medRxiv preprint doi: https://doi.org/10.1101/2020.09.26.20201814; this version posted October 12, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Evolution of COVID-19 cases in selected low- and middle-income countries: have they peaked due to high levels of infection and immunity?

Axel S. Lexmond[1]*, Carlijn J.A. Nouwen[2]*†, Othmane Fourtassi[2] and J. Paul Callan[2]*†

[1] Department of Engineering, University of Pretoria, South Africa
[2] Personal capacity

** These authors contributed equally to this work.*

† Corresponding authors.
Email: nouwencarlijn@gmail.com (C.J.A.N.); paul_callan@post.harvard.edu (J.P.C.)

25 September 2020; revised 1 October 2020 and 12 October 2020.

COVID-19 cases have peaked and declined rapidly in many low- and middle-income countries in recent months, in some cases after control measures were relaxed. For 10 such countries, the hypothesis that COVID-19 cases have declined mainly through low susceptibility levels, stemming largely from high levels of infection leading to (at least temporary) immunity, warrants serious consideration. The Reed-Frost model, perhaps the simplest description for the evolution of cases in an epidemic, with only a few constant parameters, fits the observed case data remarkably well, and yields parameter values that are reasonable. The model results give infection levels of 45% and 79%, above the herd immunity threshold for each country under their current social distancing conditions. Reproduction numbers range between 1.4 and 2.0, indicating that epidemic curves were “flattened” but not “suppressed”. Between 0.05% and 2.86% of cases have been detected according to the estimates – values which are consistent with findings from serological studies. Overall infection fatality ratios for two of three countries studied are lower than expected from reported infection fatality ratios by age (which are based on studies of several high-income countries). COVID-19 may have lower age-specific fatality risks in some countries, due to differences in immune-response, prior exposure to coronaviruses, disease characteristics or other factors. We find that the hypothesis of control through low susceptibility would not have fit the evolution of reported cases in several European countries, even just after the initial peaks; instead, these countries reduced COVID-19 cases initially through disease control measures – and subsequent resurgences of cases obviously prove that those countries have infection levels well below those required for herd immunity. Our hypothesis that the 10 countries we studied have low susceptibility levels should now be tested further through immunity studies, and efforts should continue to determine the duration and extent of immunity to SARS-CoV-2 after infection.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.26.20201814; this version posted October 12, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

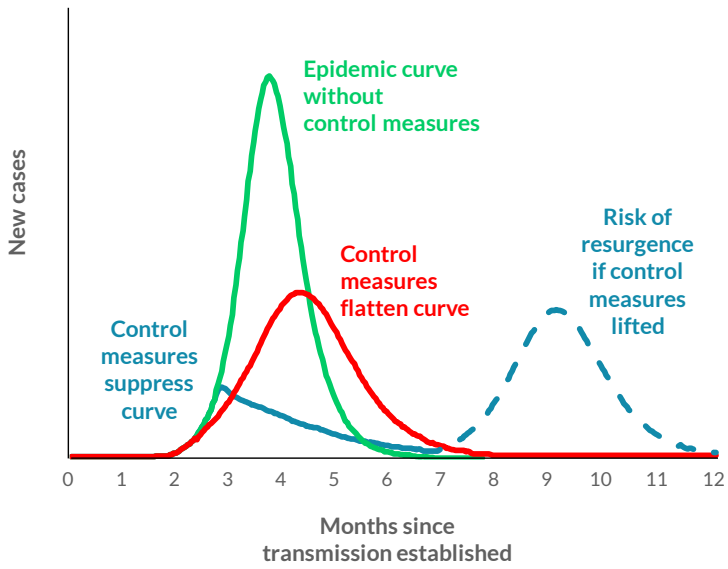
Introduction

Figure 1, based on a similar one produced by researchers at Imperial College London *[1]*, illustrates how case numbers are expected to evolve during an epidemic under different conditions. The green curve shows expected cases for an uncontrolled outbreak. This curve has three main features: initial exponential growth in new cases, followed by a single peak as the cumulative cases reach a level at which the remaining susceptible population is not large enough to sustain further growth, and an exponential decline in new cases. The curve shape is characterised by two numbers. First, the basic reproduction number, R_0, the average number of new infections caused by each current infected individual, at the outset of the disease outbreak when few people have been infected yet. Second, the mean generation time, t_g, the average time between infection of one person and when that person infects other people. The peak of the curve is reached as the proportion of the population that remains susceptible drops below $1/R_0$.

The red curve shows expected cases when governments and people take measures to control disease spread. The red curve displays largely the same features as the green – disease spread is only halted by high infection levels – but the curve is "flattened", with cases spread out more over time and fewer cases at the peak. In this case, disease control measures reduce the reproduction number to an effective basic reproduction number, R_{0_e} *[2]*. However, this number remains above 1, and the curve peaks when the proportion of the population that is susceptible reduces to $1/R_{0_e}$. Note that the "effective herd immunity threshold" (which is $1–1/R_{0_e}$) is not a fixed quantity: it depends on disease control measures and will increase when control measures are lifted; and only when the "full" or "natural" herd immunity threshold, $1–1/R_0$, is reached, through infection and/or vaccination, will the disease be constrained even in the absence of control measures.

The blue curve shows the evolution of cases when disease control measures are sufficient to bring R_{0_e} below 1. The curve is "crushed" or "suppressed"; the initial exponential growth in case numbers is halted and cases decline (almost always at a slower rate than for the green or red curves). In this case, because most people are still susceptible to the disease, it is possible for the disease to return if containment measures later allow R_{0_e} to increase above 1, as shown in the dashed part of the blue curve in Figure 1.

Figure 1. Typical disease outbreak curves in different scenarios [1].



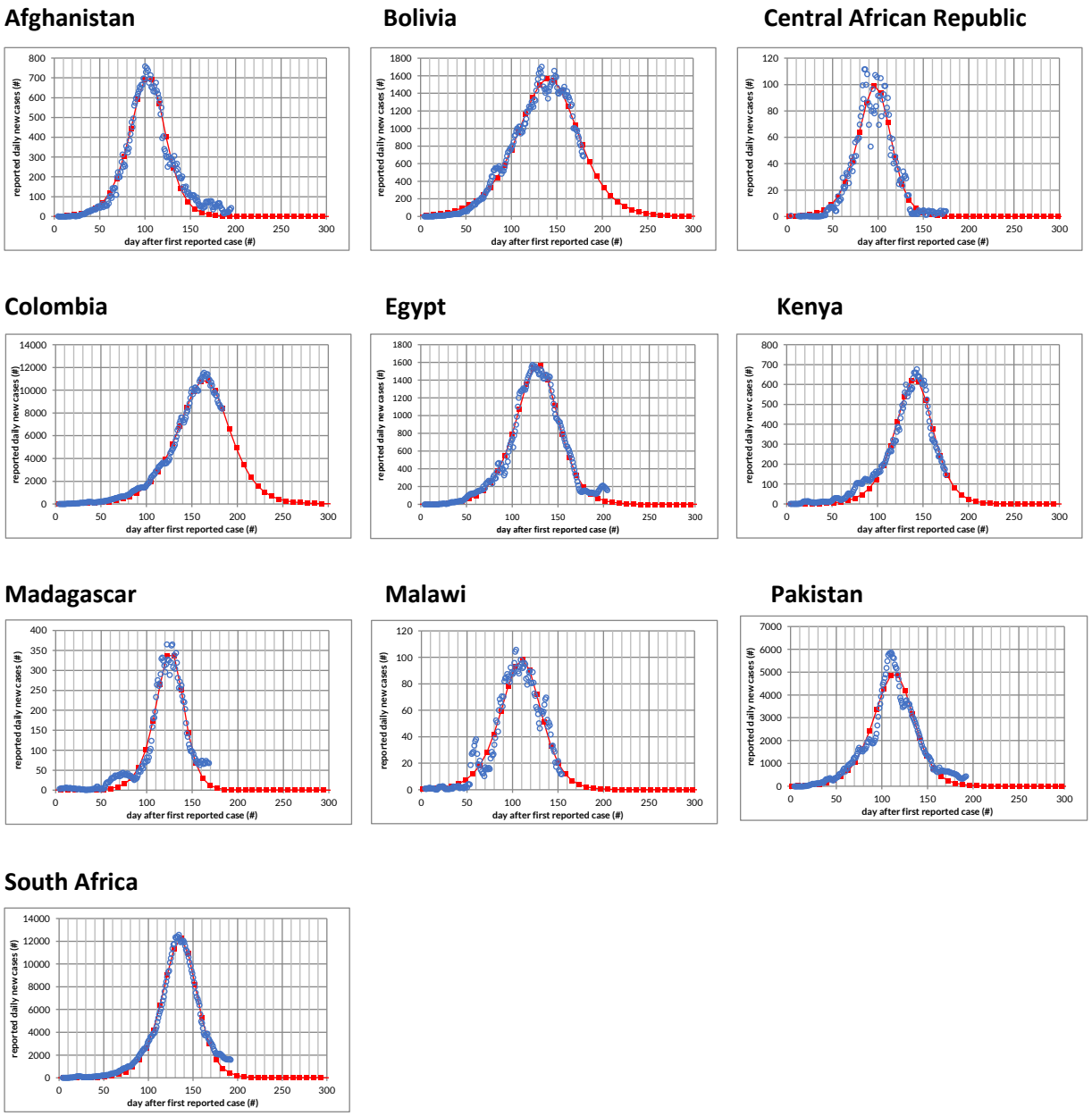
medRxiv preprint doi: https://doi.org/10.1101/2020.09.26.20201814; this version posted October 12, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Patterns in reported cases

Reported cases in several low- and middle-income countries (LMICs) have evolved in a manner that is very similar to the red and green curves of Figure 1. We show the 7-day rolling average of reported new cases for 10 such countries in Figure 2.

Figure 2. COVID-19 reported cases in 10 selected low- and middle-income countries, together with best-fit curves from simple disease spread model.

Blue circles show daily reported new cases of COVID-19 (based on a 7-day rolling average). Red squares (joined by straight lines) show best-fit outputs from linearized Reed-Frost model.



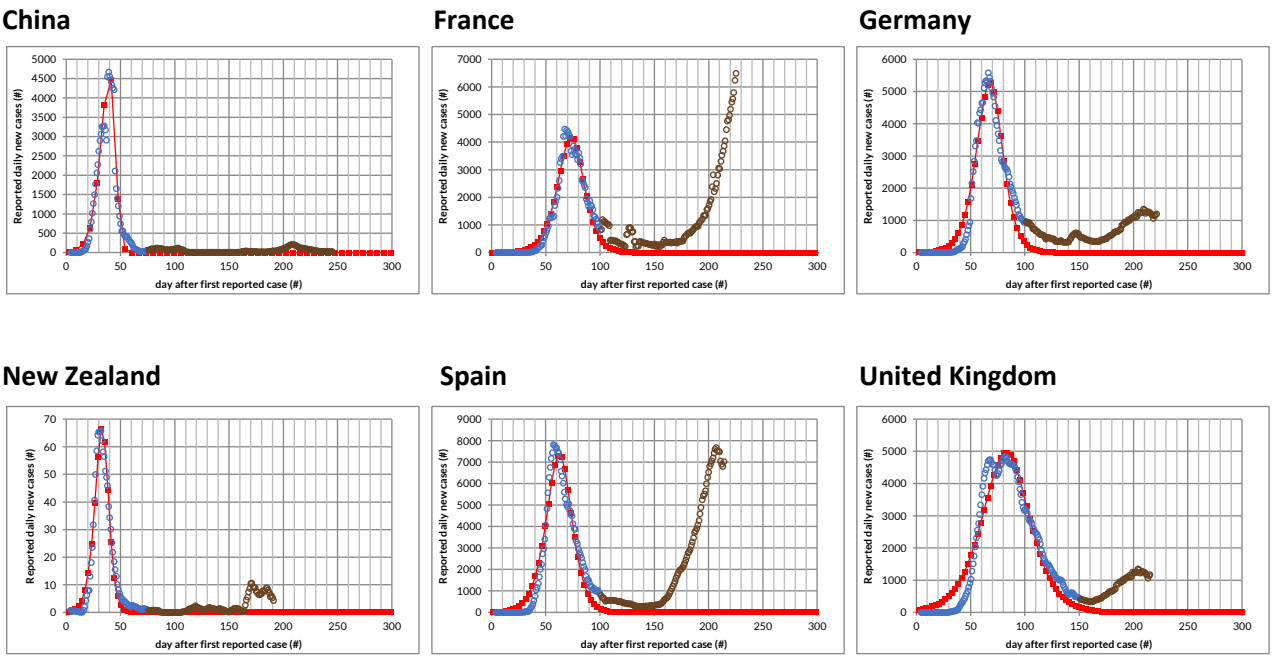
medRxiv preprint doi: https://doi.org/10.1101/2020.09.26.20201814; this version posted October 12, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Other countries show similar patterns; we have chosen to study a subset with the clearest similarities to expected outbreak curves like the red and green curves of Figure 1. In all these countries, the reported cases have (1) grown exponentially, (2) reached a single clear peak and (3) declined exponentially. Regulations were most stringent, and compliance was greatest, in most of these countries, after the designation of the global pandemic in March and have relaxed to varying degrees in recent months – but cases continued to decline. None of these countries has reported a significant increase in new cases after the peak that would indicate a second wave (although cases in some countries have only recently passed the peak). Together, these observations point to a hypothesis that the outbreaks in these countries have reached sufficiently low levels of remaining susceptibility, and that the recently observed declines in new cases are because many people are not susceptible– at least temporarily.

However, the numbers of cases in other countries – including most high-income countries (HICs) but also some LMICs – show patterns that are much different. Figure 3 shows 7-day rolling averages of reported cases for 6 such comparison countries. In these countries, cases have evolved in a manner that is similar to the first part of the blue curve of Figure 1. There have been peaks in numbers of reported cases, yet the decline is often longer and slower than the red or green curves would suggest. In some countries, there have been resurgences in cases, indicating that the initial suppression was not due to low levels of susceptibility.

Figure 3. COVID-19 reported cases in 6 selected comparison countries, together with best-fit curves from test of simple disease outbreak model.

Blue and brown circles show daily reported new cases of COVID-19 (based on a 7-day rolling average). Red squares (joined by straight lines) show best-fit outputs from linearized Reed-Frost model, fit only to the blue circles in the first peaks in cases.



medRxiv preprint doi: https://doi.org/10.1101/2020.09.26.20201814; this version posted October 12, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Fit with disease outbreak model and estimation of outbreak parameters

We test our hypothesis that the disease dynamics in these 10 countries has been driven primarily by susceptibility levels, by using a simple disease outbreak model and fitting to the reported cases. The outbreak model is a linearised Reed-Frost model *[3]*, the textbook deterministic mathematical model for an epidemic. The curves produced by this model depend on just two parameters, namely the effective basic reproduction number (R_{0_e}) and the mean generation time (t_g). Expected reported cases are calculated by scaling the Reed-Frost model's results by a detection rate (p) *[4]*. The parameters that produce the best-fit curve for reported cases in each country are determined partly analytically from the observed data and partly from least squares regression. Our model parameters (R_{0_e}, t_g and p) are constant over time – a beneficial assumption in that it avoids having too many free parameters, which might lead to good fits even if the model incorrectly describes the disease dynamics. Furthermore, in most of the countries studied, reported cases and reported deaths have followed similar trends (with changes in deaths lagging the corresponding changes in cases), even for countries with very low absolute numbers of reported cases and deaths – which suggests that the shapes of the curves likely reflect trends in actual cases and deaths, and that detection rates for both do not vary wildly over time *[5]*.

The best-fit curves are shown with red squares in Figure 2, for each of the 10 LMICs studied, and the corresponding parameters are presented in Table 1. For South Africa, R_{0_e} and t_g are calculated from the slopes and width of the observed data, and p is calculated from the sum of reported cases up to the peak divided by the total population times the expected proportion of population infected at the peak. For the other 9 countries, we use the value of t_g calculated for South Africa, and determine R_{0_e} and p from fitting to the observed case data. The fits are close, with R-squared goodness-of-fit measures between 0.94 and 1.00. (The fits are somewhat less good for Central African Republic and Malawi, due to very low numbers of reported cases, and for Pakistan, due to undulations in reported cases which may signal variations in R_{0_e} or p which the model assumes are constant.) These results demonstrate that the observed case patterns can indeed very accurately be described by an exponential outbreak halted by declining numbers of people still susceptible to infection.

Table 1 presents best estimates for R_{0_e}, p and the infection level (on 7 September), together with ranges of possible values (in parentheses). There are fairly wide ranges of reasonably possible values because the effects of R_{0_e} and t_g on the observed case curves are hard to distinguish – especially when the observed data has more "noise" or when the observed data does not include many points after the peak, and when values of R_{0_e} are close to 1 *[6]*. For South Africa, R_{0_e} has a best-estimate value of 1.74 (and a range of possible values between 1.45 and 1.90); the corresponding value of t_g is 7.8 days (4.8 days – 9.1 days) and of p is 1.41% (1.93% – 1.27%). For the other 9 countries, we determine best estimates of R_{0_e} and p by assuming that t_g = 7.8 days and fitting model outputs to data on reported cases. The value of t_g = 7.8 days is consistent with studies of serial intervals – the time between illness onset in successive cases in a transmission chain, whose mean value should equal the mean generation time t_g – by Ali *et al.* and others *[7]*, and close to the values used in other COVID-19 models *[8]*. We also conduct sensitivity analyses, determining ranges of possible values of R_{0_e} and p if t_g varies between 3 days and 11 days (wider than the range of possible values suggested in the literature *[7,8]*), and finding the values of R_{0_e} and p that produce the best fits to the reported cases for the extreme values of t_g. For all of the possible values of t_g, and corresponding values of R_{0_e} and p, the reported case curves generated do not change much, for each of the countries, and lead to the same implication that total infection levels in the countries have grown to the point where new cases are declining due to insufficient numbers of susceptible people (for the current value of R_{0_e}, i.e., under current disease control conditions).

The effective basic reproduction numbers, R_{0_e}, in Table 1 range between 1.4 (in Bolivia) and 2.0 (in Madagascar). Estimates for the basic reproduction number, R_0, the "natural" value in the absence of social distancing, for SAR-CoV-2 (the virus that causes COVID-19) in Wuhan at the outset of the global epidemic, range from 1.4 *[9]* up to 5.7 *[10]*. R_0 might be expected to be higher in low-income countries

medRxiv preprint doi: https://doi.org/10.1101/2020.09.26.20201814; this version posted October 12, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Table 1. Outbreak parameters for selected countries – effective basic reproduction number (R_{0_e}), mean generation time (t_g) and case detection rate (p) – and implied share of population infected as of 7 September.

The best estimates for the parameters correspond to the best-fit curves shown in Figure 2, and ranges of possible values for each parameter are given in parentheses (and as discussed in the text, are generated by letting t_g vary from 3 days to 11 days). The share of population infected up to 7 September is determined from the model outputs. The table also shows the values of the goodness-of-fit measure R-squared for the best-fit curves in Figure 2.

Country	**Effective basic reproduction number (R_{0_e})**	**Mean generation time (t_g)**	**Detection rate (*p*)**	**Share of population infected up to 7 Sept**	**R-squared for best-fit curve in Fig. 2** [perfect fit = 1]
Afghanistan	1.7 *(1.2–2.1)*	7.8 days *(3 days – 11 days)* [central estimate taken from calculated value for South Africa and supported by other research; range based in part on other research; see text for further discussion]	0.13% *(0.26%–0.12%)*	71% *(37%–82%)*	0.97
Bolivia	1.4 *(1.1–1.6)*		2.35% *(5.20%–1.90%)*	45% *(20%–57%)*	0.99
Central African Republic	1.8 *(1.3–2.2)*		0.13% *(0.25%–0.12%)*	74% *(39%–86%)*	0.94
Colombia	1.5 *(1.2–1.7)*		2.86% *(6.24%–2.35%)*	46% *(21%–58%)*	1.00
Egypt	1.6 *(1.2–1.9)*		0.15% *(0.30%–0.13%)*	64% *(32%–77%)*	0.98
Kenya	1.7 *(1.2–2.0)*		0.09% *(0.19%–0.08%)*	66% *(33%–77%)*	0.98
Madagascar	2.0 *(1.3–2.4)*		0.07% *(0.12%–0.06%)*	79% *(44%–89%)*	0.96
Malawi	1.6 *(1.2–1.9)*		0.05% *(0.09%–0.04%)*	64% *(32%–77%)*	0.94
Pakistan	1.6 *(1.2–2.0)*		0.20% *(0.38%–0.16%)*	66% *(34%–80%)*	0.95
South Africa	1.7 *(1.5–1.9)*	7.8 days *(4.8 –9.1 days)*	1.41% *(1.93%–1.27%)*	71% *(55%–77%)*	0.98

due to factors such as dense living conditions, lack of access to clean water and sanitation facilities, and inability of most people to work from home. Thus, our findings suggest that, for the 10 LMICs studied, social distancing measures and practices, perhaps in combination with higher levels of partial immunity from prior exposure to coronaviruses *[11,12]*, likely reduced the effective basic reproduction number and slowed the spread of the disease – but with R_{0_e} above 1, they did not "crush the curve".

Detection rates are estimated to be very low, ranging from 2.86% in Colombia to 0.05% in Malawi. These low detection rates explain how high levels of actual infections could be reached despite low numbers of reported cases, relative to total population, in all the countries studied. These low detection rates are not surprising. Serological testing results in Kenya, Pakistan and South Africa suggest that the number of people with coronavirus antibodies substantially exceeds the reported cases – by factors of up to 3,800 in Kenya (based on data in mid-May) *[13]*, of up to 540 in Pakistan (based on data from May to July) *[14]* and of up to 65 in South Africa (based on data from July to early August) *[15]*, which would correspond to detection rates of 0.03% for Kenya, 0.18% for Pakistan and 1.5% for South Africa. Furthermore, there is evidence that serological tests might underestimate effective immunity levels,

medRxiv preprint doi: https://doi.org/10.1101/2020.09.26.20201814; this version posted October 12, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

because people may have partial immunity due to T-cell responses even if infection by SARS-CoV-2 did not produce antibodies or if those antibodies subsequently waned *[11,12]*.

In all these countries, the analysis indicates that significant percentages of their populations have been infected and have become immune – at least temporarily. The best estimates of total infection levels on 7 September derived from the fitted curves range from 45% in Bolivia to 79% in Madagascar *[16]*. Note that the infection levels required to curb the outbreak (the herd immunity threshold) normally quote the percentage of population infected at the peak of the curve, but significant numbers of people continue to be infected after this point, even as the numbers of new cases decline.

The infection fatality ratio (IFR), or the percentage of deaths from COVID-19 among those infected with the SARS-CoV-2 virus, can be estimated for countries with reliable estimates of deaths. For Bolivia, Colombia and South Africa *[17]*, the IFRs calculated from reported deaths divided by the total number of infections derived from our analysis, are 0.15%, 0.10% and 0.04%, respectively. In these three countries, estimates have been made of excess deaths due to natural causes, and, if all of these excess deaths are due to COVID-19, the IFRs for the three countries could be up to 0.57%, 0.13% and 0.11%, respectively *[18]*. All three countries are expected to have a lower overall IFR, compared to European countries, because their populations have a higher share of young people, who are significantly less likely to die from COVID-19 if they contract the virus. Differences in population age profiles explain the estimated IFR for Bolivia, but not for Colombia or South Africa. If reported infection fatality ratios by age, based on data from several HICs *[19]*, were valid for these countries, the expected overall IFRs for Bolivia, Colombia and South Africa would be 0.57%, 0.63% and 0.33%, respectively. Possible explanations for why mortality risk for COVID-19 might be lower in Colombia or South Africa, compared to the (mainly) European countries from which IFRs by age are derived, could include differences in immune-system response (already observed, for example, between men and women in some HICs), partial immunity to COVID-19 due to prior exposure to other coronaviruses, differences in lethality and prevalence of different virus strains, and different infection levels for different age groups.

To test the robustness of our approach, we applied the same methodology to fit curves to the first peaks in the comparison countries shown in Figure 3 (represented as red squares in this figure). Researchers at Imperial College London and others argued convincingly in June that European countries have not reached herd immunity *[20]*, and subsequent increases in cases have proven their point. Applying our model to fit curves just to the first peaks (shown with blue circles in Figure 3) *[21]*, and hence assuming that the peaks were due to herd immunity, we find that the best-fit curves match the observed data for the first peaks fairly well in all cases, but with values for the disease parameters that are implausible. For example, for France, the best-fitting curve yields R_{0_e} = 1.4, t_g = 3 days (the lowest permitted value) and p = 0.39%. The detection rate is well below the detection rates of between 7% and 18% suggested by serological studies in European countries *[22]*. New Zealand is well-known for "crushing" the curve – and the parameters associated with the best-fit "herd immunity curve" to its reported cases would be R_{0_e} = 1.9, t_g = 3 days and a highly improbable p = 0.03%. Thus, our approach leads to a conclusion for the comparison countries that the evolution of reported cases was <u>not</u> due to herd immunity (but instead must have been due to control measures). With this check, we increase our confidence in the hypothesis that the outbreaks in the 10 LMICs studied <u>are</u> declining due to herd immunity, which generates well-fitting curves with plausible parameters.

Discussion

Prominent models of the epidemic from teams at Imperial College London (ICL) *[23]* and the University of Washington Institute for Health Metrics and Evaluation (IHME) *[24]* use SEIR simulations and determine key parameters – especially the effective reproduction number, which can vary over time – by fitting the models' results for deaths to the reported numbers of deaths from COVID-19 and utilising age-specific IFRs from recent studies (in HICs). These models estimate that total infections to date for

medRxiv preprint doi: https://doi.org/10.1101/2020.09.26.20201814; this version posted October 12, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

the 10 LMICs we studied are much greater than reported, but much smaller than our analysis suggests. For example, the models estimated total infection levels for South Africa of 7.7% (ICL) and 9.5% (IHME), and corresponding case detection rates of 13.9% and 11.2%, as of early September *[25]* – implying infection levels significantly below the levels of up to 40% suggested by serological study findings in Cape Town from July to early August *[15]*. It has previously been observed, by the IHME COVID-19 Model Comparison Team, that the predictive performance of seven COVID-19 models, including those of ICL and IHME, shows significantly higher errors for Sub-Saharan Africa, South Asia and Latin America and the Caribbean, compared to their performance for HICs *[26]*. Our research suggests why this might be the case. The ICL, IHME and other models are well-suited to HICs: reported deaths for such countries are likely to be reasonably close to actual deaths; the IFRs used in the models are based mainly on studies conducted in HICs; and it is clear from the evolution of reported cases that these countries have not reached herd immunity *[20]* and that their effective basic reproduction numbers have varied significantly over time as disease control measures have been introduced and adjusted (necessitating the additional granularity of SEIR modelling). However, for some LMICs: reported deaths from COVID-19 are likely to understate actual deaths by large factors *[18,26]*; age-specific IFRs might differ substantially from those in Europe and North America; and a simple model, using the approximation that effective basic reproduction numbers and detection rates remain constant over time, may be sufficient to describe the evolution of reported cases well (at least for the 10 countries we studied).

Systematic studies of representative population samples should be conducted in the 10 LMICs discussed here (and perhaps other countries as well) to determine the percentages of people who have been infected and are immune – and consequently test directly the primary conclusion that the overall infection levels are high and likely exceed effective herd immunity thresholds.

Even when susceptibility levels are sufficiently low that the virus can no longer spread exponentially, it will not be gone completely. Individuals may still contract the virus if they are not immune. Isolated communities, such as rural areas far from urban centres, may have much lower infection levels than the overall population, and may still experience localised outbreaks. New general outbreaks might happen as control measures are relaxed, and such outbreaks could be large for countries where R_{0_e} is currently close to 1 and/or for which cases have not mostly declined from the current peak – because relaxing disease control measures will increase R_{0_e} and could shift the effective herd immunity threshold to new values that are substantially greater than the current share of the population with immunity.

It is not yet known how long immunity from SARS-CoV-2 lasts. Even if a population is protected due to a high immunity level today, it is possible that this could be lost over time, which might lead to future outbreaks – the severity of which would depend on the share of people losing immunity, the amount of variation in timing of when people lose immunity, and whether susceptibility to reinfection is equal to the susceptibility to first infection. New strains of SARS-CoV-2 have emerged, and it is conceivable that future mutations could allow the virus to evade immune systems, and thus render previously immune populations susceptible again to the disease – but there is no evidence yet of any such immunity-evading strains.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.26.20201814; this version posted October 12, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

References and Notes

1. See https://www.imperial.ac.uk/news/195976/individual-response-covid-19-important-government-action/.

2. We introduce the effective basic reproduction number, R_{0_e} in this paper, because it is important to distinguish between changes in the reproduction number brought about by disease control measures and changes caused by increases in immunity levels in the population. The literature usually presents the effective reproduction number, R_e, defined as the actual average number of new infections caused by each current infected individual at any given time. R_e thus is affected by both disease control measures and immunity levels.

3. Rothman KJ. Epidemiology. Oxford University Press; 2012. See Chapter 6 of this book or other textbooks on epidemiology. Our linearized version of the Reed-Frost model uses the formula $n_{i+1} = n_i * R_{0_e} * s_i / n_{tot}$, where n_i is the number of infected people for the time period i, s_i is the number of susceptible people without immunity for the time period i, and n_{tot} is the total population. The mean generation time t_g is the duration of each time period in the model. Expected reported cases are equal to $p * n_i$ for each time step i in the model, where p is the case detection rate.

4. One more parameter is required in fitting the curve, namely the difference between the actual first case and the first reported case. This parameter can only shift the model curve to earlier or later times, without changing the shape of the curve and without impacting any of the other parameters.

5. In practice, one might expect R_{0_e}, t_g and p to vary due to introduction or relaxation of disease control measures, seasonal changes in behaviours, changes in testing protocols, and so on. However, such variations for the 10 selected countries appear not to be so great that they cause the output of the simple linearised Reed-Frost model to deviate from the observed case data. The evolution of reported cases and deaths in other LMICs, besides the 10 countries studied here, show features that are suggestive of dynamics determined mainly or partly by levels of infection and susceptibility over time – but the reported cases (as of 7 September) for many of these other LMICs did not always display simple, smooth, single peak curves, perhaps because the deviations from the key assumptions of the simple Reed-Frost model, including constancy of R_{0_e}, t_g and p over time, immunity being sustained for a substantial period of time after infection, and homogeneity in disease spread across the population, are sufficiently large to cause the real behaviour of the disease to deviate from the model outputs. It is in theory possible that variations in reproduction numbers, generation times and case detection rates – over time and across a population – could produce reported case curves that match those from the linearised Reed-Frost model, even though the actual cases were evolving in a different and more complex manner. However, such behaviour would require very specific changes in R_{0_e}, t_g and p, which seem unlikely to take place in practice: for example, the reductions in reported cases in South Africa from mid-July to early-September might be explained by a continuous reduction in case detection rate of more than a factor of 1,000 during that time, but this seems unlikely given that the total number of daily tests decreased by only a factor of about 3 over the same period.

6. For a given shape of the curve of disease cases over time – characterised by the exponential rates of increase before the peak and of decrease after the peak and the width of the curve – there is a unique combination of R_{0_e} and t_g which leads to the observed curve. However, for any given curve, there are a range of combinations of R_{0_e} and t_g which can produce very similar curves. (This issue is why it is very hard to determine the basic reproduction number, R_0, for any new disease, even when the case doubling time is well-known.) In the case of South Africa, the observed data is sufficiently "clean" that we are able to calculate R_{0_e} and t_g. For the other countries, the observed data is more "noisy" or does not have much data after the peak, and

medRxiv preprint doi: https://doi.org/10.1101/2020.09.26.20201814; this version posted October 12, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

there can be multiple combinations of R_{0_e} and t_g which produce reasonable fits to the observed data.

7. Ali ST, Wang L, Lau EHY, Xu X-K, Du Z, Wu Y, et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. Science, 2020;369(6507):1106-1109. https://doi.org/10.1126/science.abc9004. Griffin J, Collins AB, Hunt K, Casey M, McEvoy D, Byrne AW, et al. A rapid review of available evidence on the serial interval and generation time of COVID-19. (Preprint) https://doi.org/10.1101/2020.05.08.20095075. Ali et al. find, from data on transmission pairs in mainland China in January and February 2020, that the mean serial interval was 7.8 days early in the epidemic but shortened substantially to 2.6 days within a month. Griffin et al. found 22 estimates for the mean serial intervals between 3.1 and 7.5 days, and 3 estimates of the mean generation time between 3.9 and 5.2 days. Studies of serial intervals and mean generation time rely on symptomatic cases identified within health systems, and Ali et al. find that early case isolation decreases the mean serial interval. Therefore, it is reasonable to assume that the mean generation time, equal to the mean serial interval, for complete populations should be greater than those reported in the literature. The best estimate for t_g calculated from the reported cases curve in South Africa, namely 7.8 days, is the same value as the mean serial interval in the early days of the epidemic in China reported by Ali et al., and is a reasonable value to use for in our model for the other 9 LMICs that we consider. For the sensitivity analysis, we use a range of t_g from 3 to 11 days, which is wider than the full range of possible values of the mean generation time suggested in the literature to date.

8. For example, the Imperial College London COVID-19 Response Team's models assume a gamma-distributed serial interval with mean 6.48 days and standard deviation of 3.83 days. See https://mrc-ide.github.io/covid19-short-term-forecasts/index.html and Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. https://doi.org/10.25561/77482.

9. Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019-nCoV. Euro Surveill. 2020;25(4):pii=2000058. https://doi.org/10.2807/1560-7917.ES.2020.25.4.2000058.

10. Sanche S, Lin Y, Xu C, Romero-Severson E, Hengartner N, Ke R. High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. Emerg Infect Dis. 2020;26(7):1470-1477. https://dx.doi.org/10.3201/eid2607.200282.

11. Sekine T, Perez-Potti A, Rivera-Ballesteros O, Strålin K, Gorin J-B, Olsson A, et al. Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19. Cell 2020;183(1):158-168. https://doi.org/10.1016/j.cell.2020.08.017.

12. Braun J, Loyal L, Frentsch M, Wendisch D, Georg P, Kurth F, et al. SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. Nature 2020; published online (29 July 2020). https://doi.org/10.1038/s41586-020-2598-9.

13. KEMRI/Wellcome Trust. Preliminary Report of SARS-CoV-2 antibody prevalence among blood donors in Kenya. 28 June 2020.

14. Zaidi S, Rizwan F, Riaz Q, Siddiqui A, Khawaja S, Imam M, et al. Seroprevalence of anti-SARS-CoV-2 antibodies in residents of Karachi—challenges in acquiring herd immunity for COVID 19. J Public Health 2020; fdaa170. https://doi.org/10.1093/pubmed/fdaa170.

15. Press reports of a seroprevalence study of 2,700 public healthcare users (pregnant women and HIV patients) in Cape Town, conducted by Prof. Mary-Ann Davies of the Centre for Infectious

medRxiv preprint doi: https://doi.org/10.1101/2020.09.26.20201814; this version posted October 12, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Disease Epidemiology and Research at the University of Cape Town in July and early August, which found an overall seroprevalence level of 40% (Spotlight, Adele Baleta, 4 September 2020: https://www.spotlightnsp.co.za/2020/09/04/covid-19-high-prevalence-found-in-cape-town-antibody-study/).

16. All numbers apply to the portion of a country that has had some level of exposure. If specific physically concentrated groups in a country are fully shielded, the total proportion of the national population that has been infected to date, would be lower than the numbers listed here. For most countries, the numbers are as of 7 September; for some, they are as of 5 or 6 September.

17. The other countries studied have not reported excess mortality, and are likely to under-report deaths from COVID-19 significantly, and by differing proportions in each country. As a result, we limit our estimates of IFRs to the cases of Bolivia, Colombia and South Africa.

18. For all three countries, we use reported deaths as of 21 September. We estimate the IFRs expected if all excess deaths from natural causes are due to COVID-19, from the ratios of excess deaths to reported deaths as of the dates for which estimates of excess deaths are available. For Bolivia, as of 31 August, reported deaths due to COVID-19 were 5,027 and underreported deaths attributed to COVID-19 were 14,508, according to the Servicio de Registro Cívico (SERECI) (as reported in https://muywaso.com/14-mil-muertes-que-no-se-pueden-ocultar-y-una-tasa-de-subregistro-del-75/). For Colombia, from 18 May to 2 August, reported deaths due to COVID-19 were 12,393 and total excess deaths were 15,728, as reported by Minsalud (https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/VSP/estimacion-exceso-mortalidad-colombia-covid19.pdf). For South Africa, reported deaths due to COVID-19 were 15,499 as of 15 September, while excess deaths due to natural causes (which may not all be due to COVID-19 but are assumed to be for this calculation) between 6 May and 15 September are estimated to be 44,481 by the South African Medical Research Council (SAMRC) (https://www.samrc.ac.za/reports/report-weekly-deaths-south-africa?bc=254).

19. Levin AT, Hanage WP, Owusu-Boaitey N, Cochran KB, Walsh SP, Meyerowitz-Katz G. Assessing the Age Specificity of Infection Fatality Rates for COVID-19: Meta-Analysis & Public Policy Implications. (Preprint) https://doi.org/10.1101/2020.07.23.20160895.

20. Okell LC, Verity R, Watson OJ, Mishra S, Walker P, Whittaker C, et al. Have deaths from COVID-19 in Europe plateaued due to herd immunity? The Lancet 2020;395(10241):E110-E111. https://doi.org/10.1016/S0140-6736(20)31357-X.

21. The fits are made to only the first 100 days of reported cases for each of France, Germany and Spain, the first 150 days of reported cases for the United Kingdom, and the first 75 days of reported cases for China and New Zealand.

22. Calculated by comparing findings from seroprevalence studies in four countries with reported cases from around the times of the studies. The seroprevalence studies are as follow. England: Ward H, Atchison CJ, Whitaker M, Ainslie KEC, Elliott J, Okell LC, et al. Antibody prevalence for SARS-CoV-2 in England following first peak of the pandemic: REACT2 study in 100,000 adults. (Preprint) https://doi.org/10.1101/2020.08.12.20173690. Italy: National Institute of Statistics. Primi risultati dell'indagine di sieroprevalenza sul SARS-CoV-2. 2020. Portugal : National Institute of Health. Relatório de Apresentação dos Resultados Preliminares do Primeiro Inquérito Serológico Nacional COVID-19. 2020. Spain: Pastor-Barriuso R, Perez-Gomez B, Hernan MA, Perez-Olmeda M, Yotti R, Oteo J, et al. Infection fatality risk for SARS-CoV-2: a nationwide seroepidemiological study in the non-institutionalized population of Spain. (Preprint) https://doi.org/10.1101/2020.08.06.20169722.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.26.20201814; this version posted October 12, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

23. The ICL model is described at https://mrc-ide.github.io/global-lmic-reports/parameters.html. See also Walker PGT, Whittaker C, Watson OJ, Baguelin M, Winskill P, Hamlet A, et al. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. Science 2020;369(6502):413-422. https://doi.org/ 10.1126/science.abc0035.

24. The IHME model results are at https://covid19.healthdata.org/global.

25. Derived from the total reported cases divided by the total mean estimates of new infections from the models, up to 8 September, as reported on https://ourworldindata.org/covid-models.

26. Friedman J, Liu P, Gakidou E, IHME COVID-19 Model Comparison Team. Predictive performance of international COVID-19 mortality forecasting models. (Preprint) https://doi.org/10.1101/2020.07.13.20151233.

27. Watson OJ, Alhaffar M, Mehchy Z, Whittaker C, Akil Z, Imperial College COVID-19 Response Team, et al. Report 31: Estimating the burden of COVID-19 in Damascus, Syria: an analysis of novel data sources to infer mortality under-ascertainment. Imperial College London 2020. https://doi.org/10.25561/82443. This paper estimated that only 1.25% of deaths (range: 1%-3%) due to COVID-19 have been reported in Damascus, Syria.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.26.20201814; this version posted October 12, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Acknowledgements

We wish to thank Muhannad Alramlawi and Yohann Sequeira for help with parts of the modelling and research. We also acknowledge many useful discussions with Partners and consultants at Dalberg Advisors, in particular Edwin Macharia.

Funding: No funding was received to support this work.

Author contributions: A.S.L. conceived the first analysis of South Africa data, and J.P.C. conceived the idea to study a range of LMICs and comparator countries. A.S.L., C.J.A.N. and O.F. undertook the modelling. J.P.C. conducted the comparison of model parameters and assumptions with other research and with other epidemiological models. J.P.C. and C.J.A.N. prepared the manuscript.

Competing interests: C.J.A.N., O.F., J.P.C. work at Dalberg Advisors, a management consultancy whose clients include multilateral agencies, foundations, international development agencies, governments, companies and NGOs. C.J.A.N., O.F. and J.P.C. have prepared this article in a personal capacity, and the work was not funded by any client of Dalberg Advisors.

Data and materials availability: All data used in this study can be freely downloaded from the cited sources. The supplementary materials provide additional details on methods, including the equations used and the process for determining the parameters, and other notes and observations.

Supplementary Materials

A separate document provides supplementary materials, including methods and other notes.