A total of 41 fatalities have been reported in this age group since the onset of the pandemic in SA. Fatality rates increase significantly over the age of 50. Data for the Western Cape indicates that diabetes, hypertension and HIV were the most important co-mobidities.

Co-morbidities in the Western Cape per age group

Age group (years)	Diabetes (%)	HIV (%)	Hypertension (%)	Chronic kidney disease (%)
20 - 39	30	25	20	Unreported
40 - 49	51	33	28	Unreported
50 - 59	70	27	63	Unreported
60 - 69	58	8	74	Unreported
70 and above	46	0	58	46

Data above shows that while South Africa's population is skewed towards the younger population, the distribution of deaths is skewed towards older age group. A natural question to ask is whether this is a result of lower infection rates among younger age groups or whether younger patients recover at a higher rate or combination of both reasons. In order to make policy decisions that affect children and scholars, more data is required to determine the age distribution of infections rate (i.e. the number of tests done per age group and the number of positive tests per age group). This information is also crucial for more accurate modeling the pandemic. This data is important when deciding on the opening up of schools, institutions of higher learning and the economy.