International Health 2021; **13**: 514–519
doi:10.1093/inthealth/ihab006 Advance Access publication 3 March 2021

Global epidemiology and socio-economic development correlates of the reproductive ratio of COVID-19

Lutz P. Breitling *

Augsburg University Hospital, Department of Gastroenterology, Rheumatology and Infectiology, Stenglinstr. 2, 86156 Augsburg, Germany

*Corresponding author: Tel: +0049-821-400-2351; Fax: +0049-821-400-3331; E-mail: lutz.breitling@uk-augsburg.de

Received 25 August 2020; revised 25 November 2020; editorial decision 26 January 2021; accepted 12 February 2021

Background: The most commonly cited argument for imposing or lifting various restrictions in the context of the coronavirus disease 2019 (COVID-19) pandemic is an assumed impact on the reproductive ratio of the pathogen. It has furthermore been suggested that less-developed countries are particularly affected by this pandemic. Empirical evidence for this is lacking.

Methods: Based on a dataset covering 170 countries, patterns of empirical 7-d reproductive ratios during the first months of the COVID-19 pandemic were analysed. Time trends and associations with socio-economic development indicators, such as gross domestic product per capita, physicians per population, extreme poverty prevalence and maternal mortality ratio, were analysed in mixed linear regression models using log-transformed reproductive ratios as the dependent variable.

Results: Reproductive ratios during the early phase of a pandemic exhibited high fluctuations and overall strong declines. Stable estimates were observed only several weeks into the pandemic, with a median reproductive ratio of 0.96 (interquartile range 0.72–1.34) 6 weeks into the analysis period. Unfavourable socio-economic indicators showed consistent associations with higher reproductive ratios, which were elevated by a factor of 1.29 (95% confidence interval 1.15 to 1.46), for example, in the countries in the highest compared with the lowest tertile of extreme poverty prevalence.

Conclusions: The COVID-19 pandemic has allowed for the first time description of the global patterns of reproductive ratios of a novel pathogen during pandemic spread. The present study reports the first quantitative empirical evidence that COVID-19 net transmissibility remains less controlled in socio-economically disadvantaged countries, even months into the pandemic. This needs to be addressed by the global scientific community as well as international politics.

Keywords: SARS-CoV-2, ecological study, health disparities.

Introduction

The basic reproduction number R_0 indicates the average number of individuals infected by each case of an infectious disease introduced into a fully susceptible population.[1] Once the transmission dynamics of a novel pathogen are understood in detail, reliable estimates of R_0 are particularly useful for theoretical analyses modelling the impact of control interventions.[2–4] On the other hand, the empirical reproductive ratio of new cases per time unit divided by cases during the preceding time unit (R_e) is a directly observable epidemiologic correlate of disease dynamics in the real world.[5] In the early course of a pandemic caused by a novel pathogen, little is known about issues such as the timing of infectiousness and generation of new cases. Thus only measures such as R_e are available to inform policymakers for decision making and justifying the introduction or lifting of interventions that may reduce disease spread but that may come at substantial social and economic costs.[2,3]

Governments worldwide have taken unprecedented measures to slow the spread of coronavirus disease 2019 (COVID-19).[3,6] It has been suggested that the pandemic might be particularly detrimental in less-developed countries.[7] However, empirical evidence on whether COVID-19 net reproduction is higher in disadvantaged countries is lacking.

As outlined in a pertinent World Health Organization (WHO) global report in 2012, the relationships between poverty and infectious disease risk are manifold and encompass factors such as reduced access to healthcare, lack of general and health education, malnutrition and environmental and living conditions that increase the risk of contracting disease.[8] A comprehensive literature review found household crowding, which is closely related with household size and lower income, contributes substantially to the burden of gastrointestinal and respiratory disease.[9] Gross domestic product per capita was found to be a useful indicator of infectious disease risk in a study focussing on economic downturns in Europe.[10]

With the pandemic spread of COVID-19 in 2019–2020, data for global analysis of the distribution and evolution of R_e estimates during a pandemic event have become available for the first time. The purpose of the present study was 2-fold, namely to explore the global pattern of observed R_e values during the first months of the COVID-19 pandemic and to analyse if COVID-19 transmission dynamics as measured by R_e are less controlled in socio-economically disadvantaged countries, which could result in a vicious cycle leading to yet greater inequality and hampered socio-economic development.

Methods

Data on confirmed COVID-19 cases by date and country were obtained from the COVID-19 Government Response Tracker project, which also compiles information on implemented control strategies such as physical distancing and contact tracing.[6,11] These data were combined with relevant country-level socio-economic, disease burden and development indicators obtained from a variety of recognized curated sources.[12–15]

In brief, the World Bank provides access to a large compilation of socio-economic country-level indicators but also incorporates health-related data. This encompasses data of the WHO Global Health Expenditure Database, which is based on a regularly conducted questionnaire study and provides estimates of health expenditures per capita[16]; data of the WHO Global Health Workforce Statistics, which estimates the number of physicians per population, based on a standardized national reporting framework[17]; data on maternal mortality from a joint study of multinational organizations[18]; data on fertility rates, combining information collected from the United Nations Population Division, the United Nations Statistical Division and national and international statistical offices; data on the mortality rate of children <5 y of age estimated by the Group for Child Mortality Estimation, a cooperation of international agencies (www.childmortality.org); standardized and comparable data on the prevalence of child malnutrition based on the joint child malnutrition estimates provided by the United Nations Children's Fund, the WHO and the World Bank. Household size data were estimated by the United Nations based on a compilation of censuses and household surveys. The country-level burden of disability-adjusted life years was derived by the Global Burden of Disease Study 2017, which estimated these measures based on a multitude of sources, for example, survey data, inpatient admission records and health insurance claims.[19] The World Bank Gini coefficient was calculated by the World Bank Development Research Group and provides an estimate of how income is distributed among the population of a country; a Gini coefficient of 0 implies perfect equality, whereas a Gini coefficient of 100 implies perfect inequality.

The academic Our World In Data (OWID) charity project provides another meta-database that allows access to tables combining topical data from national sources, such as the number of COVID-19 tests conducted by country and the date collected, with relevant data from the World Bank, WHO etc.[20] For the present study, the aforementioned COVID-19 testing data were downloaded from OWID along with World Bank data on gross domestic product per capita, population density, median age, proportion of people >65 y of age, life expectancy and prevalence of extreme poverty, that is, the proportion of the population living on <1.90 international dollars per day.

The empirical 7-d reproductive ratio for any day and country was calculated as the number of newly confirmed cases during the last week divided by the number during the week before. This 7-d R_e holds immediate appeal and is widely used because it levels out weekend–weekday differences in reporting.[5] More elaborate definitions of R_e are sometimes used that try to obtain more accurate R_e values by taking into account factors such as the distribution of reporting delays[5,21]; for the present global analysis, such data were not available.

The distribution of R_e was analysed using standard graphical approaches and calculating medians (interquartile ranges [IQRs]) for selected time points. Infinite or zero R_e were considered undefined and treated as missing values, as were 54 negative values that could result from retrospective revisions of national case counts. Detailed analyses were restricted to a 'robust' subset of the data, which was defined country-wise as the time period from first exceeding a total of 500 confirmed cases, including at least 100 cases during the last week ('robust' day 0), until first dropping below 50 cases during the last week. This was furthermore considered to imply mostly autochthonous transmission within each country.

To further investigate the development of R_e estimates over time, a mixed linear regression model predicting log-transformed R_e was developed, controlling for linear and cubic time trends and accounting for repeated measurements by including country as a random effect.[22] In this model, the $\log(R_e)$ was modelled as the sum of a country-specific random intercept, the time effect, the effect of additional predictors, such as tertile categories of the socio-economic indicators, and an error term. Since the R_e values within each country are not independent measurements, but are less correlated the more time has elapsed between two measurements, a continuous autocorrelated error structure was used to correctly estimate regression parameters and confidence intervals (CIs).[22] All models were fit using the robust time period as the time scale. To examine the role of country-level government interventions for R_e time trends, additional predictors were introduced in the model one at a time, indicating whether each intervention had been active 21 d before.

A total of 17 indicators covering socio-economic wealth, general and healthcare-related development and economic inequality were analysed. The association of each indicator with R_e was estimated by including a categorical predictor defined by its tertiles in the regression model. This approach was taken to allow for non-linear associations, while avoiding problems with too small numbers in each category. Based on exploratory results of the time trends, these models were restricted to the data beyond the 28th robust day, when R_e estimates within country appeared stable and there were no significant time trends left.

In sensitivity analyses, stricter robustness criteria (2000 confirmed cases, 250 in last week, no drop below 100 cases) were applied, the intervention time gap was varied from 28 to 0 d or the stable time period was defined differently. A significance level of 0.05 was used throughout. All analyses were done using R 3.3.3 (R Foundation for Statistical Computing, Vienna, Austria) and an extension package for mixed regression analysis.[22]

Results

Values of R_e could be calculated for 174 countries. As shown in Supplemental Figure S1, observed R_e values showed strong fluctuations, particularly during the first weeks of disease spread in many countries, with a tendency to excessively large values. Data from 140 countries with a total population >7 billion fulfilled the criteria for robust analyses. The median time from first confirmed case to reaching robustness was 35.5 d (IQR 26–55). The evolution of these more robust R_e values over time is shown in Figure 1. The spread of observed R_e narrowed substantially over time, with an IQR of 1.07–2.59 on the day 7 (median R_e 1.57) of the robust period and 0.72–1.34 on day 42 (median R_e 0.96).

The observed R_e overall tended to decline for about 2 months and then approached a lower asymptote (Figure 1). In the mixed linear regression model predicting log(R_e), the time trends were clearly significant and the 95% CIs around the estimated coefficients for day and day^2 clearly excluded the null effect ($\beta_{day} = -0.035$ [95% CI -0.039 to -0.031]; $\beta_{day^2} = 0.000\,24$ [95% CI 0.000 21 to 0.000 28]). These coefficients remained essentially unchanged when countrywide interventions were controlled for in the model (Supplemental Table S1). For those interventions with CIs excluding the null effect, the associated reduction of R_e was by a factor comparable to the change associated with advancing 2 d in the early robust period. Notably, the significance of the interventions themselves was strongly dependent on the choice of time lag (Supplemental Figure S2). When applying stricter robustness criteria, the time-trend coefficients were attenuated to $\beta_{day} = -0.026$ (95% CI -0.030 to -0.022) and $\beta_{day^2} = 0.000\,18$ (95% CI 0.000 15 to 0.000 22) with altogether unchanged patterns (Supplemental Table S1).

To facilitate the interpretation of the regression results for the socio-economic indicators, the regression coefficient estimating the log(R_e) difference between the countries in the highest and lowest tertile of each indicator was exponentiated. The exponentiated coefficient then indicates the ratio of the average R_e in the highest vs lowest tertile. For example, an exponentiated coefficient of 1.5 would mean that the average R_e in the highest tertile is 50% higher than in the lowest tertile (i.e. COVID-19 transmission is less controlled in the highest tertile). As shown in Table 1, the effect estimates of most development indicators analysed featured CIs excluding the null effect. The direction of every single association was consistent with a higher R_e in more disadvantaged countries, regardless of whether indicators relevant to health infrastructure, disease burden or other aspects of human well-being and development were considered. For example, the average R_e in countries in the highest tertile of average household size was 24% higher than in the countries in the lowest household size tertile. The strongest association was seen for extreme poverty prevalence, where the highest tertile was associated with a 29% higher average R_e. In contrast, for indicators such as gross domestic product or health expenditure per capita, the highest tertile was consistently associated with a lower average R_e (-13% and -15%, respectively, for these examples). These patterns were robust in sensitivity analyses (Supplemental Table S2).

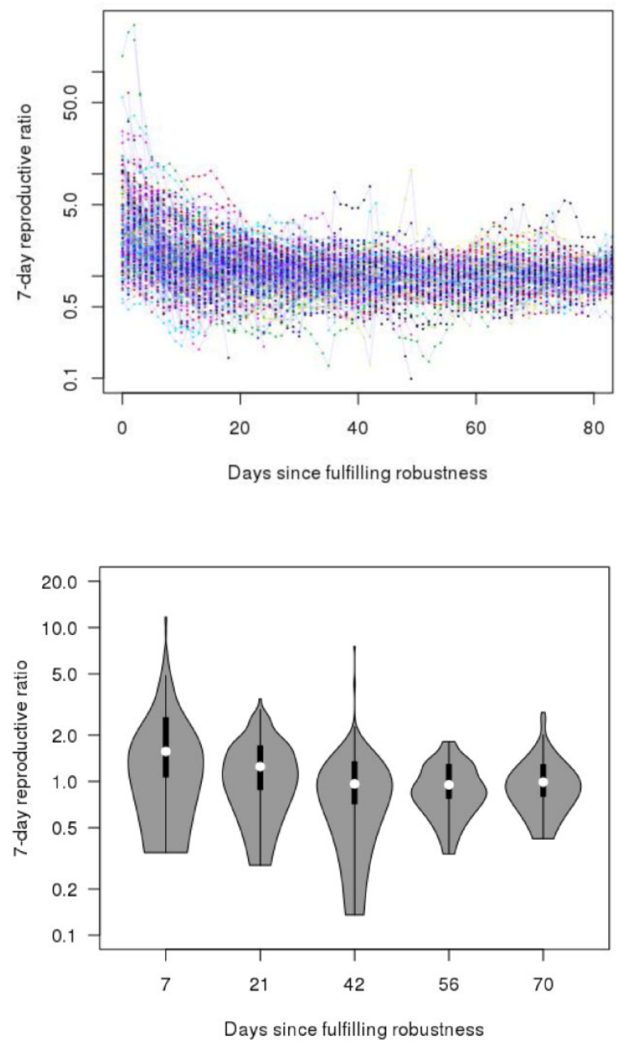

Figure 1. Distribution of the empirical COVID-19 reproductive ratio over time in 140 countries fulfilling criteria for robust analysis. The upper panel shows the data of each country connected by lines, whereas the bottom panel shows violin plots of the reproductive ratios at selected time points.

Discussion

The present data reveal that stable estimates of the reproductive ratio of a novel pathogen emerge only several weeks into pandemic spread. Months into the COVID-19 pandemic, transmission dynamics as estimated by the 7-d R_e remained less controlled in socio-economically disadvantaged countries, which is worrisome on numerous levels.

Table 1. Association of country-level socio-economic indicators with COVID-19 7-d empirical reproductive ratios (R_e). The dataset analysed included 6894 observations from 113 countries. Except for average household size (n=97), extreme poverty prevalence (n=80) and child malnutrition prevalence (n=68), data were available for >100 countries

Socio-economic indicator	Tertile cut-offs	R_e fold-change in highest vs lowest tertile (95% CI)
Gross domestic product per capita (dollars)	>21 546 vs ≤6367	0.87 (0.79 to 0.97)
Population density (people per km²)	>119 vs ≤48	0.96 (0.87 to 1.06)
Average household size, n	>4.5 vs ≤3.2	1.24 (1.12 to 1.37)
Median age (years)	>34 vs ≤25	0.85 (0.77 to 0.94)
People >65 y of age (%)	>10 vs ≤4	0.83 (0.76 to 0.92)
Life expectancy (years)	>76.9 vs ≤70.8	0.84 (0.76 to 0.93)
Health expenditures (dollars per capita)	>671 vs ≤115	0.85 (0.76 to 0.94)
Physicians per 1000 population, n	>2.5 vs ≤0.7	0.85 (0.76 to 0.93)
Disability-adjusted life years lost due to any (rate per 100 000)	>35 536 vs ≤27 396	1.06 (0.96 to 1.17)
Disability-adjusted life years lost due to communicable, maternal, neonatal and nutritional diseases (rate per 100 000)	>7818 vs <2272	1.18 (1.07 to 1.30)
Disability-adjusted life years lost due to non-communicable diseases (rate per 100 000)	>21 657 vs <16 841	0.85 0.77 to 0.94)
Total fertility rate (births per woman)	>2.8 vs <1.8	1.18 (1.07 to 1.29)
Extreme poverty prevalence (%)	>8.6 vs ≤1.0	1.29 (1.15 to 1.46)
Child malnutrition prevalence (%)	>29.2 vs <12.3	1.03 (0.91 to 1.18)
Mortality rate in children <5 y of age (per 1000 live births)	>30 vs ≤8	1.15 (1.04 to 1.28)
Maternal mortality ratio (per 100 000 live births)	>142 vs ≤19	1.13 (1.02 to 1.25)
World Bank Gini coefficient	>41.4 vs ≤33.8	1.21 (1.10 to 1.34)

The steep decline in observed R_e during the first weeks of the pandemic spread of COVID-19 probably cannot be explained entirely by altered transmission dynamics per se, given the minute intervention effect estimates compared with the overall time trends. In a recent interrupted time-series analysis of an earlier version of the Government Response Tracker dataset, the introduction of physical distancing reduced COVID-19 incidence by 13%.[6] As most interventions were introduced early during the pandemic (Supplemental Figure S3), when observed R_e showed a strongly negative correlation with time, time and intervention effects may be hard to disentangle in a reliable way. This is also supported by the smallest p-values being observed for implausibly short time lags in the present study, although immediate impacts on the epidemic curve have also been described by others.[23] Given the altogether small proportion of the population that presumably had experienced disease during the first few months of the pandemic,[7] the time trends also cannot be explained by a depletion of susceptible individuals due to acquired immunity. The consistent early R_e declines presumably result from a complex interplay of behavioural changes—partially caused by formal interventions—with increasing awareness, testing and reporting yielding more and more complete denominators.

Taken together, these findings urge caution when relying on observed R_e for policymaking in the early phase of a pandemic. When trying to overcome this issue by modelling and simulation, it is equally important to minimize potential biases in early estimates of transmissibility,[24] which in turn may lead to an overestimation when projecting case numbers or intervention effects.

With respect to the association of transmission dynamics with development indicators, the current results present the first empirical evidence that COVID-19 remains less controlled in the most disadvantaged countries even months into the pandemic. Uncontrolled spread in less-developed countries will be an ongoing source of new cases spilling over to other regions. More importantly, it will foster global inequalities, putting increasing strains on the least resourceful populations. This needs to be addressed on the level of world politics, and the present results may serve as a grim reminder that substantial proportions of the global population face more serious adversity than toilet paper missing in the supermarket.

It has long been recognized that poverty and infectious diseases are part of a vicious cycle and despite remaining a challenge in countries around the globe, COVID-19 may become yet another 'disease of poverty'.[8] The correlation of R_e with socio-economic development was very robust in the present study and seemed consistent for almost all indicators examined. Differences between the different indicators should not be overinterpreted and longitudinal microdata are needed to better understand the detailed causal relationships producing these patterns. Some additional remarks are nonetheless warranted. The rather general indicators of socio-economic prosperity, such as gross domestic product per capita, median age and life expectancy, consistently showed a positive association with disease control. On a structural level, this may reflect the general challenge of public health authorities needing sufficient resources to implement pertinent interventions,[25] and the R_e was also significantly lower in the countries with higher health expenditures and more physicians per population. On an individual level, economic disadvantages are associated with reduced compliance with shelter-in-place protocols, even in highly developed settings.[26]

Average household size is a development indicator that has a particularly close link to infectious disease transmission processes.[27] Household studies have been conducted for COVID-19 and found a substantially elevated risk of infection of household contacts.[28] The present results also showed that populations already experiencing a higher burden of poverty-related ill health struggle more with controlling COVID-19, i.e. the R_e remained higher in countries with more disability-adjusted life years lost due to communicable, maternal, neonatal and nutritional disease or higher mortality rates for children <5 y of age and maternal mortality ratios. Apart from average household size, the strongest associations with higher R_e were observed for more prevalent extreme poverty and for greater economic inequality as measured by the World Bank Gini coefficient. Interestingly, a similar association has recently been described for the state-level Gini coefficient with COVID-19 incidence and mortality in Brazil.[29] All these aspects of development, health and well-being are highly interrelated and the associations should not be interpreted as indicating causal relationships.

Limitations of this work include the ecological and observational nature of the data and a lack of information on face mask wearing and other hygiene recommendations. Whereas all data for this study were obtained from reputable organizations and curated databases, the heterogeneity of data sources presents another limitation.

Conclusions

Infectious diseases have always been a particular burden for socio-economically disadvantaged populations.[8] The results of the current work suggest that COVID-19 remained less controlled in countries with worse socio-economic development indicators, even months into the pandemic. There certainly is a danger that this might worsen global inequalities in the short as well as longer term. This needs to be addressed by policymakers and the international health community alike. Decisive support for disadvantaged populations in the context of incipient vaccination programs could be a starting point. In the long run, continuous efforts need to be maintained to reduce the burden of poverty-related disease by promoting public health structures and comprehensive socio-economic justice and well-being around the world.

Supplementary data

Supplementary data are available at *International Health* online (http://inthealth.oxfordjournals.org).

Author's contributions: LPB conceived the study questions, analysed the data and wrote the manuscript.

Acknowledgments: The great efforts by all subjects and organizations involved in compiling the public data bases used are gratefully acknowledged.

Funding: None.

Competing interests: None declared.

Data availability: All data used in the present work can be obtained from publicly accessible databases as detailed in the text.

References

1 Anderson RM, May RM. Infectious diseases of humans: dynamics and control. New York: Oxford University Press; 1991.
2 Hsiang S, Allen D, Annan-Phan S, et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. Nature. 2020;584(7820):262–7.
3 Flaxman S, Mishra S, Gandy A, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature. 2020;584(7820):257–61.
4 Britton T, Ball F, Trapman P. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. Science. 2020;369(6505):846–9.
5 an der Heiden M, Hamouda O. Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie in Deutschland – Nowcasting. Epid Bull. 2020;17:10–6.
6 Islam N, Sharp SJ, Chowell G, et al. Physical distancing interventions and incidence of coronavirus disease 2019: natural experiment in 149 countries. BMJ. 2020;370:m2743.
7 Walker PGT, Whittaker C, Watson OJ, et al. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. Science. 2020;369(6502):413–22.
8 World Health Organization. Global report for research on infectious diseases of poverty. Geneva: World health Organization; 2012.
9 Baker MG, McDonald A, Zhang J, et al. Infectious diseases attributable to household crowding in New Zealand: a systematic review and burden of disease estimate. Wellington: He Kainga Oranga/Housing and Health Research Programme, University of Otago, 2013.
10 Hunter PR, Colon-Gonzalez FJ, Brainard J, et al. Can economic indicators predict infectious disease spread? A cross-country panel analysis of 13 European countries. Scand J Public Health. 2020;48(4):351–61.
11 Hale T, Petherick A, Phillips T, et al. Coronavirus Government Response Tracker. Available from: www.bsg.ox.ac.uk/covidtracker [accessed 19 July 2020)]
12 Our World In Data. Coronavirus pandemic (COVID-19). Available from: https://ourworldindata.org/coronavirus.
13 Institute for Health Metrics and Evaluation. GBD Results Tool. Available from: http://ghdx.healthdata.org/gbd-results-tool.
14 World Bank. World Bank Open Data. Available from: https://data.worldbank.org.
15 United Nations, Department of Economic and Social Affairs, Population Division. Database on Household Size and Composition 2019. Available from: https://population.un.org/Household/index.html#/countries/840.
16 World Health Organization. Methodology for the update of the Global Health Expenditure Database, 2000–2017. Geneva: World Health Organization; 2019.

17 World Health Organization. National health workforce accounts: a handbook. Geneva: World Health Organization; 2017.

18 World Health Organization, United Nations Children's Fund, United Nations Population Fund, World Bank and the United Nations Population Division. Trends in maternal mortality 2000 to 2017. Available from: www.who.int/reproductivehealth/publications/maternal-mortality-2017/en/.

19 GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392(10159): 1789–858.

20 Hasell J, Mathieu E, Beltekian D, et al. A cross-country database of COVID-19 testing. Sci Data. 2020;7:345.

21 Lawless JF. Adjustments for reporting delays and the prediction of occurred but not reported events. Can J Stat. 1994;22(1):15–31.

22 Pinheiro JC, Bates DM. Mixed-effects models in S and S-plus. New York: Springer; 2004.

23 Pan A, Liu L, Wang C, et al. Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. JAMA. 2020;323(19):1915–23.

24 Kucharski AJ, Russell TW, Diamond C, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis. 2020;20(5):553–8.

25 Mooketsane KS, Phirinyane MB. Health governance in sub-Saharan Africa. Global Soc Policy. 2015;15(3):345–8.

26 Wright AL, Sonin K, Driscoll J, et al. Poverty and economic dislocation reduce compliance with COVID-19 shelter-in-place protocols. J Econ Behav Organ. 2020;180:544–54.

27 Becker NG, Dietz K. The effect of household distribution on transmission and control of highly infectious diseases. Math Biosci. 1995;127(2):207–19.

28 Lei H, Xu X, Xiao S, et al. Household transmission of COVID-19-a systematic review and meta-analysis. J Infect. 2020;81(6):979–97.

29 Demenech LM, Dumith SC, Vieira MECD, et al. Income inequality and risk of infection and death by COVID-19 in Brazil. Rev Bras Epidemiol. 2020;23:e200095.