

Philosophy *of* Medicine

Analysis: Philosophical
Perspectives on Covid-19

Reviewing the Reproduction Number R in Covid-19 Models

Maria Cristina Amoretti^{1*} and Elisabetta Lalumera²

¹ Department of Classics, Philosophy, and History, University of Genoa, Genoa, Italy.

² Department for Life Quality Studies, University of Bologna, Rimini, Italy.

*Correspondence: cristina.amoretti@unige.it

Abstract

Most of the epidemiological models of the Covid-19 pandemic contain the *reproduction number* (R) as a parameter. In this article we focus on some shortcomings regarding its role in driving health policies and political decisions. First, we summarize what R is and what it is used for. Second, we introduce a three-question matrix for the evaluation of any construct or parameter within a model. We then review the main literature about R to highlight some of its shortcomings and apply to them our three-question matrix. Finally, we argue that these shortcomings are important for an epistemic and political evaluation of R .

1. Introduction

National and local institutions mostly base their strategies and policies on epidemiological models of the pandemic. One of the common features of most of these models is that they contain R , the *reproduction number*, as one of their parameters. In some countries, especially at the beginning of the pandemic, and notably in media communication, the value of R has been given a key role. On the one hand, R has been implicitly associated with the dangerousness of the virus and the overall health situation (the greater R is, the more dangerous the virus and the overall health situation are); on the other hand, R has been presented as a number that can be simply “read off from nature” and can give us “scientific” and “objective” information about the spread of the virus. However, both assumptions are far from precise.

In this article we focus on some epistemic shortcomings of R regarding its role in driving health policies and political decisions. The points we raise are not new in the epidemiological literature, but we believe that they were generally unnoticed in many cases of the naïve use of models by journalists, content creators, and politicians. Also, we do not intend to suggest that other measures and parameters would fare better than R —in fact,



This work is published by [Pitt Open Library Publishing](#) and is licensed under a [Creative Commons Attribution 4.0 International License](#). © The Author(s).

they all have epistemic shortcomings (for a discussion, see Amoretti and Lalumera 2021a, 2021b). Rather, our broad aim is in line with that of other philosophers assessing Covid-19 science and especially epidemiology, such as Alex Broadbent and Benjamin Smart (2020), Jonathan Fuller (2020, 2021), and Saana Jukola and Stefano Canali (2021). We think that assessing the epistemic credentials of current scientific models, concepts, and practices from the point of view of the philosophy of science is not a way of downgrading the authority of science in the public discourse, but rather a useful methodological contribution that philosophers can make to science itself.

The structure of the article is as follows: first, we briefly summarize what R is and what it is used for in the current pandemic (section 2). Second, we introduce a three-question matrix for the evaluation of any construct or parameter within a scientific model and apply it to R . The matrix we propose involves three main elements: validity, practicality, and adequacy for purpose (section 3). We then review the main literature about R to highlight some of its shortcomings, and apply our three-question evaluation matrix to these findings (section 4). Finally, we argue that the apparent shortcomings are important for both an epistemic and a political evaluation of the epidemiological models using R for representing the current pandemic and its possible evolutions (section 5).

2. What Is R and What Is It Used For?

Originally, R was introduced in demography to measure the reproduction of people (Heesterbeek 2002), to estimate whether a certain population was growing or not. In epidemiology, R has a similar meaning as it measures the spread of an infectious disease in a population. It is a function of both the biological mechanism of transmission of the virus and of the rate of interaction between members of the targeted population. As R represents the *average* number of people each person with an infection goes on to infect, it is an indicator of the contagiousness or transmissibility of an infectious disease, not of the speed of the infection. R is typically reported as a single numeric value and its interpretation is often taken to be straightforward: an infectious disease is expected to keep on spreading if R has a value > 1 while it is expected to decrease and eventually end if R has a value < 1 (Anderson and May 1992). For instance, if R is 2, then, on average, an infected person will infect two others, who will infect four others, who will infect eight others, and so on.

R can have different variants. Specifically, R_0 —the *basic reproduction number*—assumes that everybody in a population is susceptible to infection. R_0 thus represents the average expected number of secondary cases that are generated by a single case in a completely susceptible population (Diekmann, Heesterbeek, and Metz 1990; Van den Driessche 2017). The assumption that the whole population is equally susceptible to infection is typically false. Even at the beginning of a pandemic when a new virus emerges, it is difficult to establish to what extent natural immunity, or immunity developed for other similar viruses, can offer protection from the new one. On the other hand, the moment when a new virus is identified can be much later than when it starts circulating and thus it is not easy to figure out what the time zero is—for example, Covid-19 has been found in blood samples taken as early as autumn 2019 (Althoff et al. 2021), while the first models forecasting its spread appeared in spring 2020 (see, for instance, Ferguson et al. 2020).

In contrast, R_t —the *effective reproduction number*—measures the average expected number of secondary cases generated by a single case in a population whose individuals may have gained some immunity (because they have recovered from the infection or have

been vaccinated), or where some control measures have been implemented (such as lockdown, social distancing, travel restrictions, and so on). R_t can thus be specified at a particular time and be used to trace changes in R as the number of immune individuals increases, or new restrictive measures are implemented (Nishiura and Chowell 2009; Mercer, Glass, and Becker 2011). During the Covid-19 pandemic, in many countries R_t has been one of the numbers that are reported daily and used to ground health policies and social measures.

Although both R_o and R_t are a “biological reality,” they are usually calculated by applying mathematical models to a complex set of different assumptions (for a review of methods for calculating R , see, for instance, Heffernan, Smith, and Wahl 2005). In principle, epidemiologists can calculate R using individual-level contact tracing; that is, tracing and testing all contacts of infectious individuals and then averaging over the number of secondary cases (this, in fact, would perfectly fit the definition of R). However, unless an entire country’s population is regularly tested, R cannot be measured directly but only retrospectively, considering population-level data using various parameters, such as the current and previous number of positives, hospitalizations, and deaths as a result of the infectious disease, the duration of contagiousness, the likelihood of infection per contact, and the effective contact rate (Dietz 1993; Lloyd-Smith et al. 2005). To describe more complex scenarios, additional parameters can be considered, such as the availability of public health resources, particular health policies, or containment measures. For instance, as both R_o and R_t vary with the social dynamics of a population, even an easily transmitted virus such as SARS-CoV-2 has trouble spreading when people hardly meet.

Considering these data and trends, suppositions must be made by the modeler to estimate what infection numbers could possibly explain them (Delameter et al. 2019). Thus, the estimated values of both R_o and R_t generated by population-level epidemiological models are constitutively dependent on numerous decisions made by the modelers (Artalejo and Lopez-Herrero 2013). Furthermore, mathematical methods, typically based on ordinary differential equations (ODEs), which have minor differences in structure and assumptions, might still deliver different values for R even if the same epidemiological data are used as inputs (Delameter et al. 2019). Jing Li, Daniel Blakeley, and Robert Smith (2011), for instance, consider a simple model for malaria and show that different methods for calculating R_o (specifically, the survival function, the Jacobian matrix, the constant term of the characteristic polynomial, the next-generation method, and the graph-theoretic method) can produce different results for R_o . More recently, Roya Nikbakht et al. (2019) ran a simulation based on data in Canada to evaluate six different methods for calculating R_o , proving that they can produce different results. Moreover, not only can individual- and population-level estimates “not lead to the same answer, as they depend on the efficiency of contact tracing, the use of test results (and their accuracy) vs. symptoms, etc.” (Pandit 2020, 1643) but different individual-level models having exactly the same expectations of the corresponding population-level variables (incidence and prevalence) can also produce different values for R (Breban, Vardavas, and Blower 2007).

Generally speaking, epidemiological models are currently used by national and local institutions as bases for their strategies—for example, the report commissioned by the European Union on the United States, France, Germany, Italy, Spain, the United Kingdom, and the United States (European Commission 2020). Prominent models from the beginning of the pandemic are, for instance, the Imperial College model, the Institute for Health

Metrics and Evaluation (IHME) model, and the London School of Hygiene & Tropical Medicine (LSHTM) model—and the Youyang Gu (YYG) model should be added for China. Most of these models use R .

The first group of models has the structure of susceptible-infected-recovered (SIR), susceptible-exposed-infectious-removed (SEIR) models, or compartmental models—they represent individuals in a population as moving through three/four distinct states (or compartments): susceptible to the disease; exposed but not yet infectious; infectious; and removed (recovered or deceased). To fill in these compartments and specify the likelihood of transitions between them, SIR and SEIR models may use data on new positives, hospitalizations, deaths as a result of the infectious disease, the duration of contagiousness, the likelihood of infection per contact, the effective contact rate, and so on. Typically, SIR and SEIR models first calculate R using ODEs and then use it to figure out the spread of a disease across a population.

A second group of models are “agent-based models,” which estimate how a disease spreads across a population by simulating various agents that interact in different social settings. This means that rather than applying the same rules to whole groups of individuals within different compartments, they explicitly model each single agent. To do so, they may use data on human mobility, activity surveys, mobile location, public transportation, census, and so on. Agent-based models can either calculate R by themselves or compute it separately; with respect to SEIR models, however, R is calculated per agent, not over whole populations.

Epidemiological models that use R may have the goal of testing the effect of mitigation strategies or other policies enforced by governments, as well as the effect of other factors, such as schools reopening or seasonal changes in infections and deaths (a classic example is the Wuhan study by An Pan et al. 2020). In such cases, R is a dependent variable for the study. Alternatively, epidemiological models employing R may have the goal of predicting attack rates, infection rates, or death rates for a time and population. In this case, R is the independent variable, and dependent variables may be new infections or deaths. In other words—to put it simply— R is either “generated” by a model when it is a dependent variable, or it is what the model bases its prediction on when it is an independent variable. In either case, epidemiological models are what make R significant for policies.

To better understand some possible interactions between model assumptions and R , let us briefly consider when R is used in an epidemiological model as an independent variable to predict, for example, attack rates, infection rates, or death rates. In a simple SIR model, R is usually calculated as the ratio between the disease transmission speed and the recovery speed. However, when the models are more complex, using age-structured populations, including demographic variables, considering public health interventions, and so on, it becomes far from obvious how to calculate R and how to interpret it. As Benjamin Ridenhour, Jessica Kowalik, and David Shay point out, focusing on R_o , “a model that includes age structure, population structure, and vaccination status could easily have more than 100 parameters. What does R_o represent in such a model? ... The salient point is that there are different methods [to calculate R_o], and that each method can potentially produce a different estimate of R_o ” (2014, 449). This implies that employing R with the goal of predicting attack rates, infection rates, or death rates for a time and population is dependent on the models and methods that have been chosen at the outset.

Some countries, at least at the beginning of the pandemic, either explicitly or implicitly based and evaluated their policies and political decisions on the values of R . A widely cited opinion article in the *Journal of the American Medical Association* recommended this strategy:

The US Centers for Disease Control and Prevention (CDC) should regularly report on the R_t for the US and for each of the 50 states so that political and public health leaders can gauge how well the combined organizational and individual social distancing measures in place around the country are working to diminish transmission of this virus. The CDC should then communicate this transparently to the public to increase public buy-in and understanding of the actions being taken to slow the spread of Covid-19. (Inglesby 2020)

Germany explicitly followed this kind of advice. On 16 April 2020, Angela Merkel famously explained in a video interview that R needed to be kept under 1.1, otherwise hospitals would be overwhelmed (Mahase 2021; Adam 2020; Hyde 2021). It has been argued that Merkel's explanation involving R can also be seen as a powerful narrative—a cognitive-rhetorical device—which helped the implementation of mitigation strategies in Germany at that time (Mintrom et al. 2021).

A further example is the case of Italy, where risk indicators for determining different scenarios and restrictive provisions for the twenty regions were defined in the ministerial decree of 30 April 2020. Among the 21 risk indicators, R_t —which is calculated based on the Istituto Superiore di Sanità's integrated surveillance—played a key role, as an R_t value > 1 was considered dangerous and led to four different scenarios and mitigation strategies. In the first scenario, a

regional R_t above the threshold [$= 1$ or > 1] for limited periods (less than 1 month) and low incidence ... [Second:] regional R_t values systematically and significantly between $R_t = 1$ and $R_t = 1.25$... [Third] regional R_t values systematically and significantly between $R_t = 1.25$ and $R_t = 1.5$... [Fourth] regional R_t values systematically and significantly higher than 1.5. (Ministry of Health et al. 2020, 18)

Notice that in the four scenarios above—as well as in the case of Germany—the values of R_t were implicitly associated with increasing values of dangerousness of the virus. There was therefore a small but significant conceptual step in the use of R : what the definition of the four scenarios implicitly showed is a new meaning of the parameter R_t ; namely, from reproduction rate of the virus to dangerousness of the virus. As mentioned earlier, the risk factors included parameters other than R_t but R_t was taken as measuring dangerousness of the virus.

3. Assessing Parameters: A Three-Question Matrix

Having introduced what R is and what it has been used for in epidemiological models during the pandemic, in this section we prepare the ground for the evaluative question: is R good enough for its job? Or, more modestly, is there any shortcoming that an epistemic reflection on R could bring to light? These questions are surely too vague to be tackled in their current form. Our strategy is therefore to present a three-question matrix for the evaluation of any

construct or parameter in a scientific model, and then to apply it to R . We do not see the matrix as particularly controversial, as it is built on various insights taken from the philosophy of measurement and psychometry (Tal 2020), and from discussions of prediction and explanation in scientific models (see, for example, Northcott 2017). The three elements of our proposed matrix are: validity, practicality, and adequacy for purpose.

3.1 Validity

Does the construct/parameter/measurement really measure or represent what it is intended to measure or represent? We might call this the question of validity, knowing that there are many senses of the term currently in use—our notion of validity skips over most of the subtleties of the debate, pointing to the idea that “validity is a concept like truth: It represents an ideal or desirable situation” (Borsboom, Mellenbergh, and Van Heerden 2004, 1063).

For example, it has long been debated whether IQ (intelligence quotient) measures what it is intended to measure; namely, general human intelligence. The parameter and the corresponding tests were initially introduced with the aim of predicting educational achievement of children but over time IQ scores tended to be popularized as measures of intelligence (Evans and Waites 1981). Putting aside the cultural and political issues that IQ also raises, objections have also been raised that general intelligence is a far more complex phenomenon than what IQ is able to track, when considered in everyday life (Richardson and Norgate 2014; Serpico 2018).

A slightly different example of a validity question can be raised about the ZIP code as a proxy of socioeconomic status (SES). Building on the assumption that some city areas and towns are more expensive to live in, social scientists and epidemiologists use ZIP codes as a proxy for socioeconomic status. It may then be concluded, for example, that ZIP codes correlate with physical activity, access to healthcare facilities, and various health parameters (Drewnowski, Rehm, and Solet 2007). Here, the relationship between ZIP codes and SES is not that of measuring but rather of representing, but the question can still be raised: are ZIP codes valid for representing SES? The answer is not obvious, as it might be objected that this is true in the United States but not for all world areas—in Europe, for example, old city centers may easily contain both gentrified areas and poor surroundings within the same ZIP code, as neighborhoods are less segregated. Thus, factors other than ZIP codes may be usefully integrated to represent SES, such as income, earning, occupational status, and so on (Ross and Mirowsky 2008).

A much less controversial case of validity in our sense is temperature as measured with a thermometer in a person’s armpit as a proxy for that person’s body temperature. The result can be more or less accurate (see next point), but it is generally agreed that it measures what it is intended to measure. Body temperature is then taken to be informative of many different diseases and health conditions.

3.2 Practicality

Can the construct/parameter/measurement be applied easily and adequately enough in its intended contexts of application? Is it reliable and accurate enough?

Suppose we are reasonably sure that a construct or parameter really measures what we intend to measure—that is, our validation theory is strong enough. The question remains whether the intended measure can be carried out easily and adequately in the contexts

where we need to gather our data. Our practicality question covers a range of different issues. By “easily” we mean that technological complexity, time, and cost are proportionate to the task. With “adequately” we refer to both accuracy and reliability. Accuracy represents the degree of closeness of measurements of a quantity to that quantity’s true value. Accurate measurements presuppose reliability, which is a further property of yielding converging results from different instances of measuring, and/or from different people completing the task.

For example, measuring body temperature with a thermometer under a person’s armpit is a fair compromise between ease and accuracy, and it is arguably very practical in our sense. Self-reported body weight is less accurate than hospital-scale weighing but it is also easier and less costly in terms of time and money, and hospital-scale weighing is not feasible in many contexts where researchers need to collect data about people’s weight.

3.3 Adequacy for Purpose

Is the construct/parameter/measurement what we need to reach the intended practical goal of the scientific model we are building?

This question zooms out from assessing the individual parameters to evaluating its fit with what a model is for. If our goal is to make the temperature of a dining room as pleasant as possible for most people, knowing the body temperature values of all individuals present in the room is not adequate for this purpose. More information is needed (such as subjective evaluations of the room condition and the room temperature) and individual body temperature provides information on a variety of conditions that are irrelevant to the thermal pleasantness of the room. Adequacy for purpose has been employed in the evaluation of parameters in climate change models (Parker 2009).

4. Reviewing the Literature for Problems with R

In this section we review some shortcomings of R that have been discussed in both scientific and philosophical literature and apply the three-question evaluation matrix to our findings.

(1) The value of R is notoriously difficult to measure properly. On the one hand, individual-level estimates (which best fit the definition of R) are difficult to obtain and may not correspond to population-level estimates. On the other hand, R not only depends on the biological features and mechanism of transmission of a virus, which may be partly unknown, at least at the beginning of an outbreak, but also on the social dynamic of a population and the rate of contact or interaction between its members, which may vary from region to region and time to time. Thus, assumptions and “educated guesses” must be made by the modelers. In Italy, for instance, R is calculated without considering positive and asymptomatic people, even if it is now recognized that the so-called silent spreaders play an important role in the transmission of SARS-CoV-2, and assuming that people whose infections have been contracted abroad and identified upon their subsequent return to Italy can transmit exactly like people who are infected in Italy, even if in certain periods quarantine has been compulsory for people coming from specific foreign countries (Istituto Superiore Sanità 2021). These assumptions may be not shared by other models. Moreover, the quality of data the models are built on may vary: they can be good or bad quality, scarce or sufficient, freshly collected or retrieved from registries and repositories filled out in the past; they may depend on nonepistemic factors, and so on (see points 5 and 6 below). The result is that the value of R may easily fluctuate as models vary in their complexity and data

in their quality. In line with what we proposed in section 3, this may be seen as a problem of validity, as the population-level estimates may not capture what the definition of R describes, and of practicality, as the values of R seem to lack both accuracy and reliability.

(2) As outlined in section 2, R represents the *average* number of susceptible people each infected person with a disease goes on to infect. Being about averages, R is not always adequate for understanding phenomena characterized by widely varying behaviors, such as the current Covid-19 pandemic. For example, let us imagine that Jeff Bezos or Elon Musk walks into a bar crowded with 100 regular people; if so, the average wealth in that bar suddenly increases, probably exceeding \$1 billion. However, if a philosophy professor walks into the same bar, the average wealth would not change that much. This example makes it easy to appreciate that the average is not a useful number to evaluate the distribution of wealth in such a bar, nor what is needed to change it. A similar thing happens with Covid-19: a few infected people, the so-called superspreaders, disproportionately cause most of the new infections, while most infected people infect only a few susceptible people. For instance, a recent study reported that “19% (15–24%) of cases were responsible for 80% of all SARS-CoV-2 transmission in Hong Kong ... while 69% (65–71%) of cases did not transmit to anyone” (Adam et al. 2020, 1718; see also Endo et al. 2020; Fang et al. 2021; Lau et al. 2020). This means that superspreaders are more infectious than an average individual, whose infectiousness is represented by R . In terms of our three-question matrix, this can represent a problem of validity, as it is dubious that R is truly able to measure the spread of an infectious disease in a population characterized by the presence of superspreaders.

(3) Another shortcoming linked to the fact that R represents an average across a certain area is that it does not adequately consider localized clusters and local variation, such as, for instance, the cluster of infections that characterized Bergamo in Lombardy, Italy, in March 2020. On the other hand, a high incidence of infection in a small spatially distinct subpopulation can affect the R value of a larger region, as happened in Germany, when about 2,000 people, the majority of who were linked to the Tönnies meat-processing factory, contracted Covid-19 in the Gütersloh area, thus making Germany’s national R value jump from just more than 1 to 2.88 in late June 2020 (Adam 2020). Even if it were possible to calculate R for subregions and smaller areas, however, its value would be less accurate, especially when absolute infections are low, as it would apply to smaller populations. This, again, is a problem of validity, as it is dubious that R is truly able to measure the spread of an infectious disease in a population where clusters are present, but also of practicality, as the value of R seems to be neither accurate nor reliable enough.

(4) When infection numbers are low, R can spike up and down for statistical reasons because it is an average number (Adams 2020). This means that when the number of existing and new cases is small, the value of R becomes questionable and useless in terms of providing information about an ongoing pandemic. For example, Carl Gwinnutt points out that

if the number of cases on day x is 100, and on day $x + 1$ is 150, this gives an R_o of 1.5, but only 50 new cases ... if on day x the number of cases is 50,000 and on day $x + 1$ is 60,000, not unreasonable looking at recent figures, the R_o value is 1.2 but there are 10,000 new cases. It would seem that when the R_o is used for small numbers of existing and new cases then its value is questionable. (2021, 26)

In terms of our three-question matrix, these fluctuations of R when the numbers of existing and new cases are low raise a problem of validity, as values of R in small populations cannot be reliably compared with those in larger ones, and of adequacy for purpose, as variations of R in small populations do not reflect variations of infectiousness that are significant for policies.

(5) As it is unlikely that an entire country's population is regularly tested, we have seen that typically R cannot be measured directly, but only retrospectively, considering different parameters, such as the current and previous number of positives, hospitalizations, and deaths; the duration of contagiousness; the likelihood of infection per contact; and the effective contact rate. This means that R does not capture the current status of an epidemic. The consequences of such time lag can be minimized by mathematical tools, which, however, cannot take into account other relevant considerations, such as the extra time needed to register deaths as a result of Covid-19 (Adam 2020). This possible delay creates a problem of adequacy for purpose, since real-time decisions are what one is trying to attain. Above all, however, practicality is the main issue here, as R is not easy to measure, and its value sums up all the problems of accuracy and reliability of the other measures it is inferred from.

(6) Among the data used to retrospectively calculate R are mortality data. However, in certain countries deaths resulting from Covid-19 are poorly registered, as possibly happened on the Indian subcontinent (Biswas, Afiaz, and Huq 2020; Pulla 2020). Moreover, mortality data are themselves not transparent, as a statement on a death certificate that identifies Covid-19 as the underlying cause of death is not a purely descriptive predicate, as “it is grounded on both factual (causal chains and the patient's medical conditions before and at the time of death) and non-factual reasons (the importance of prevention and the epidemiological clause exception)” (Amoretti and Lalumera 2021a, 3–4). As accuracy and reliability are involved here, this can be seen as a problem of practicality.

(7) A final shortcoming can be related to the general idea that a value of $R > 1$ points to an epidemic. A common understanding of R is that “if R_o is greater than one then the outbreak will lead to an epidemic, and if R_o is less than one then the outbreak will become extinct” (Anderson and May 1992, 61), thus assuming that R_o is a threshold parameter able to determine whether an outbreak will cease or persist, and how much control effort is needed to eliminate the infection from a population. Thus, R_o has been widely understood as a measure of disease strength and dangerousness and used to estimate the efficiency of control measures and health policies.

However, it has been argued that R_o is not always an epidemic threshold parameter because there are some diseases that can persist with $R_o < 1$, while diseases with $R_o > 1$ can die out. This can be the result of the presence of backward bifurcation, where a stable endemic equilibrium coexists with a stable disease-free equilibrium when $R_o < 1$. In this case, making $R_o < 1$ is a necessary—but not sufficient—condition for effectively controlling the spread of a disease outbreak (Li, Blakeley, and Smith 2011; for a discussion related to Covid-19, see also Nadim and Chattopadhyay 2020). Moreover, according to Nicolas Bacaër and M. Gabriela M. Gomes (2009), when the infectious disease is affected by seasonality, relatively large epidemics can occur even when $R_o < 1$ and the final epidemic size may not be an increasing function of R_o .

More importantly, the threshold value calculated by epidemiologists using population-level models typically based on ODEs, which describe the dynamics of the expected population size in different disease stages, does not always correspond to the average number of secondary infections; that is, to R_o , obtained from contact tracing and individual-level models. This means that not only have many threshold values been denoted by R_o , even if they do not really refer to the average number of secondary infections (Breban, Vardavas, and Blower 2007; Li, Blakeley, and Smith 2011), but also that

R_o may not be an accurate measure of the severity of an outbreak since R_o may fail to represent an epidemic threshold parameter. Thus, measuring R_o through contact tracing ... may not help in predicting the severity of the outbreak and may not be a useful measure for determining the strength of the necessary control interventions. Only an epidemic threshold parameter can be used to design control strategies. (Breban, Vardavas, and Blower 2007, 282)

Such a threshold parameter is different to R_o . This may be especially evident when populations are not homogeneous (Roberts and Heesterbeek 2003; Heesterbeek and Roberts 2007). This can be seen as a problem of adequacy for purpose, as far as R is used as a threshold parameter for a population-level model to make decisions about health policies and the strength of control interventions.

5. Conclusions

We have seen that R has several shortcomings, which can be differently related to our three-question evaluation matrix.

Validity: Does R really measure what it is intended to measure—the spread of Covid-19 in a particular population? Being about averages, R can be insufficient to properly measure a phenomenon such as the current pandemic, which is characterized by widely varying behaviors, such as those determined by superspreaders, localized clusters, and local variation. Moreover, since when infection numbers are low R can spike up and down, the value of R becomes questionable as a measure of the spread of Covid-19. Finally, some ways to measure R may contrast with its own definition. This means that R does not really capture what is going on with the current Covid-19 pandemic.

Practicality: Can R be applied easily and adequately enough in its intended contexts of application? Is R reliable and accurate enough? The success of R is probably because of the simplicity of its definition and the ease with which it can be applied to the current situation. Adequacy is less evident, though. We have seen that the value of R is difficult to properly measure and may fluctuate since mathematical models used to calculate it vary in their complexity and data in their quality. Moreover, the value of R may not be accurate enough when absolute infections are low, as in small populations. Reliability and accuracy can also be threatened by bad quality data, and the fact that some data may not be transparent, being determined by the interrelation of facts and values. Even if R were able to capture what is going on in the current pandemic, it can hardly be measured accurately and reliably enough.

Adequacy for purpose: Is R what we need to reach the intended practical goal of the scientific model we are building? We have seen that when infection numbers are low, R is unable to give relevant information about an ongoing pandemic. Moreover, as R cannot be measured directly, but only retrospectively, it can hardly capture the current status of an

epidemic and used to make real-time decisions, which is what an epidemiological model is currently used for by politicians and decision-makers. Finally, if the aim is predicting the severity of an outbreak and thus determining the strength of the necessary control interventions, R is not ideal for guiding political choices. So, even if R were able to capture what is going on and were accurately and reliably measurable, it remains questionable that R is the best tool to gain relevant information about the current pandemic, make real-time decisions, and determine successful control interventions.

Of course, all these problems and shortcomings are well known to epidemiologists—and, in fact, the epidemiological literature we have reviewed here clearly highlights and addresses them in various interesting ways. Still, they are under-discussed in the philosophical debate and often ignored in the public sphere, where R is generally used by politicians and media communicators in a very naïve, uncritical, and epidemiologically uninformed way, with the main desire to have a single “magic number” that is able to “really” define the development of the pandemic in a “scientific” and “objective” way.

Given the several shortcomings of R , our review can contribute to the argument that R is not the best candidate to be used by politicians and decision-makers to determine public health responses or national and local policies during the current pandemic. The attention on R , which has been strengthened by the media and social communicators, may obscure other relevant data and epidemiological data, such as fatality rates, transmission rates, infectious periods, and latent periods, the number of infected people, new infections, deaths, hospital admissions, and intensive care patients, as well as their disaggregation by age, sex, comorbidities, and so on.

On the one hand, this does not mean that any specific kind of information, or some other epidemiological rate different from R , would be able to mirror exactly the “reality” of disease development, or to magically inform health policies and mitigation strategies. Let us consider, for example, infection fatality rate (IFR), which estimates the proportion of deaths among all infected individuals, or case fatality rate (CFR), which estimates the proportion of deaths among identified confirmed cases (for a further philosophical discussion on these parameters, see, for instance, Amoretti and Lalumera 2021b). Even if they may seem more accurate and direct ways of assessing disease development, they not only inherit, so to speak, the problems related to mortality data we highlighted in the previous section but also suffer from systematic biases and substantial delays when the aim is to design and implement health policies and mitigation strategies. On the other hand, we do not want to deny the possible usefulness of R —given it is assessed in a critically and epidemiologically informed way and considered as a part of a collection of estimated epidemic features—but only to point out that policies based on it must not be taken as purely data driven.

To sum up, our (rather modest) point here is that in order to estimate the possibility of preventing deaths and thus to decide whether to reopen schools, restaurants, and other public services, establish social distancing and quarantine periods, support flexible working, and so on, multiple information, and not an exclusive focus on R or some other specific parameter, can *all* be useful. Of course, practical considerations should also be pondered, such as the fact that R is an intuitive property of an epidemic and is relatively easy to calculate but such considerations should be made explicit, along with a critical and public assessment of R and alternative epidemiological estimates and data. This may be especially true in the case of Covid-19, as mortality is unevenly distributed and disproportionately affects the elderly.

We conclude with two general comments. First, the general and somehow uncritical reliance on the “magic” parameter R can be seen as a case of “scientized politics” (Goldenberg 2021, 91). Maya J. Goldenberg argues that there is a tendency for governments and institutions to base policies on supposedly value-free scientific evidence, and “rather than appealing to explicit values that are likely not shared by all, ‘the evidence’ is proposed to adjudicate between competing claims” (Goldenberg 2021, 95). But scientized politics rests on an unrealistic view of science and, most importantly, it cancels the role of public debate about values and priorities.

Second, this article aims to contribute to a recent body of philosophical work that points out errors, biases, and hidden values in Covid-19 science; we assume, however, that such issues are possible (if not likely) in scientific practice, but they can be addressed without implying or suggesting that science should not be trusted, within its proper limits.

Acknowledgments

We would like to thank the two anonymous reviewers for *Philosophy of Medicine* for their careful, competent, and insightful comments and constructive criticisms.

Disclosure Statement

No competing interests were reported by the authors.

References

Adam, David. 2020. “A Guide to R : The Pandemic’s Misunderstood Metric.” *Nature* 583: 346–348. <https://doi.org/10.1038/d41586-020-02009-w>.

Adam, Dillon C., Peng Wu, Jessica Y. Wong, Eric H.Y. Lau, Tim K. Tsang, Simon Cauchemez, Gabriel M. Leung, and Benjamin J. Cowling. 2020. “Clustering and Superspreading Potential of SARS-CoV-2 Infections in Hong Kong.” *Nature Medicine* 26, no. 11: 1714–1719. <https://doi.org/10.1038/s41591-020-1092-0>.

Althoff, Keri N., David J. Schlueter, Hoda Anton-Culver, James Cherry, Joshua C. Denny, Isaac Thomsen, Elizabeth W. Karlson, et al. 2021. “Antibodies to Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in All of Us Research Program Participants, 2 January to 18 March 2020.” *Clinical Infectious Diseases* 74, no. 4: 584–590. <https://doi.org/10.1093/cid/ciab519>.

Amoretti, Maria Cristina and Elisabetta Lalumera. 2021a. “COVID-19 as the Underlying Cause of Death: Disentangling Facts and Values.” *History and Philosophy of the Life Sciences* 43, no. 4: 1–4. <https://doi.org/10.1007/s40656-020-00355-6>.

———. 2021b. “Non-Epistemic Factors in Epidemiological Models: The Case of Mortality Data.” *Mefisto: Rivista di medicina, filosofia, storia* 5, no. 1: 65–78. http://www.edizioniets.com/scheda.asp?n=9788846761194&from=&fk_s= (accessed 29 March 2022).

Anderson, Roy M. and Robert M. May. 1992. *Infectious Diseases of Humans*. Oxford: Oxford University Press.

Artalejo, J.R. and M.J. Lopez-Herrero. 2013. “On the Exact Measure of Disease Spread in Stochastic Epidemic Models.” *Bulletin of Mathematical Biology* 75, no. 7: 1031–1050. <https://doi.org/10.1007/s11538-013-9836-3>.

- Bacaër, Nicolas and M. Gabriela M. Gomes. 2009. "On the Final Size of Epidemics with Seasonality." *Bulletin of Mathematical Biology* 71, no. 8: 1954–1966. <https://doi.org/10.1007/s11538-009-9433-7>.
- Biswas, Raaj Kishore, Awan Afiaz, and Samin Huq. 2020. "Underreporting COVID-19: The Curious Case of the Indian Subcontinent." *Epidemiology and Infection* 148, e207: 1–5. <https://dx.doi.org/10.1017%2FS0950268820002095>.
- Borsboom, Denny, Gideon J. Mellenbergh, and Jaap van Heerden. 2004. "The Concept of Validity." *Psychological Review* 111, no. 4: 1061–1071. <http://dx.doi.org/10.1037/0033-295X.111.4.1061>.
- Breban, Romulus, Raffaele Vardavas, and Sally Blower. 2007. "Theory versus Data: How to Calculate R_0 ?" *PLoS ONE* 2, no. 3: e282. <https://doi.org/10.1371/journal.pone.0000282>.
- Broadbent, Alex and Benjamin T.H. Smart. 2020. "Why a One-Size-Fits-All Approach to COVID-19 Could Have Lethal Consequences." *The Conversation*, 23 March. <https://theconversation.com/why-a-one-size-fits-all-approach-to-covid-19-could-have-lethal-consequences-134252> (accessed 29 March 2022).
- Delamater, Paul L., Erica J. Street, Timothy F. Leslie, Y. Tony Yang, and Kathryn H. Jacobsen. 2019. "Complexity of the Basic Reproduction Number (R_0)." *Emerging Infectious Diseases* 25, no. 1: 1–4. <https://doi.org/10.3201/eid2501.171901>.
- Diekmann, Odo, J.A.P. Heesterbeek, and Johan A.J. Metz. 1990. "On the Definition and the Computation of the Basic Reproduction Ratio R_0 in Models for Infectious Diseases in Heterogeneous Populations." *Journal of Mathematical Biology* 28, no. 4: 365–382. <https://doi.org/10.1007/BF00178324>.
- Dietz, Klaus. 1993. "The Estimation of the Basic Reproduction Number for Infectious Diseases." *Statistical Methods in Medical Research* 2, no. 1: 23–41. <https://doi.org/10.1177%2F096228029300200103>.
- Drewnowski, Adam, Colin D. Rehm, and David Solet. 2007. "Disparities in Obesity Rates: Analysis by ZIP Code Area." *Social Science & Medicine* 65, no. 12: 2458–2463. <https://doi.org/10.1016/j.socscimed.2007.07.001>.
- Endo, Akira and Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group (Sam Abbott, Adam J. Kucharski, and Sebastian Funk). 2020. "Estimating the Overdispersion in COVID-19 Transmission Using Outbreak Sizes outside China." *Wellcome Open Research* 5: 67. <https://doi.org/10.12688/wellcomeopenres.15842.3>.
- European Commission. 2020. "Study on Informed Public Policy-Making on Base of Policy Modelling and Simulation – ANNEX Case Studies". <https://joinup.ec.europa.eu/sites/default/files/custom-page/attachment/2020-07/Study%20on%20informed%20public%20policymaking%20on%20base%20of%20policy%20modelling%20and%20simulation%20-%20ANNEX%20Case%20studies.pdf> (accessed 29 March 2022).
- Evans, Brian and Bernard Waites. 1981. *IQ and Mental Testing: An Unnatural Science and Its Social History*. New York: Macmillan International Higher Education.
- Ferguson, Neil M., Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, et al. 2020. "Report 9: Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand." *Imperial College London* 10, no. 77482: 491–497. <http://dx.doi.org/10.25561/77482>.

Fang, Ferric C., Constance A. Benson, Carlos Del Rio, Kathryn M. Edwards, Vance G. Fowler, David N. Fredricks, Ajit P. Limaye, et al. 2021. "COVID-19: Lessons Learned and Questions Remaining." *Clinical Infectious Diseases* 72, no. 12: 2225–2240. <https://doi.org/10.1093/cid/ciaa1654>.

Fuller, Jonathan. 2020. "What's Missing in Pandemic Models." *Nautilus*, 6 May. <http://nautil.us/issue/84/outbreak/whats-missing-in-pandemic-models> (accessed 21 August 2021).

———. 2021. "What Are the COVID-19 Models Modeling (Philosophically Speaking)?" *History and Philosophy of the Life Sciences* 43, 47. <https://doi.org/10.1007/s40656-021-00407-5>.

Goldenberg, Maya J. 2021. *Vaccine Hesitancy: Public Trust, Expertise, and the War on Science*. Pittsburgh: University of Pittsburgh Press.

Gwinnutt, Carl. 2021. "Measuring R_0 ." *Anaesthesia* 76, suppl. 3: 26. <https://doi.org/10.1111/anae.15351>.

Heesterbeek, J.A.P. 2002. "A Brief History of R_0 and a Recipe for Its Calculation." *Acta Biotheoretica* 50: 189–204. <https://doi.org/10.1023/A:1016599411804>.

Heesterbeek, J.A.P. and M.G. Roberts. 2007. "The Type-Reproduction Number T in Models for Infectious Disease Control." *Mathematical Biosciences* 206, no. 1: 3–10. <https://doi.org/10.1016/j.mbs.2004.10.013>.

Heffernan, J.M., R.J. Smith, and L.M. Wahl. 2005. "Perspectives on the Basic Reproductive Ratio." *Journal of the Royal Society Interface* 2, no. 4: 281–293. <https://dx.doi.org/10.1098%2Frsif.2005.0042>.

Hyde, Rob. 2021. "Controversy Surrounds Merkel's New Lockdown Powers." *The Lancet* 397, no. 10285: 1610. [https://doi.org/10.1016/s0140-6736\(21\)00992-2](https://doi.org/10.1016/s0140-6736(21)00992-2).

Istituto Superiore di Sanità. 2021. "ISS per COVID-19." https://www.iss.it/coronavirus/-/asset_publisher/1SRKHcCJJQ7E/content/faq-sul-calcolo-del-rt (accessed 21 August 2021).

Inglesby, Thomas V. 2020. "Public Health Measures and the Reproduction Number of SARS-CoV-2." *JAMA* 323, no. 21: 2186–2187. <https://doi.org/10.1001/jama.2020.7878>.

Jukola, Saana and Stefano Canali. 2021. "On Evidence Fiascos and Judgments in COVID-19 Policy." *History and Philosophy of the Life Sciences* 43, no. 2: 61. <https://dx.doi.org/10.1007%2Fs40656-021-00410-w>.

Lau, Max S.Y., Bryan Grenfell, Michael Thomas, Michael Bryan, Kristin Nelson, and Ben Lopman. 2020. "Characterizing Superspreading Events and Age-Specific Infectiousness of SARS-CoV-2 Transmission in Georgia, USA." *Proceedings of the National Academy of Sciences* 117, no. 36: 22430–22435. <https://doi.org/10.1073/pnas.2011802117>.

Li, Jing, Daniel Blakeley, and Robert J. Smith. 2011. "The Failure of R_0 ." *Computational and Mathematical Methods in Medicine*. <https://doi.org/10.1155/2011/527610>.

Lloyd-Smith, J.O., S.J. Schreiber, P.E. Kopp, and W.M. Getz. 2005. "Superspreading and the Effect of Individual Variation on Disease Emergence." *Nature* 438: 355–359. <https://doi.org/10.1038/nature04153>.

Mahase, Elizabeth. 2021. "Covid-19: Where Are We on Vaccines and Variants?" *British Medical Journal* 372. <https://doi.org/10.1136/bmj.n597>.

- Mercer, G.N., K. Glass, and N.G. Becker. 2011. "Effective Reproduction Numbers Are Commonly Overestimated Early in a Disease Outbreak." *Statistics in Medicine* 30, no. 9: 984–994. <https://doi.org/10.1002/sim.4174>.
- Ministry of Health, Istituto Superiore di Sanità, Consiglio Superiore di Sanità, Dipartimento della Protezione Civile, Istituto Nazionale Assicurazione Infortuni sul Lavoro, Fondazione Bruno Kessler, Conferenza Stato Regioni, et al. 2020. "Prevention and Response to COVID-19: Evolution of Strategy and Planning in the Transition Phase for the Autumn-Winter Season." Ministerial Decree of 30 April. https://www.iss.it/documents/5430402/o/COVID+19+_strategy_ISS_MoH+%281%29.pdf/fod91693-c7ce-880b-e554-643c049eaf3?t=1604675600974 (accessed 21 August 2021).
- Mintrom, Michael, Maria Rost Rublee, Matteo R. Bonotti, and Steven T. Zech. 2021. "Policy Narratives, Localisation, and Public Justification: Responses to COVID-19." *Journal of European Public Policy* 28, no. 8: 1219–1237. <https://doi.org/10.1080/13501763.2021.1942154>.
- Nadim, Sk Shahid and Joydev Chattopadhyay. 2020. "Occurrence of Backward Bifurcation and Prediction of Disease Transmission with Imperfect Lockdown: A Case Study on COVID-19." *Chaos, Solitons, and Fractals* 140, 110163. <https://doi.org/10.1016/j.chaos.2020.110163>.
- Nikbakht, Roya, Mohammad Reza Baneshi, Abbas Bahrapour, and Abolfazl Hosseinnataj. 2019. "Comparison of Methods to Estimate Basic Reproduction Number (R_0) of Influenza, Using Canada 2009 and 2017–18 A (H1N1) Data." *Journal of Research in Medical Sciences* 24, no. 1: 67. https://dx.doi.org/10.4103%2Fjrms.JRMS_888_18.
- Nishiura, Hiroshi and Gerardo Chowell. 2009. "The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends." *Mathematical and Statistical Estimation Approaches in Epidemiology*: 103–121. https://dx.doi.org/10.1007%2F978-90-481-2313-1_5.
- Northcott, Robert. 2017. "When Are Purely Predictive Models Best?" *Disputatio* 9, no. 47: 631–656. <http://dx.doi.org/10.1515/disp-2017-0021>.
- Pan, An, Li Liu, Chaolong Wang, Huan Guo, Xingjie Hao, Qi Wang, Jiao Huang, et al. 2020. "Association of Public Health Interventions with the Epidemiology of the COVID-19 Outbreak in Wuhan, China." *JAMA* 323, no. 19: 1915–1923. <https://doi.org/10.1001/jama.2020.6130>.
- Pandit, J.J. 2020. "Managing the R_0 of COVID-19: Mathematics Fights Back." *Anaesthesia* 75, no. 12: 1643–1647. <https://doi.org/10.1111/anae.15151>.
- Parker, Wendy S. 2009. "Confirmation and Adequacy-for-Purpose in Climate Modelling." *Aristotelian Society Supplementary Volume* 83, no. 1: 233–249. <https://doi.org/10.1111/j.1467-8349.2009.00180.x>.
- Pulla, Priyanka. 2020. "What Counts As a COVID-19 Death?" *BMJ* 370, m2859. <https://doi.org/10.1136/bmj.m2859>.
- Richardson, Ken and Sarah Norgate. 2014. "Does IQ Measure Ability for Complex Cognition?" *Theory & Psychology* 24, no. 6: 795–812. <http://dx.doi.org/10.1177/0959354314551163>.
- Ridenhour, Benjamin, Jessica M. Kowalik, and David K. Shay. 2014. "Unraveling R_0 : Considerations for Public Health Applications." *American Journal of Public Health* 104, no. 2: e32–e41. <https://doi.org/10.2105/AJPH.2013.301704>.

Roberts, M.G. and J.A.P. Heesterbeek. 2003. “A New Method for Estimating the Effort Required to Control an Infectious Disease.” *Proceedings of the Royal Society: Biological Sciences* 270, no. 1522: 1359–1364. <https://dx.doi.org/10.1098%2Frspsb.2003.2339>.

Ross, Catherine E. and John Mirowsky. 2008. “Neighborhood Socioeconomic Status and Health: Context or Composition?” *City & Community* 7, no. 2: 163–179. <http://dx.doi.org/10.1111/j.1540-6040.2008.00251.x>.

Serpico, Davide. 2018. “What Kind of Kind Is Intelligence?” *Philosophical Psychology* 31, no. 2: 232–252. <https://doi.org/10.1080/09515089.2017.1401706>.

Tal, Eran. 2020. “Measurement in Science.” In *The Stanford Encyclopedia of Philosophy*, edited by E.N. Zalta. <https://plato.stanford.edu/archives/fall2020/entries/measurement-science/> (accessed 19 August 2021).

Van den Driessche, Pauline. 2017. “Reproduction Numbers of Infectious Disease Models.” *Infectious Disease Modelling* 2, no. 3: 288–303. <https://dx.doi.org/10.1016%2Fj.idm.2017.06.002>.