Unevenness of geographical spread of COVID-19 at country and continental scales

The longitudinal variation of COVID-19 metrics in terms of cases, deaths, etc. is well documented. This is done in raw terms or in per capita terms. However, the spatial variation in the spread of the pandemic, when the globe itself or its continental regions are stratified according to geographical or geopolitical entities, is rarely reported. Here, using raw metrics (cumulative deaths and population) and per capita metrics (cumulative deaths per million of population), we quantify the diffusion of COVID-19 as a pandemic across the globe at continental scales and within continents at country scales. We find that Europe, North America and South America have reached relatively high evenness of spread, but in Asia and Africa there is still much larger concentration in a few areas.

If D is the number of cumulative deaths and P the population, then the cumulative indicator based on deaths per million of population, $N = D/P$, captures one component of the progress of the epidemic. It is size-independent and reflects the textbook-style sigmoidal shape of the early progress of the epidemic. Figure 1 is based on Our World in Data (OWID), where the COVID-19 data are updated daily and include information on confirmed cases, deaths and testing, as well as other variables of potential interest (https://github.com/owid/covid-19-data/tree/master/public/data).

We downloaded the cumulative death rates per million of population N, from 1 April 2020 to 31 December 2021, so that data are available for all countries and continental regions in the dataset. We choose 1 April 2020 as a reasonable date by which most countries in OWID had registered cases and deaths due to the new pandemic. Figure 1 shows the growth of the cumulative number of deaths per million of population, N, for the continental regions in real time, counted in days from 1 April 2020. For ease of interpretation, the cohorts are segregated into two sets by the world line (marked as black dashed lines), the first showing the three continents that are ahead of the world average and the other two that lag behind. Oceania has been left out because its contribution is miniscule, and the countries are few and often remote from each other. In real time, it is difficult to conclude that there were distinctly identifiable transitions from the first wave to the second wave. Only Europe shows near textbook-style behaviour – a prolonged levelling-off period (a plateau lasting almost a 100 days) between the first and second wave, followed by the second wave. For the other continental regions, this textbook feature is interrupted by lockdowns, other pandemic-mitigating protocols, and the emergence of new variants. The picture then becomes complex with multiple wave-like interferences. As of 31 December 2021, N takes the following values: Africa (165.93), Asia (269.19), world (680.36), Europe (2043.51), North America (2048.89) and South America (2744.96). There is a 17-fold difference between Africa and South America, as currently reported. This raises intriguing research questions: even after 641 days of counting, is this difference due to remoteness, or undercounting, or any other or all of these factors?

As there are no high-fidelity data of this quality available on earlier pandemics or epidemics, it is only possible to raise these as research questions. At present, no immediate recommendations can be made to contribute to public health action and aid control of COVID-19 spread or future epidemics of this scale. The population data are real time as recorded in OWID.

The D and N metrics related to COVID-19 for any geography will be dispersed over a wide range because of varying population sizes, densities and mobilities, population age structures, quality of healthcare facilities, per capita income, reproduction numbers, connectivities (by air and surface), etc. D and N are customarily plotted as functions of real time and this is how they are usually reported. The wave-like behaviour is discerned from these patterns. However, the spatial spread is rarely monitored. We will use data from OWID to observe what happens at continental scales globally and at country scales within continents using a Lorenz curve representation for the spatial spread. By spatial we mean how the epidemic spreads from one geographical location to another, and this depends on the size of the country, and the internal air and surface transport linkages.

The unevenness of spatial spread of N over all countries within a region at any point in time can be easily computed using a simple formula. If D_i is the number of deaths in country i with a population P_i in a cohort of M countries, then one can compute a second-order term $X = (\sum D_i)^2/\sum P_i$ and another second-order term $E = (\sum(D_i^2/P_i))$, where the summation is carried over $i = 1$ to M. Prathap[1,2] showed that the simplest measure of evenness or equality of distribution is given by $\eta = X/E$. Note that only simple mathematical operations are involved.

It will be useful to add a note on this measure of unevenness and the representation of the spread as a two-dimensional

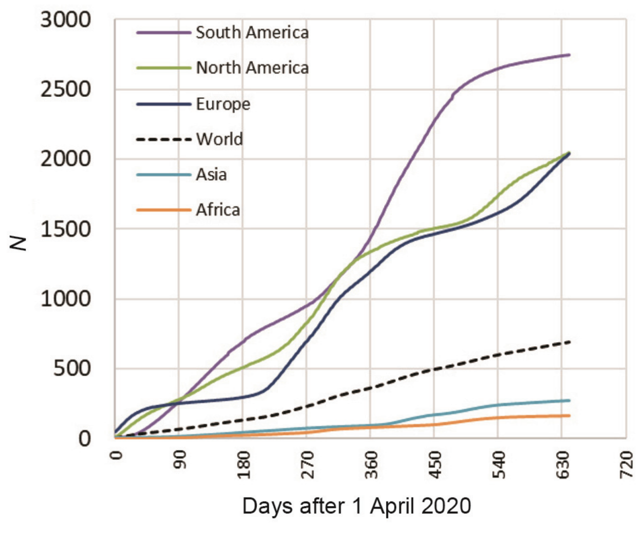

Figure 1. The longitudinal progress of cumulative deaths per million of population N for the continental regions.

Table 1. Unevenness of spatial spread of N over all countries within a region on 1 April 2020, 31 December 2020 and 31 December 2021

	N			Unevenness		
Region	1 April 2020	31 December 2020	31 December 2021	1 April 2020	31 December 2020	31 December 2021
Europe	47.31	728.37	2043.51	0.69	0.19	0.09
North America	11.68	859.42	2048.89	0.39	0.13	0.11
South America	1.32	962.93	2744.76	0.52	0.27	0.16
Asia	1.59	72.22	269.19	0.90	0.68	0.57
World	6.44	238.83	690.36	0.94	0.72	0.66
Africa	0.17	47.67	165.93	0.80	0.83	0.83

N, Cumulative deaths per million of population.

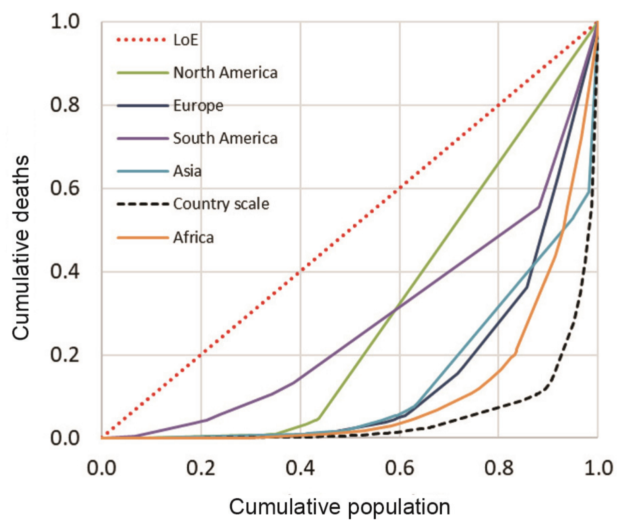

Figure 2. Unevenness of spatial spread of cumulative deaths per million of population N for the continental regions as of 1 April 2020.

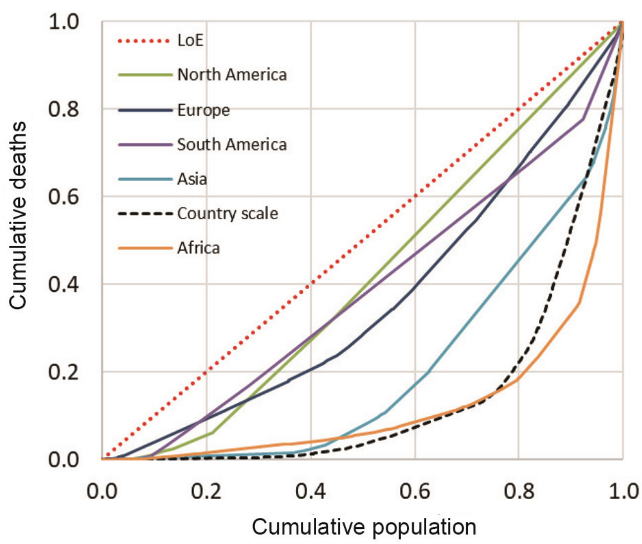

Figure 3. Unevenness of spatial spread of cumulative deaths per million of population N for the continental regions as of 31 December 2020.

chart. The Lorenz curve is used by econometricians to display such uneven distributions, e.g. income or wealth in a population. The Gini index is usually computed from the sequence $N_i = D_i/P_i$, reordered in monotonically increasing values from the lowest to the highest. The cumulative values of D_i and P_i are progressively added up, ranking the population in the same order and then plotted in the form of a graph (Figures 2–4). The areas under the curve are computed. Figures 2–4 are known as Lorenz curves and show snapshots of the unevenness of spatial spread of cumulative deaths per million of population N for the continental

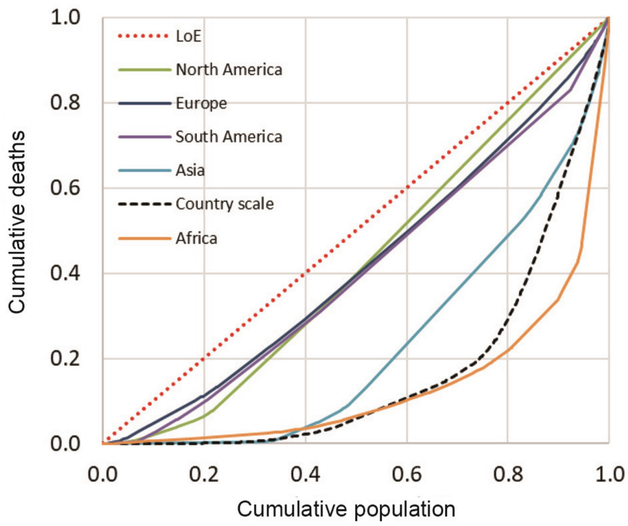

Figure 4. Unevenness of spatial spread of cumulative deaths per million of population *N* for the continental regions as of 31 December 2021.

regions as of 1 April 2020, 31 December 2020 and 31 December 2021 respectively. We chose 31 December 2020 as the cut-off for the end of the first wave and the beginning of the second wave. Table 1 compiles the unevenness of spatial spread of *N* over all countries using our formula within a continental region or the world as a whole on 1 April 2020, 31 December 2020 and 31 December 2021. The evenness index η varies from 0 (extreme inequality) to 1, where there is perfect equality or equity. If one defines unevenness index as $\text{II} = 1 - \eta$, then like the Gini index, II varies from 0 (perfect equality, the curve coincides with the line of equality (LoE)) to 1 (perfect inequality). We can see from Table 1 and Figures 2–4 that COVID-19 had spread much more uniformly in Europe and the Americas, and unevenly in Asia and Africa. This is both a reflection on the variation in different factors across nations in that cohort, and particularly of the poor connectivity to remote regions. Since there is no headline statistic available for connectivity, it is impossible to quantify this relationship. Another confounding factor is that of gross undercounting, but this is also difficult to quantify. We also see from Table 1 that with progress in time, the spatial spread has become more even in the continents with high values of *N*. This does not bode well for Africa or Asia, where it appears that there are vast swathes of the continent where COVID-19 has not swept through.

Currently, Europe, South America and North America show high degrees of homogeneity, with Europe and South America having changed rapidly from high concentration to greater diffusion of COVID-19 mortality during this period. If these are considered to be desirable levels where a pandemic becomes endemic (high seropositivity and high levels of herd immunity), then Asia and Africa have a long way to go. It is possible that in these two continental regions, there is not only poor connectivity which offers natural barriers to transmission, but also high levels of undercounting.

1. Prathap, G., *Scientometrics*, 2011, **87**(3), 515–524.
2. Prathap, G., *Scientometrics*, 2011, **91**(3), 997–1003.

ACKNOWLEDGEMENT. I thank my colleagues Ronald Rousseau, Ajit Haridas and P. Jafarali, and the anonymous reviewer for their critical comments on earlier drafts.

Received 4 January 2022; accepted 11 January 2022

GANGAN PRATHAP

A. P. J. Abdul Kalam Technological University,
Thiruvananthapuram 695 016, India and
Institute of Computer Science and Digital Innovation,
UCSI University, 56000 Cheras,
Kuala Lumpur, Malaysia
e-mail: gangan_prathap@hotmail.com