COMMENTARY

American Journal of Human Biology WILEY

Transmission-dynamics models for the SARS Coronavirus-2

James Holland Jones[1] | **Ashley Hazel[1]** | **Zack Almquist[2]**

[1]Department of Earth System Science, Stanford University, Stanford, California

[2]Department of Sociology, University of Washington, Seattle, Washington

Correspondence
James Holland Jones, Department of Earth System Science, Stanford University, Stanford, CA, USA.
Email: jhj1@stanford.edu

1 | INTRODUCTION

Infectious disease is central to the human experience. It has served as a major source of selection on human populations (Armelagos & Dewey, 1970) and consequently many of the regions of the human genome that have been under particularly strong selection in the recent past (ie, since Holocene) are immune-related and presumably a function of infectious disease (Karlsson, Kwiatkowski, & Sabeti, 2014). Infectious disease is clearly central to our understanding of the peopling of the world (Black, 1975; Cockburn, 1971). Continuing to the present, infectious disease is a major determinant of population health (Heesterbeek et al., 2015) and a key contributor to persistent socio-economic inequality (Bonds, Keenan, Rohani, & Sachs, 2010; Sachs & Malaney, 2002).

It is therefore surprising that there has been relatively little formal work on infectious disease dynamics and their consequences for populations and individuals embedded within populations within the fields of Human Biology and Biological Anthropology. Human biology remains concerned with the consequences of infectious disease (eg, Dinkel et al., 2020; Gurven, Kaplan, & Supa, 2007; Houldcroft, Ramond, Rifkin, & Underdown, 2017; Houldcroft & Underdown, 2016; McDade et al., 2012), but with precious few exceptions (Hazel & Jones, 2018; Hazel, Marino, & Simon, 2015; Hazel, Ponnaluri-Wears, Davis, Low, & Foxman, 2014; McGrath, 1988; Nakazawa, Ohmae, Ishii, & Leafasia, 1998; O'Neil & Sattenspiel, 2010; Sattenspiel, Koopman, Simon, & Jacquez, 1990; Upham, 1986) generally does not integrate the rich body of theory and methodology derived from mathematical models of transmission dynamics.

The COVID-19 pandemic presents an opportunity to engage these research communities. We will focus our mini-review on topics that we think are likely to resonate with the readership of AJHB. After introducing the formalisms of transmission-dynamics models for infectious disease and how these models have been used to gain insight into the origin, amplification, and dissemination of the SARS Coronavirus-2, the causative agent of COVID-19, we will turn to topics of particular interest for anthropologists: the role population structure plays in shaping transmission dynamics, geography and mobility, simple models of socio-economic and health inequality and their implications for epidemic control, and the consequences for structured interpersonal relations, as formalized using networks, for disease transmission and control.

2 | TRANSMISSION DYNAMICS

Every disease transmission event has a social cause. At the heart of every transmission, and the formal machinery for modeling infectious disease dynamics, is a social interaction. Susceptible individuals need to come in contact with infectious ones. For the great majority of models, this is about as far as the model of social behavior goes: contacts are assumed to happen at random, proportional in a bilinear way to the number of susceptibles and number infected in the population.

Epidemics are characterized simultaneously by extreme uncertainty and extreme structure. Early in an outbreak, when the number of infections is low, randomness dominates and prediction is very difficult. As an

outbreak gets larger, it starts to take on momentum and becomes more predictable. Once an epidemic is underway, it enters an exponential growth regime that is quite predictable. A robustly-growing epidemic, in which cases double in short intervals, takes on a great deal of inertia, often before people notice that the epidemic is serious. This is why epidemic controls needs to be effected early if it is to have the greatest effect. Unfortunately, early in the epidemic is when uncertainty is highest and pushback from both politicians and the general public is likely to be strong.

With some assumptions about population structure, the behavior of hosts, and the properties of the pathogen, we can construct a *compartmental model* which tracks the stocks and flows between different elements of a population (ie, the "compartments"). These models typically take the form of a collection of coupled, nonlinear ordinary differential equations. The canonical compartmental model is probably the SIR model. For a closed population of N individuals, where S are susceptible, I infected, and R are removed, the dynamics are characterized by three equations:

$$\frac{dS}{dt} = -\beta SI/N \tag{1}$$

$$\frac{dI}{dt} = \beta SI/N - \gamma I \tag{2}$$

$$\frac{dr}{dt} = \gamma I \tag{3}$$

where $\beta = \tau \bar{c}$ and is known as the *effective contact rate* and is comprised of an average rate of per-capita contact ($\bar{c}$) and a transmission probability conditional on contact (τ), and γ is the removal rate. By assumption all rates are constant. This means that the expected duration of infection is simply the inverse of the removal rate: $d = \gamma^{-1}$.

These equations are coupled because the outputs of one serve as inputs for others. They are nonlinear because the incidence (dI/dt) is driven by the multiplicative interaction of the susceptible and infectious compartments—this is the social behavior that underlies the model. The solution to Equations (1)–(3) leads to an asymmetric bell-shaped incidence curve and approximately symmetric sigmoid functions of susceptibles (declining) and removed (increasing). The exponential-growth phase of the epidemic curve can be seen in Figure 3. The decline of the epidemic curves is typically quite a bit slower than exponential, giving epidemic curves their asymmetric bell-shape.

The formulation of the SIR model presented in Equations (1)–(3) assumes that there are neither births nor
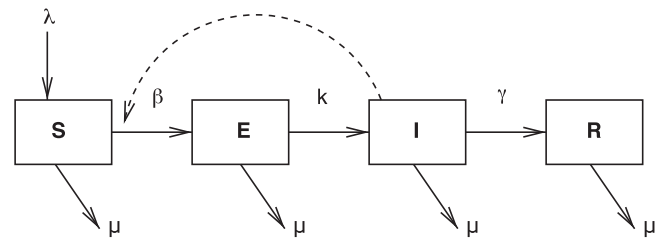

FIGURE 1 Susceptible-exposed-infected-removed model state diagram

deaths in the population (ie, the population is "closed"). While this may seem like an absurd assumption to make, it translates into the more reasonable interpretation that the dynamics of the epidemic are much faster than the vital dynamics of the population. This is clearly not a valid assumption to make for endemic infectious diseases. Based in part on the immunological profile of COVID-19 survivors, there is some suggestion that COVID-19 might become an endemic disease and understanding the continuing dynamics of the disease will require open models. Bjornstad (2018) provides and excellent introduction to SIR-type models for open populations.

Compartmental models can get much more complex than the SIR model. An important elaboration of the basic SIR adds a compartment for people who have been infected but are not yet themselves infectious. This extract compartment, typically called "Exposed," accounts for the incubation period of the infection and builds in a lag in observed infections. Most serious research applications of mathematical models for understanding the dynamics of COVID-19 have taken the SEIR form. Figure 1 presents a state diagram for the SEIR model. This particular formulation of the SEIR model includes births (λ) and deaths (μ). The rate of movement from the exposed to the infectious compartment is given by k. Otherwise, it is largely similar to the SIR model.

2.1 | The basic reproduction number of an infectious disease

The basic reproduction number, denoted R_0, is the expected number of secondary cases produced by a single typical infection early in the epidemic. It is closely related to the R_0 more familiar to demographers, namely the net reproduction rate (Heesterbeek, 2002) in that it reflects the *per generation* ratio of population size. In the case of an epidemic, however, the population size that matters is that of infectious individuals. This means that the essential identity of demography, namely $R_0 = exp\{rT\}$, where r is the intrinsic rate of increase and T is generation length

(ie, mean age of childbearing), applies. In epidemiology, this relationship holds as well, only we interpret r as the exponential rate of increase in infections and T is known as the *serial interval*, the mean duration between the onset of symptoms between index and secondary infections. This relationship provides an easy way to estimate R_0 from case data. The Taylor series expansion of an exponential function is $e^x = 1 + x$ to the first order. This allows us to write $R_0 \approx 1 + rV$, where we have written V for the serial interval.

R_0 is a threshold parameter. From the SIR equations, solve for $dI/dt > 0$, assuming at the outset of an epidemic $S/N = 1$. What results is a simple ratio of the rate at which new infections are added to the population to the rate at which infections are removed from the population, namely, $R_0 = \beta/\nu$. Note that β is a composite parameter, incorporating both the contact rate of people in the population and transmissibility of the pathogen, $\beta = \bar{c}\tau$, and that $1/\nu = d$, the expected duration of infectiousness. This makes the basic reproduction number the product of three elements: (1) the contact rate between susceptible and infectious people, (2) the transmissibility of the pathogen, and (3) the duration of infectiousness. This decomposition largely provides the theory of infectious-disease control. An epidemic can only increase when $R_0 > 1$; control efforts should therefore focus on bringing this quantity below unity. There are really only three ways to do this: (1) reduce the contact-rate between susceptible and infectious people (sheltering-in-place, quarantine, travel restrictions), (2) reduce transmissibility (use of personal protective equipment, vaccination, therapeutics which reduce viral shedding), or (3) reduce the duration of infectiousness (make sick people well).

In addition to its fundamental role as an epidemic threshold parameter, R_0 determines a number of other fundamental features of epidemics. Of particular interest for the COVID-19 pandemic, R_0 tells us about the *critical vaccination threshold* or level of herd immunity (ie, the level of immunity that prevents the epidemic from increasing in the presence of a small number of infections), and the *final size* of the epidemic. The critical vaccination threshold is found simply by noting that an epidemic cannot take place if $R_0 < 1$. Solve the equation $R = R_0(1 - p)$ for p, the proportion of the population removed from the population through vaccination. The critical fraction is clearly $p_c = 1 - 1/R_0$, where we have denoted the critical fraction of the population p_c. If $R_0 = 2.5$, the critical vaccination threshold is 60% of the population. This is similarly the threshold for herd immunity. At the time of writing, the highest prevalence of antibodies for SARS-CoV-2 in the United States is 14% in New York State, with an estimate of 22.7% in New York city (Rosenberg et al., 2020). While testing is highly

variable and plagued by sampling biases, the prevalence of antibodies appears to be much lower in most of the country (for Disease Control and Prevention, 2020a). Clearly, at seven months into the pandemic, we are a long way from herd immunity, at least for anything resembling a well-mixed population.

If we assume that the parameters of the SIR model remain constant, we can calculate the fraction of the total population that will become infected by the end of an epidemic. This is called the *final size* of the epidemic and it is calculated by dividing Equation (2) by Equation (1) and integrating. The resulting equation is:

$$\log(s_\infty) = R_0(s_\infty - 1), \qquad (4)$$

where s_∞ is the fraction of the population still susceptible at the end of the epidemic. This equation always has a solution at $s_\infty = 1$, while for $R_0 > 1$, it also has a solution $s_\infty < 1$. The complement of this value is the final size.

Figure 2 shows the final size for a range of values of R_0 that correspond to the typical range of observed basic reproduction numbers for COVID-19. An important observation to make about the final size of an epidemic: it will generally be larger than the herd-immunity threshold. For example, if $R_0 = 2.5$, the threshold for herd immunity is 60% but the final size of the epidemic is 89.7%. How can that be? The simple answer is that an epidemic cannot start when 60% of the population are removed. However, the momentum of an ongoing epidemic will carry it well through this threshold.

We have limited our discussion so far to R_0 for the simple, unstructured SIR model (Figure 3) because of the
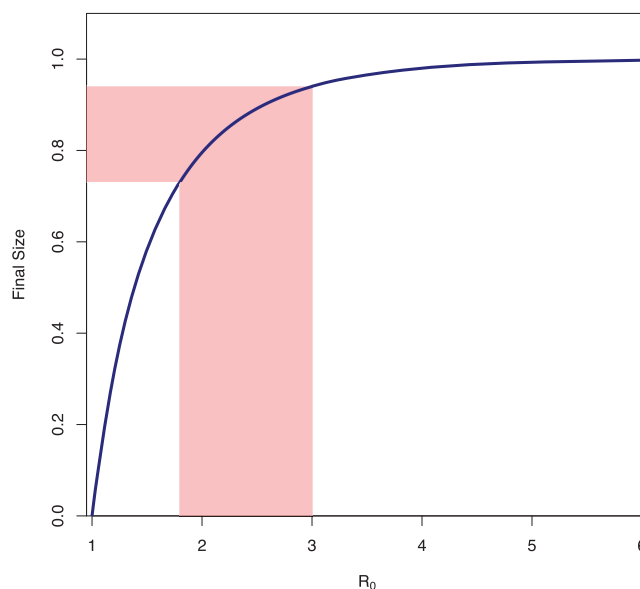

FIGURE 2 Final size calculations for a range of R_0 values consistent with COVID-19
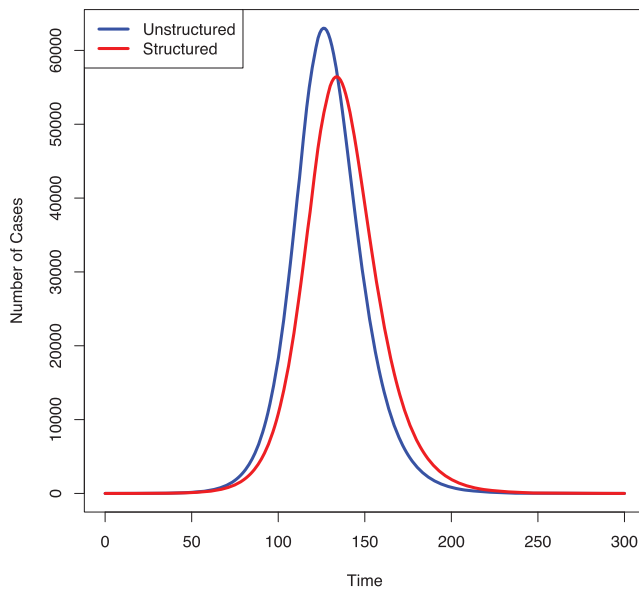
FIGURE 3 Epidemic curves for simple structured and unstructured SIR models

heuristic value of its simple product form, $R_0 = \bar{c}\tau d$. R_0 will be different for different epidemic models. While it is beyond the scope of this brief review, there is a highly-elaborated theory for calculating R_0 in more complex models (Diekmann, Heesterbeek, & Metz, 1990). Briefly, for an infection with discrete disease states, R_0 is calculated as the dominant eigenvalue of a square $k \times k$ matrix (where k indicates the number of distinct disease states) known as the *next-generation matrix*, **G**. This matrix is composed of elements g_{ij}, which can be thought of as a type-specific reproduction numbers, accounting for the number of type i infections caused by infectious contact with individuals of type j. Taking the eigenvalue of this matrix effectively averages over the different ways that infections can be created in a structured model, allowing these models to conform to the notion that R_0 is the expected number of secondary infections produced by a single, *typical* index infection. More detailed notes on the calculation of R_0 can be found in Jones (2020).

It is important to note that R_0 is *not* a property of the pathogen exclusively. It is a property of the epidemic and it incorporates, even if only in a highly-stylized manner the social structure and behavior of the host organism (Arthur, Gurley, Salje, Bloomfield, & Jones, 2017).

3 | SPECIFIC COVID-19 INSIGHTS FROM EPIDEMIC MODELS

Transmission-dynamics models have been used to gain critical insights into the behavior of the COVID-19 pandemic. In particular, models allowed epidemiologists and governments to get a handle on the size and scope of the epidemic at a time of great uncertainty. Furthermore, a number of studies have focused on the efficacy of non-pharmaceutical interventions (NPI), a critical question at the outset of an epidemic before there is a vaccine or effective therapeutic treatment. Most of the published models focus on the early outbreak in China. Several key studies estimated fundamental epidemiological parameters that have been essential in formulating models. Based on epidemiological investigation in China, Li, Guan, et al. (2020) and Li, Pei, et al. (2020) estimated the COVID-19 incubation period at 5.2 days, the serial interval of 7.5 days, $R_0 = 2.2$, and a doubling time of 7.4 days. Zhang et al. (2020) estimated the same incubation period but a shorted serial interval, both at 5.2 days. Following the end of January, they estimate that $R_t < 1$ for all provinces for which they had sufficient data to estimate it.

Models have allowed us to ascertain the scope of the epidemic and the features of the transmission dynamics of the SARS CoV-2 that make it so difficult to control. Wu, Leung, et al. (2020) and Wu, Nethery, et al. (2020) used an SEIR framework early in the epidemic in China to suggest that there were more than 75 000 infections at a time when the official cumulative prevalence was 4528, a ratio of 16 infections per case. They estimated an $R_0 = 2.68$ and noted that the infection had already been disseminated well beyond Wuhan. Li, Guan, et al. (2020) and Li, Pei, et al. (2020) linked an SEIR metapopulation model built on an explicit network with Bayesian inference to estimate that 86% of the infections in China were undocumented prior to China imposing severe travel restrictions. Per person, undocumented infections accounted for fewer secondary infections, but because of their large number, accounted for the great majority of total transmissions. The ability to infer the presence of a large degree of pre- or asymptomatic transmission would not have been possible without the aid of the mathematical model of transmission dynamics.

Chinazzi et al. (2020) found that intense travel restrictions imposed in Wuhan had the effect of delaying the epidemic by 3 to 5 days in mainland China, and also had a substantial effect on slowing the spread internationally. Importantly, they found that travel restrictions were not effective unless accompanied by substantial (~50%) reductions in transmission rate. Similarly, Prem et al. (2020) employed an SEIR model parameterized with movement data from Wuhan, China to show that movement restrictions have the greatest impact in delaying the epidemic and reducing the final size when they were eased in a gradual manner over an extended time frame. The sudden easing of restrictions would lead to a second major wave of infections. Lai et al. (2020) used anonymized movement data to simulate counterfactual

epidemic scenarios in China. They find evidence for the remarkable effectiveness of NPI. In the absence of NPI, the epidemic would likely have been 67-times larger by the end of February than it was. Consistent with historical results of Hatchett, Mecher, and Lipsitch (2007), who analyzed the effect of timing of NPI on the 1918 to 1919 influenza pandemic, speed of response in identifying and isolating infections had the largest effect on reducing the size of the epidemic, but combining this with other NPIs such as reducing contacts and imposing travel restrictions had the greatest total impact.

Moving on from understanding the early epidemic in China, research teams are increasingly looking at the later stages of the pandemic. Using stochastic simulations based on an SEIR model, Hellewell et al. (2020) show that under most reasonable conditions, contact tracing, coupled with testing and isolation, is sufficient to control COVID-19 once R_0 has been brought down to the range of 1.5. Kissler, Tedijanto, Goldstein, Grad, and Lipsitch (2020) predict recurrent winter outbreaks following the initial pandemic wave. They anticipate that intermittent social distancing will be necessary to keep critical cases below the medical-capacity threshold through at least 2022. Moghadas et al. (2020) found that with $R_0 = 2.5$, the US would need nearly four times the number of ICU beds than it has to accommodate the increased demand from severe cases. They then found that self-isolation by 20% of infections would reduce the number of ICU beds needed at peak by nearly half. Reducing R_0 by half a secondary infection meant that ICU demand would only be twice capacity and that self-isolation of 20% reduced peak demand by nearly 75%. These results suggest that self-isolation is a highly efficacious and cost-effective NPI for COVID-19.

4 | THE IMPACT OF HUMAN BEHAVIOR AND SOCIAL STRUCTURE ON EPIDEMIC OUTCOMES

Epidemic models provide us with powerful tools for understanding the consequences of behavior, social structure, and inequality on epidemic outcomes.

4.1 | Structure

Epidemic models, like the SIR model, involve dyads of individuals—one susceptible and one infectious—coming together at a specified rate and generating a new infection with a specified probability. We say that a population is *well-mixed* if all infectious-susceptible dyads in the

population have approximately the same probability of occurring. This is obviously a very strong assumption. For example, a student at Stanford is probably much more likely to encounter another Stanford student infected with the COVID-19 than she is an infected student at, say, the University of Washington.

We can relax the assumption of a population being well-mixed by adding structure to it. The resulting model will be characterized by *structured mixing*, meaning that there are potentially quite different probabilities associated with dyads that can be formed from the various elements of structure (Morris, 1991, 1993). This structure can represent geographic location (eg, Palo Alto, California vs Seattle, Washington) or it can represent various mechanisms by which social or cultural attributes affect the way people interact, for example, race/ethnicity, age, gender, occupation, social class, income quartile, and so on.

A structured model will always be slower than the well-mixed case and will typically be smaller as well. This sounds like unmitigated good news: slower epidemics mean there is more time to intervene and smaller epidemics mean there is less morbidity and, presumably, mortality associated with the epidemic. While these are true, there is a darker side to structure. First, structure can create pockets of high-prevalence in a population. These pockets can then serve as sources from which new infections can invade the broader population. Such source-sink dynamics are analogous to the so-called "rescue effect" in metapopulation biology and are the main explanation for why diseases such as gonorrhea persist in rich countries where the behavior of the typical person would not support endemicity (Hethcote & Yorke, 1984). Second, and related to this, is the effect of heterogeneity on R_0 itself. Nold (1980) noted that heterogeneity in contact rate and transmissibility can increase disease prevalence even if the means of these parameters are not altered. Anderson, Medley, May, and Johnson (1986) showed that heterogeneity in contact rates effectively increases R_0, deriving a relationship that expresses this effect. Assuming a linear effect of number of contacts on transmission, Anderson et al. (1986) show that $R_0 = R_0^{(m)}(1 + c^2)$, where $R_0^{(m)}$ is the basic reproduction number estimated from the mean values of contact, transmissibility, and duration of infection, and c is the coefficient of variation in contact rates. This result shows that the effective R_0 is directly proportional to variance in contact rates. When population are structured into groups where contact is heterogeneous, the effective R_0 increases. A paradoxical consequence of this increase in the effective R_0 is that the final size of the epidemic will be considerably smaller. As noted by Anderson and May (1991), May and Lloyd (2001), and discussed in the

context of network models for STIs by Handcock and Jones (2006), as $c \to \infty$, the final size approaches zero. This dual effect of heterogeneity making epidemics more likely but ultimately smaller than the well-mixed case may prove very important for understanding COVID-19.

4.2 | The population impact of inequality on an epidemic

An important dimension along which populations are structured is socioeconomic inequality (Subramanian, Belli, & Kawachi, 2002). We can use the mathematical formalism of infectious disease to help us understand the consequences for the epidemic on the existence of health inequalities. We will assume a very simple model where the population is divided into two components, wealthy and poor, and leading to a 2×2 next-generation matrix. Homophily and especially residential segregation lead to the diagonal elements (ie, infections within a category) being greater than the off-diagonal elements. There is mixing in the population such that both wealthy and poor can infect others who are wealthy and poor, but mixing is asymmetric. This asymmetry arises from a number of social features of the socioeconomic landscape. Of particular relevance to COVID-19, poor people are more likely to be engaged in high-risk, "frontline" labor, making it more likely that they will receive infection from both wealthy and poor. Furthermore, poor people live in more crowded housing, making NPIs such as sheltering-in-place structurally less effective than they are for wealthy people (Richardson et al., 2020). This latter observation leads to the expectation that the type-specific reproduction number for poor people will be substantially greater than the analogous number for wealthy people. The next-generation matrix is thus:

$$\mathbf{G} = \begin{bmatrix} g_{w \leftarrow w} & g_{w \leftarrow p} \\ g_{p \leftarrow w} & g_{p \leftarrow p} \end{bmatrix}. \quad (5)$$

The assumptions articulated above translate into the following ranking of type-specific reproduction numbers:

$$g_{pp} > g_{ww} \gg g_{pw} > g_{wp}$$

Following the logic of the ordering of elements, assume that $g_{ww} = g_{pp}/k$ for some risk ratio, $k > 0$. Further assume that $g_{pp}, g_{ww} \gg g_{wp}, g_{pw}$ so that the number of infections generated within a compartment greatly exceeds the number generated between compartments.

With these assumptions, $\mathbf{G}$ is effectively a diagonal matrix and the basic reproduction number for this system is the larger of the two diagonal elements.

When $k \gg 1$, R_0 is dominated by g_{pp}. This follows from the fact that the eigenvalues of a diagonal matrix are simply the diagonal elements of the matrix and the dominant eigenvalue will be largest diagonal element. If the diagonals are greater than the off-diagonals, as we have assumed, $\mathbf{G}$ approximates a diagonal matrix. If more infections are generated within the poor compartment, more effort should be allocated to controlling the epidemic in that compartment.

This discussion has focused on $k \gg 1$. It is desirable to know the sensitivity to changes in the within- and between-compartment reproduction numbers. Differentiate the characteristic equation for the eigenvalue of the next-generation matrix with respect to g_{pp} and scale the derivative by g_{pp}/R_0, yielding a proportional sensitivity, or elasticity, of R_0. Elasticities have the convenient property that the sum across all elasticities is unity. The elasticity of a particular element therefore represents the relative effectiveness of a perturbation for reducing R_0. Assuming that $R_0 > 1$, Figure 4 plots the elasticity of R_0 with respect to g_{pp}. As k increases, the elasticity of R_0 with respect to a change in g_{pp} approaches unity. Increasing the number of expected cross-compartment infections slows this approach, but the qualitative behavior remains largely the same. If $k = 2$ and transmission is dominated by the diagonal, more than 50% of the total elasticity is accounted for by g_{pp}, as indicated by the dashed line.

The upshot of this simple model is that when there is structured heterogeneity in infection risk in a population,
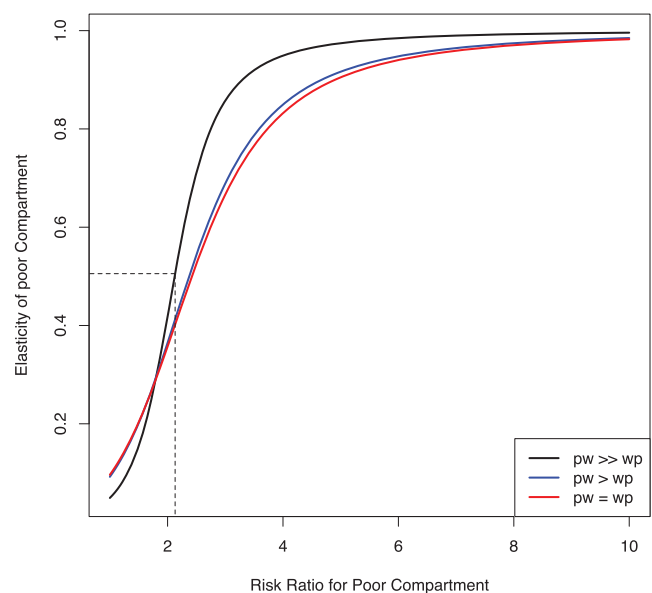

FIGURE 4 Elasticity of R_0 with respect to element g_{pp} of the next-generation matrix

the most efficient way to bring R_0 below the epidemic threshold is to focus control on the highest-risk segments of the population. In the context of COVID-19, this involves providing resources that allow frontline workers, especially those living in conditions that make sheltering-in-place effectively impossible. Providing free rooms in unused hotels to allow self-isolation for contacts who live in crowded housing and replacing lost wages for self-isolators are two specific policies for reducing g_{pp}. Mitigating the conditions that make meat-processing plants and prisons vessels for super-spreading could also disproportionately reduce the epidemic (Leclerc et al., 2020).

4.2.1 | Pollution and differential exposure

Pollution and poor ventilation are additional drivers of transmission and disease severity for COVID-19, and the distribution of exposure to both pollution and poorly-ventilated homes and work environments is highly unevenly distributed across populations. Including environmental exposures to high levels of pollution in formal models may be essential for understanding the differential impact of infection on different groups. Wu, Leung, et al. (2020) and Wu, Nethery, et al. (2020) attribute an 8% increase in death rate to just 1 μg/m^3 in PM2.5 exposure increase, according to an analysis of COVID-19-attributed deaths across 3000 US counties, which is consistent with pollution effects on death rates during the 2003 SARS (Cui et al., 2003). In Italy, the northern provinces at the epicenter of the country's epidemic have some of the worst air pollution in Europe, and chronic air pollution was significantly associated with COVID-19 cases across the 71 Italian provinces (Fattorini & Regoli, 2020). In both the US and Italy, long-term exposure had a more robust influence on COVID-19 death rates than short-term exposure.

African Americans mortality rates from COVID-19 are more than three times as high as white Americans (Gross et al., 2020), and death rates in the hardest-hit cities are disproportionately higher in non-white majority communities, regardless of residential density (Disease Control and Prevention, 2020b). Among several aspects of elevated risk that African Americans face, such as denser communities, greater reliance on public transportation, and a higher rate of front-line employment, long-term chronic pollution exposure is an important contributing factor (Brandt, Beck, & Mersha, 2020).

Pollution increases disease severity and risk of death from respiratory illness by instigating chronic inflammation and cell damage in lung tissue as well as suppressing the early immune response (Wu, Leung, et al., 2020; Wu,

Nethery, et al., 2020). While the hyper-inflammatory and respiratory and cardiac problems associated with COVID-19 are by now well documented, it is still unclear how the SARS-CoV-2 virus interacts with particulate matter (PM) during seroconversion and over the course of disease (Fattorini & Regoli, 2020) and whether PM directly inhibits the lungs' ability to clear the virus (Brandt et al., 2020).

4.3 | Transmission heterogeneity and network sampling

Since R_0 is an expectation of secondary cases, we can think about how the distribution of secondary cases might affect R_0 and subsequent dynamics of an infection.

COVID-19 is characterized by robust reproduction numbers. There have been many estimates now, with the majority of them falling between $R_0 = 2 - 3$. However, the early spread of the infection was, at times, halting. Why? In a provocative preprint, Grantz, Metcalf, and Lessler (2020) suggested that this might be due to transmission heterogeneity. These results have since been confirmed by Endo et al. (2020).

Coronaviruses consistently feature substantially skewed transmission (Munster, Koopmans, van Doremalen, van Riel, & de Wit, 2020). That is, coronavirus transmission dynamics are characterized by the presence of super-spreading events (Lloyd-Smith, Schreiber, Kopp, & Getz, 2005). While the presence of super-spreading is obviously not good, there is an upside. In particular, the presence of a few super-spreaders can drag the expected number of secondary cases (which is what R_0 is at the outset of an epidemic) out toward the tail that they define. The only way you can have super-spreading events, where dozens of secondary cases are created, and a value of R_0 on the order of 1 to 3 (or even 4-6) is for most people to infect a very small number of new people. The distribution of secondary cases is highly skewed and the mode probably less than one (Leclerc et al., 2020).

The intuition behind Grantz and colleagues' explanation of the epidemiological facts of COVID-19 is that, if we assume that the expected number of secondary cases a person is likely to generate is a feature of their physiology or the circumstances of their infection (ie, it can be thought of as a trait they take with them) and you pick people at random with respect to this distribution, you are likely to sample mostly people who are not going to generate many secondary cases. As a result, the infection chains emanating from them are more likely to die out quickly and the amount of epidemic dispersal will be limited.

This interpretation has a lot in common with the problem of sampling networks. Indeed, it can actually be
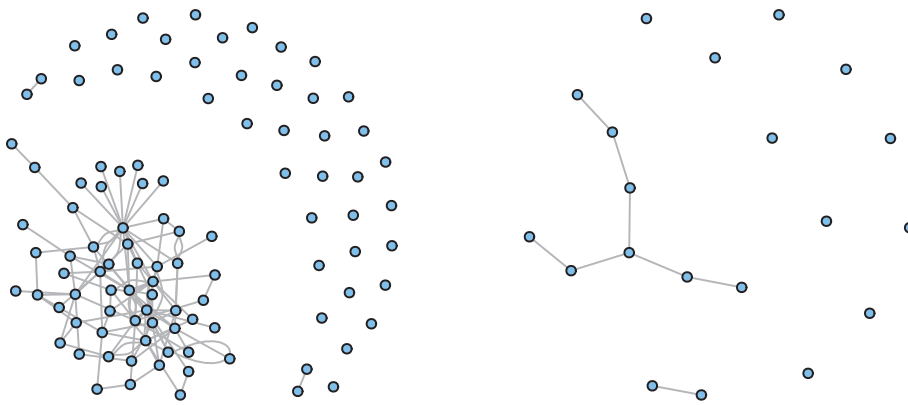
FIGURE 5 Heterogeneous network of 100 nodes and sampled network from a random sample of 20 nodes

thought of as a network-sampling problem. It is well known that a random sample of the vertices of a network will lead to a biased sample of the network's edges, and vice-versa (though this can be solved by re-weighting the sample, see Gjoka, Kurant, Butts, & Markopoulou, 2010). This is the basis of the famous friendship paradox, first noted by Feld (1991), namely that your friends have more friends than you do. When the degree distribution of a graph is heterogeneous, and your sample is random with respect to vertex, you will likely under-sample the edges of the graph, making the induced subgraph arising from the sampling possibly far less connected than the parent graph. Consider a model for a network that is a graph comprised of a set of vertices and edges connecting vertices: $\mathcal{G} = \{V, E\}$. An undirected edge between two individuals i and j indicates the possibility of transmission across the dyad. We will start with a random graph of 100 vertices drawn from a skewed degree distribution and then draw a random sample of 20 vertices from this graph.

The sampled graph (right panel of Figure 5) is far less connected than that of the graph of the population from which it is sampled (left panel of Figure 5). There is nothing inevitable about this. If we had the ability to sample edges or perhaps weight our sample by the degree distribution (as, eg, in some adaptive network sampling schemes: Salganik & Heckathorn, 2004), we could generate a more strongly-connected sample (Figure 6).

The model we have presented here is somewhat abstract, but it is not difficult to see how we could represent the number of potential secondary transmission events from a given case as a network. An epidemic is indicated if the resulting graph is strongly-connected. Sampling the vertices of the heterogeneous network (analogous to moving away from the epidemic focus in Wuhan) leads to an unconnected subgraph and the epidemic dies out. These insights take on paramount importance as we move into later parts of the pandemic where stuttering transmission will again be an issue and a disproportionate fraction of all new cases are likely to arise from super-spreading events.
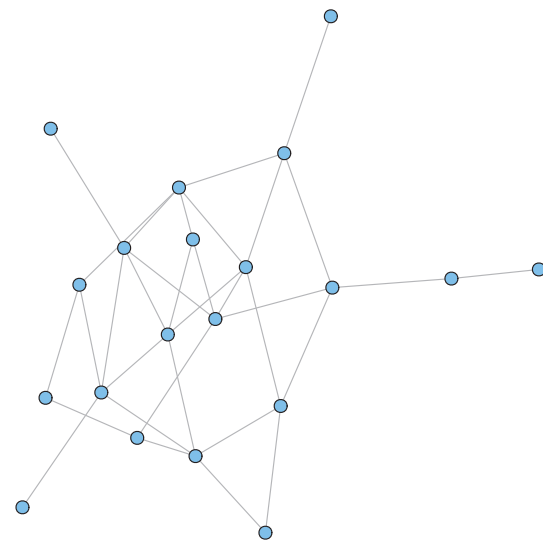

FIGURE 6 Degree-weighted sample of 20 nodes, showing that sparseness of the sampled network is not inevitable but arises from the bias introduced by sampling edges from a sample of nodes

These methods were actually developed to count sub-populations where a sampling frame was not available (eg, the number of individuals who have contracted HIV). Another practical use of network sampling methods would be to estimate the number of infected and recovered individuals of COVID-19. See Gile and Handcock (2010) for thorough review.

4.4 | Networks, control, and behavior change

Networks represent a way to conceptualize human interaction and disease spread. They can also provide strategies for transmission mitigation (eg, social distancing). Network models have been used to understand some of the heterogeneity in transmission that interact nontrivially with spatial heterogeneity of human populations (see Thomas et al. (2020)). Further, network models can

FIGURE 7 Social distancing pushes the network to forefront of COVID-19 modeling. We start with no social distancing, figure A; focusing on a network with mean degree of 5. Contrast this with an empty network (perfect social distancing) in B. Now, let's consider some loosening of social distancing with essential works at 10% of the population, C. Last, we what if we allow household mixing of just one household member (at random) we get figure D

(A) Mean Degree of 5

(B) No Connection

(C) Essential Workers
(10% of the population)

(D) One Household Mixing
and Essential Workers
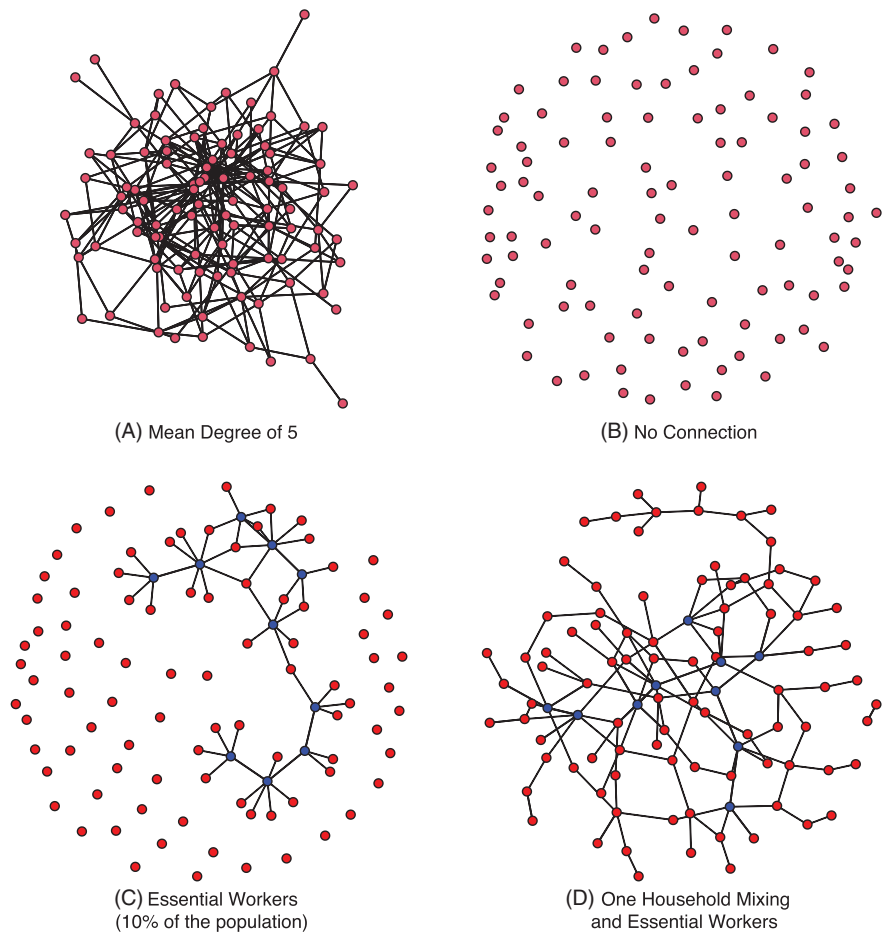
be used to understand how fragile social distancing strategies can be.

Another counterfactual to consider is that of the network effects of *social distancing*. In Figure 7 we illustrate the Goodreau, Pollock, Birnbaum, Hamilton, and M (2020) effect of social distancing which they refer to as "can't I visit just one friend?". We start this exercise with no social distancing (Figure 7A), which we illustrate with a random network that has a mean degree of five people (eg, each person in the network has on average five interactions which could lead to spread of COVID-19). Next, we consider the case of *perfect* social distancing (Figure 7B). In this case, we go from almost everybody being reachable/transmissible (largest component divided by total population) of 100 to 0. Now, let's consider allowing for essential workers. We set this to 10% of the population (Figure 7C). This results in an increase of 0 connections to 46. What happens if we loosen social distancing just a little more, such that we allow just one household member to mix with another household member? This results in an increase from 46 to 98 or almost a factor 4 times as many potential infections. Given the minimal exposure requirement for the spread of COVID-19 this sort of

simulation shows that it does not take much mixing to break down the effect of social distancing.

4.5 | Seasonality of COVID-19?

There has been much speculation about the underlying environmental drivers of COVID-19, particularly whether the COVID-19 epidemic will wane in the summer months. It is too early to know if SARS-CoV-2 has a survival advantage in cold weather, but the strong seasonality of endemic respiratory viruses, particularly influenza and other endemic (non-SARS) coronaviruses (CoV), in temperate countries is instructive. However, important differences in virulence and immunological interactions limit the utility of these inferences (Yang et al., 2018). Additionally, the environmental drivers of established respiratory viruses are either understudied (CoVs) or highly complicated (influenza), and are themselves active areas of scientific debate. Nevertheless, epidemic forecasting is a critical public health tool because accurate fine-grained models can steer life-saving changes in policy and outreach (Kramer & Shaman, 2019).

The seasonality of influenza and CoV is well established in temperate conditions. Although, the United States Centers for Disease Control (US CDC) only began reporting CoV infection rates in 2018, CoV illness in the US shows strong peaks in January and February with low summer incidence (Shaman & Galanti, 2020). It is unclear whether transmission of endemic seasonal respiratory viruses mostly disappear during the summer months in temperate countries and are then reintroduced by long-distance travel from locations with active outbreaks, or if undetected transmission carries on at low levels until weather conditions favor virus survival and accelerated transmission. Shaman et al. (2018) investigated virus presence among ambulatory adults in New York City during the low epidemic months (April-July) and found that 7.2% of participants were positive for at least one respiratory virus, and 21.5% of detected viruses were seasonal CoV. Depending on how symptomaticity was defined, they estimated between 57.7%and 93.3% of virus-positive people were asymptomatic. Even if there is a significant drop in virus circulation in the Northern Hemisphere's summer months, ongoing asymptomatic shedding can be an important contribution to a second wave as weather becomes colder again. If low-level transmission in the global north coincides with re-openings of economies and national borders as well as cooler weather in the Southern Hemisphere, we could see a dangerous shift in SARS-CoV-2 distribution toward countries with poorer public health infrastructure, denser urban settlements, and more remote rural populations with erratic access to hospital care. As Buckee et al. (2020) point out, large-scale mobility, not just local travel, is important to incorporate into COVID-19 models, particularly because SARS-CoV-2 is not (yet) an endemic virus, and its ongoing potential to cause epidemics worldwide will be driven by long-distance travel as well as by local livelihood-related mobilities. Buckee and various colleagues who are among the co-authors in Buckee et al. (2020) have pioneered the inclusion of remotely-sensed movement data into epidemic models (eg, Bharti et al., 2011; Buckee, Tatem, & Metcalf, 2017; Wesolowski et al., 2012, 2015, 2017; Wesolowski, Buckee, Engø-Monsen, & Metcalf, 2016).

Relative to CoV, influenza surveillance and case report data for the US are more robust and have been collected for a longer time period, so forecasts of annual epidemic peaks and intensities are relatively accurate and well-calibrated. Furthermore, extensive retrospective data enable better modeling of environmental effects so that the specific weather elements that contribute to epidemic forcing can be teased out. Transmission dynamics models have shown that small differences in transmissibility arising from seasonal variability can have surprisingly large effects on the overall dynamics and intensity of transmission, matching observed epidemic patterns remarkably well (Altizer et al., 2006; Dushoff, Plotkin, Levin, & Earn, 2004). High mean relative humidity and temperature are frequently identified as the key weather elements that drive influenza epidemics. However, as noted by Shaman and Kohn (2009), absolute humidity predicts influenza virus transmission and survival much better than relative humidity. Increased transmission occurs when humidity and temperature decrease because a rapid shift toward colder, drier air favors longer virus survival in water droplets and, as temperatures drop, people spend increasing time indoors, where crowding and poor ventilation exacerbate transmission risk (Chattopadhyay, Kiciman, Elliott, Shaman, & Rzhetsky, 2018). In the US, these epidemic conditions are best met in the southeastern states and transmission moves from coastal areas toward denser inland areas, following local travel patterns. This suggests that COVID-19 transmission is likely to increase again in the late autumn and winter of 2020 to 2021.

The seasonality of respiratory virus transmission is less clear for tropical and sub-tropical regions, partly because seasonality is less defined than in temperate regions, and partly because of the lack of longitudinal, fine-grained data. We can look at the state of influenza forecasting to understand the challenges of modeling epidemic seasonality in general but particularly for tropical countries. Most reliable influenza forecasting is conducted for the US and other temperate countries, while forecasts for tropical and sub-tropical regions are less accurate and less frequently done, with the exception of Hong Kong and Singapore, whose public health infrastructure and case surveillance and reporting abilities are not broadly representative of the socio-economic conditions of many tropical countries.

Unlike the clear seasonal waves observed in temperate regions, in tropical and subtropical countries, influenza peaks occur sporadically throughout the year. Links between influenza and tropical weather patterns have been inconsistently identified, with rainy season being the most consistent correlate (Shek & Lee, 2003; Viboud, Alonso, & Simonsen, 2006), but how precipitation drives viral circulation in tropical regions remains unclear, with some studies finding increased viral transmission during rainy and humid seasons and others not (Yang et al., 2018). Variance in data quality and availability restrict meaningful multi-national modeling and a lot of the disparities in findings come from country-specific models.

5 | DISCUSSION

In this paper, we have tried to summarize an already dauntingly large literature on the relevance of formal models of transmission dynamics to understanding the COVID-19 pandemic. Models are powerful tools for understand complex phenomena. Most of the simple models we have discussed in this review are best viewed as producing counterfactuals. When we say that the final size of an epidemic where $R_0 = 2.5$ is 89.3% of the total population, there is an implicit if-nothing-changes caveat. This leads to a common problem with the interpretation of models, namely, that they don't make good predictions. Many models, particularly early in an epidemic are not attempting to make predictions per se. They are trying to present counterfactuals for various possible scenarios. This said, simple models can be surprisingly robust over the long-term, as highlighted by the recent preprint by Carletti, Fanelli, and Piazza (2020) and made for long-term demographic forecasting by Goldstein and Stecklov (2002).

R_0 must not be naturalized as a quality of the pathogen. It *always* encapsulates both pathogen and host behavior and social structure (Arthur et al., 2017), even if only in a rudimentary way. Because every transmission event for an infectious disease ultimately has a social cause, the social behavior of the hosts must be incorporated into models of disease transmission and measures of epidemic thresholds. This social behavior may be as simple as assuming random encounters such that the product of the densities of susceptible and infectious individuals in the population, as in the simple SIR model. However, as we have outlined above, models can incorporate structure ranging from simply subdividing the number of compartments in an SIR-like model (as in the inequality example depicted in Figure 4) to the explicit description of fine-grained social structure as in the network examples. There remains a great deal of room for improved input for the best way to incorporate structure and behavior from anthropologists and human biologists.

We have sought to engage the human biology community in this review by framing the results in terms of important research areas within human biology and biological anthropology. In particular, we have focused on results relating to population heterogeneity, health inequalities, and differential environmental exposures in particular. We hope that this review can serve as a launching point for human biologists, with their unique focus on the holistic bio-social causation and integration with evolutionary explanation, into work on the dynamics of respiratory (and other) infectious diseases of people, including COVID-19. There remain major challenges to incorporating both human behavior and social structure into epidemic models (Arthur et al., 2017; Funk et al., 2015), and the COVID-19 pandemic has highlighted the large stakes associated with developing adequate models. We believe that human biologists can play an important role in meeting these challenges.

AUTHOR CONTRIBUTIONS

James Jones: Conceptualization; formal analysis; funding acquisition; methodology; visualization; writing-original draft. **Ashley Hazel:** Conceptualization; writing-original draft; writing-review and editing. **Zack Almquist:** Formal analysis; methodology; visualization; writing-original draft; writing-review and editing.

ORCID

James Holland Jones https://orcid.org/0000-0003-1680-6757
Ashley Hazel https://orcid.org/0000-0001-7680-5460
Zack Almquist https://orcid.org/0000-0002-1967-123X

REFERENCES

Altizer, S., Dobson, A., Hosseini, P., Hudson, P., Pascual, M., & Rohani, P. (2006). Seasonality and the dynamics of infectious diseases. *Ecology Letters*, *9*(4), 467–484.

Anderson, R., Medley, G., May, R. M., & Johnson, A. (1986). A preliminary study of the transmission dynamics of the Human Immunodeficiency Virus (HIV), the causative agent of AIDS. *IMA Journal of Mathematics Applied in Medicine and Biology*, *3*, 229–263.

Anderson, R. M., & May, R. M. (1991). *Infectious diseases of humans: Dynamics and control*. Oxford: Oxford University Press.

Armelagos, G. J., & Dewey, J. R. (1970). Evolutionary response to human infectious diseases. *Bioscience*, *20*(5), 271–275.

Arthur, R. F., Gurley, E. S., Salje, H., Bloomfield, L. S. P., & Jones, J. H. (2017). Contact structure, mobility, environmental impact and behaviour: the importance of social forces to infectious disease dynamics and disease ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372* (1719), 20160454.

Bharti, N., Tatem, A. J., Ferrari, M. J., Grais, R. F., Djibo, A., & Grenfell, B. T. (2011). Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. *Science*, *334*(6061), 1424–1427.

Bjornstad, O. N. (2018). *Epidemics: Models and data using R*. Cham, Switzerland: Springer.

Black, F. L. (1975). Infectious-diseases in primitive societies. *Science*, *187*(4176), 515–518.

Bonds, M. H., Keenan, D. C., Rohani, P., & Sachs, J. D. (2010). Poverty trap formed by the ecology of infectious diseases.

Proceedings of the Royal Society of London B: Biological Sciences, *277*(1685), 1185–1192.

Brandt, E. B., Beck, A. F., & Mersha, T. B. (2020). Air pollution, racial disparities, and COVID-19 mortality. *Journal of Allergy and Clinical Immunology*, *146*, 61–63.

Buckee, C. O., Balsari, S., Chan, J., Crosas, M., Dominici, F., Gasser, U., ... Schroeder, A. (2020). Aggregated mobility data could help fight COVID-19. *Science*, *368*(6487), 145–146.

Buckee, C. O., Tatem, A. J., & Metcalf, C. J. E. (2017). Seasonal population movements and the surveillance and control of infectious diseases. *Trends in Parasitology*, *33*(1), 10–20.

Carletti, T., D. Fanelli, and F. Piazza (2020). COVID-19: The unreasonable effectiveness of simple models.

Chattopadhyay, I., Kiciman, E., Elliott, J. W., Shaman, J. L., & Rzhetsky, A. (2018). Conjunction of factors triggering waves of seasonal influenza. *eLife*, *7*, e30756.

Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., ... Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, *368*(6489), 395−+.

Cockburn, T. A. (1971). Infectious diseases in ancient populations. *Current Anthropology*, *12*(1), 45–62.

Cui, Y., Zhang, Z.-F., Froines, J., Zhao, J., Wang, H., Yu, S.-Z., & Detels, R. (2003). Air pollution and case fatality of SARS in the People's Republic of China: an ecologic study. *Environmental Health*, *2*(1), 15.

Diekmann, O., Heesterbeek, J. A. P., & Metz, J. A. J. (1990). On the definition and the computation of the basic reproduction ratio Ro in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, *28*(4), 365–382.

Dinkel, K. A., Costa, M. E., Kraft, T. S., Stieglitz, J., Cummings, D. K., Gurven, M., ... Trumble, B. C. (2020). Relationship of sanitation, water boiling, and mosquito nets to health biomarkers in a rural subsistence population. *American Journal of Human Biology*, *32*(1).1–13.

Disease Control and Prevention (2020a). Commercial laboratory seroprevalence survey data. Available from https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/commercial-lab-surveys.html.

Disease Control and Prevention (2020b). COVID-19 in racial and ethnic minority groups. Available from https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/racial-ethnic-minorities.html.

Dushoff, J., Plotkin, J. B., Levin, S. A., & Earn, D. J. D. (2004). Dynamical resonance can account for seasonality of influenza epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(48), 16915–16916.

Endo, A. C., Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Abbott, S., Kucharski, A., & Funk, S. (2020). Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Research*, *5*(67).1–10.

Fattorini, D., & Regoli, F. (2020). Role of the chronic air pollution levels in the COVID-19 outbreak risk in Italy. *Environmental Pollution*, *264*, 114732.

Feld, S. L. (1991). Why your friends have more friends than you do. *American Journal of Sociology*, *96*, 1464–1477.

Funk, S., Bansal, S., Bauch, C. T., Eames, K. T. D., Edmunds, W. J., Galvani, A. P., & Klepac, P. (2015). Nine challenges in incorporating the dynamics of behaviour in infectious diseases models. *Epidemics*, *10*, 21–25.

Gile, K. J., & Handcock, M. S. (2010). Respondent-driven sampling: An assessment of current methodology. *Sociological Methodology*, *40*(1), 285–327.

Gjoka, M., M. Kurant, C. T. Butts, and A. Markopoulou (2010). *Walking in Facebook: A case study of unbiased sampling of OSNs*. 2010 *Proceedings IEEE Infocom*, IEEE, pp. 1–9.

Goldstein, J. R., & Stecklov, G. (2002). Long-range population projections made simple. *Population and Development Review*, *28*(1), 121–141.

Goodreau, S., E. Pollock, J. Birnbaum, D. Hamilton, and M. M. (2020). Can't i please just visit one friend? Visualizing social distancing networks in the era of Covid-19. Available from http://statnet.org/COVID-JustOneFriend/.

Grantz, K., C. Metcalf, and J. Lessler (2020). Dispersion vs. control. Available from https://hopkinsidd.github.io/nCoV-Sandbox/DispersionExploration.html.

Gross, C. P., U. R. Essien, S. Pasha, J. R. Gross, S.-Y. Wang, and M. Nunez-Smith (2020). Racial and ethnic disparities in population level COVID-19 mortality. *medRxiv*.

Gurven, M., Kaplan, H., & Supa, A. Z. (2007). Mortality experience of Tsimane amerindians of Bolivia: Regional variation and temporal trends. *American Journal of Human Biology*, *19*(3), 376–398.

Handcock, M. S., & Jones, J. (2006). Interval estimates for epidemic thresholds in two-sex network models. *Theoretical Population Biology*, *70*, 125–134.

Hatchett, R. J., Mecher, C. E., & Lipsitch, M. (2007). Public health interventions and epidemic intensity during the 1918 influenza pandemic. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(18), 7582–7587.

Hazel, A., & Jones, J. H. (2018). Remoteness influences access to sexual partners and drives patterns of viral sexually transmitted infection prevalence among nomadic pastoralists. *PLoS One*, *13*(1), e0191168.

Hazel, A., Marino, S., & Simon, C. (2015). An anthropologically based model of the impact of asymptomatic cases on the spread of *Neisseria gonorrhoeae*. *Journal of the Royal Society Interface*, *12*(106), 20150067.

Hazel, A., Ponnaluri-Wears, S., Davis, G. S., Low, B. S., & Foxman, B. (2014). High prevalence of *Neisseria gonorrhoeae* in a remote, undertreated population of Namibian pastoralists. *Epidemiology and Infection*, *142*(11), 2422–2432.

Heesterbeek, H., Anderson, R. M., Andreasen, V., Bansal, S., De Angelis, D., Dye, C., ... Viboud, C. (2015). Modeling infectious disease dynamics in the complex landscape of global health. *Science*, *347*(6227), aaa4339.

Heesterbeek, J. A. P. (2002). A brief history of R0 and a recipe for its calculation. *Acta Biotheoretica*, *50*(3), 189–204.

Hellewell, J., Abbott, S., Gimma, A., Bosse, N. I., Jarvis, C. I., Russell, T. W., ... Eggo, R. M. (2020). Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*, *8*(4), e488–e496.

Hethcote, H. W., & Yorke, J. A. (1984). *Gonorrhea: Transmission dynamics and control. Lecture Notes in Biomathematics* (Vol. 56). Berlin: Springer-Verlag.

Houldcroft, C. J., Ramond, J. B., Rifkin, R. F., & Underdown, S. J. (2017). Migrating microbes: what pathogens can tell us about

population movements and human evolution. *Annals of Human Biology*, *44*(5), 397–407.

Houldcroft, C. J., & Underdown, S. J. (2016). Neanderthal genomics suggests a Pleistocene time frame for the first epidemiologic transition. *American Journal of Physical Anthropology*, *160*(3), 379–388.

Jones, J. (2020). Notes on R0. Available from https://eehh-stanford.github.io/gceid/Jones-R0-notes2020. pdf.

Karlsson, E. K., Kwiatkowski, D. P., & Sabeti, P. C. (2014). Natural selection and infectious disease in human populations. *Nature Reviews Genetics*, *15*(6), 379–393.

Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H., & Lipsitch, M. (2020). Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*, eabb5793.860–868.

Kramer, S. C., & Shaman, J. (2019). Development and validation of influenza forecasting for 64 temperate and tropical countries. *PLoS Computational Biology*, *15*(2), e1006742.

Lai, S., Ruktanonchai, N. W., Zhou, L., Prosper, O., Luo, W., Floyd, J. R., ... Tatem, A. J. (2020). Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature*, *585*, 410–413.

Leclerc, Q., Fuller, N., Knight, L., CMMID COVID-19 Working Group, Funk, S., & Knight, G. (2020). What settings have been linked to SARS-CoV-2 transmission clusters? [version 1; peer review: 1 approved with reservations]. *Wellcome Open Research*, *5*(83), 83.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., ... Feng, Z. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine.*, *382*, 1199–1207.

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., & Shaman, J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*, eabb3221.*368*(6490), 489–493.

Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., & Getz, W. M. (2005). Superspreading and the effect of individual variation on disease emergence. *Nature*, *438*(7066), 355–359.

May, R. M., & Lloyd, A. L. (2001). Infection dynamics on scale-free networks. *Physical Review E*, *64*(6), 066112.

McDade, T. W., Tallman, P. S., Madimenos, F. C., Liebert, M. A., Cepon, T. J., Sugiyama, L. S., & Snodgrass, J. J. (2012). Analysis of variability of high sensitivity C-reactive protein in lowland Ecuador reveals no evidence of chronic low-grade inflammation. *American Journal of Human Biology*, *24*(5), 675–681.

McGrath, J. W. (1988). Social networks of disease spread in the lower Illinois valley: A simulation approach. *American Journal of Physical Anthropology*, *77*(4), 483–496.

Moghadas, S. M., A. Shoukat, M. C. Fitzpatrick, C. R. Wells, P. Sah, A. Pandey, J. D. Sachs, Z. Wang, L. A. Meyers, B. H. Singer, and A. P. Galvani (2020). *Projecting hospital utilization during the COVID-19 outbreaks in the United States*. Proceedings of the National Academy of Sciences, 202004064.

Morris, M. (1991). A log-linear modeling framework for selective mixing. *Mathematical Biosciences*, *107*(2), 349–377.

Morris, M. (1993). Epidemiology and social networks: Modeling structured diffusion. *Sociological Methods & Research*, *22*(1), 99–126.

Munster, V. J., Koopmans, M., van Doremalen, N., van Riel, D., & de Wit, E. (2020). A novel coronavirus emerging in China: Key questions for impact assessment. *New England Journal of Medicine*, *382*, 692–694.

Nakazawa, M., Ohmae, H., Ishii, A., & Leafasia, J. (1998). Malaria infection and human behavioral factors: A stochastic model analysis for direct observation data in the Solomon Islands. *American Journal of Human Biology*, *10*(6), 781–789.

Nold, A. (1980). Heterogeneity in disease-transmission modeling. *Mathematical Biosciences*, *52*(3), 227–240.

O'Neil, C. A., & Sattenspiel, L. (2010). Agent-based modeling of the spread of the 1918-1919 flu in three Canadian fur trading communities. *American Journal of Human Biology*, *22*(6), 757–767.

Prem, K., Liu, Y., Russell, T. W., Kucharski, A. J., Eggo, R. M., Davies, N., ... Klepac, P. (2020). The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *The Lancet Public Health*, *5*(5), e261–e270.

Richardson, E. T., M. M. Malik, W. A. Darity, A. K. Mullen, M. Malik, A. Benton, M. T. Bassett, P. E. Farmer, L. Worden, and J. H. Jones (2020). Reparations for Black American descendants of persons enslaved in the U.S. and their estimated impact on SARS-CoV-2 transmission. *medRxiv*.

Rosenberg, E. S., Tesoriero, J. M., Rosenthal, E. M., Chung, R., Barranco, M. A., Styer, L. M., ... Zucker, H. A. (2020). Cumulative incidence and diagnosis of SARS-CoV-2 infection in New York. *Annals of Epidemiology*, *48*, 23–29.

Sachs, J., & Malaney, P. (2002). The economic and social burden of malaria. *Nature*, *415*(6872), 680–685.

Salganik, M. J., & Heckathorn, D. D. (2004). Sampling and estimation in hidden populations using respondent-driven sampling. *Sociological Methodology*, *34*, 193–239.

Sattenspiel, L., Koopman, J., Simon, C., & Jacquez, J. A. (1990). The effects of population-structure on the spread of the HIV infection. *American Journal of Physical Anthropology*, *82*(4), 421–429.

Shaman, J. and M. Galanti (2020). Direct measurement of rates of asymptomatic infection and clinical care-seeking for seasonal coronavirus. *medRxiv*.

Shaman, J., & Kohn, M. (2009). Absolute humidity modulates influenza survival, transmission, and seasonality. *Proceedings of the National Academy of Sciences*, *106*(9), 3243–3248.

Shaman, J., Morita, H., Birger, R., Boyle, M., Comito, D., Lane, B., ... Planet, P. (2018). Asymptomatic summertime shedding of respiratory viruses. *The Journal of Infectious Diseases*, *217*(7), 1074–1077.

Shek, L. P.-C., & Lee, B.-W. (2003). Epidemiology and seasonality of respiratory tract virus infections in the tropics. *Paediatric Respiratory Reviews*, *4*(2), 105–111.

Subramanian, S. V., Belli, P., & Kawachi, I. (2002). The macroeconomic determinants of health. *Annual Review of Public Health*, *23*, 287–302.

Thomas, L. J., P. Huang, F. Yin, X. I. Luo, Z. W. Almquist, J. R. Hipp, and C. T. Butts (2020). Spatial heterogeneity can lead to substantial local variations in COVID-19 timing and severity. arXiv:2005.09850.

Upham, S. (1986). Smallpox and climate in the American Southwest. *American Anthropologist*, *88*(1), 115–128.

Viboud, C., Alonso, W. J., & Simonsen, L. (2006). Influenza in tropical regions. *PLoS Medicine*, *3*(4), e89.

Wesolowski, A., Buckee, C. O., Engø-Monsen, K., & Metcalf, C. J. E. (2016). Connecting mobility to infectious diseases: The promise and limits of mobile phone data. *The Journal of Infectious Diseases*, *214*(Suppl 4), S414–S420.

Wesolowski, A., Eagle, N., Tatem, A. J., Smith, D. L., Noor, A. M., Snow, R. W., & Buckee, C. O. (2012). Quantifying the impact of human mobility on malaria. *Science*, *338*(6104), 267–270.

Wesolowski, A., Metcalf, C. J. E., Eagle, N., Kombich, J., Grenfell, B. T., Bjørnstad, O. N., … Buckee, C. O. (2015). Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. *Proceedings of the National Academy of Sciences*, *112*(35), 11114–11119.

Wesolowski, A., zu Erbach-Schoenberg, E., Tatem, A. J., Lourenco, C., Viboud, C., Charu, V., … Metcalf, C. J. E. (2017). Multinational patterns of seasonal asymmetry in human movement influence infectious disease dynamics. *Nature Communications*, *8*, 2069.

Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *The Lancet*, *395*, 689–697.

Wu, X., R. C. Nethery, B. M. Sabath, D. Braun, and F. Dominici (2020). Exposure to air pollution and COVID-19 mortality in the United States. *medRxiv*.

Yang, W., Cummings, M. J., Bakamutumaho, B., Kayiwa, J., Owor, N., Namagambo, B., … Shaman, J. (2018). Dynamics of influenza in tropical Africa: Temperature, humidity, and co-circulating (sub) types. *Influenza and Other Respiratory Viruses*, *12*(4), 446–456.

Zhang, J., Litvinova, M., Wang, W., Wang, Y., Deng, X., Chen, X., … Yu, H. (2020). Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *The Lancet Infectious Diseases*, *20*, 793–802.