QUESTIONING ESTIMATES OF NATURAL PANDEMIC RISK

David Manheim

The central argument in this article is that the probability of very large natural pandemics is more uncertain than either previous analyses or the historical record suggest. In public health and health security analyses, global catastrophic biological risks (GCBRs) have the potential to cause “sudden, extraordinary, widespread disaster,” with “tens to hundreds of millions of fatalities.” Recent analyses focusing on extreme events presume that the most extreme natural events are less likely than artificial sources of GCBRs and should receive proportionately less attention. These earlier analyses relied on an informal Bayesian analysis of naturally occurring GCBRs in the historical record and conclude that the near absence of such events demonstrates that they are rare. This ignores key uncertainties about both selection biases inherent in historical data and underlying causes of the nonstationary risk. The uncertainty is addressed here by first reconsidering the assumptions in earlier Bayesian analyses, then outlining a more complete analysis accounting for several previously omitted factors. Finally, relationships are suggested between available evidence and the uncertain question at hand, allowing more rigorous future estimates.

Keywords: Risk estimates, Pandemics, Existential risk, Global catastrophic biological risk

“Humanity has survived what we might call natural existential risks for hundreds of thousands of years; thus it is *prima facie* unlikely that any of them will do us in within the next hundred.”¹

—Nick Bostrom

REVISITING THE CASE NOT TO WORRY

BOSTROM IS CONCERNED with risks that are far larger than those included in most discussions of biosecurity, but the same probabilistic argument has been used to discuss natural pandemics.²⁻⁶ The argument for the rarity of natural global catastrophic biological risks (GCBRs) can be formulated to use the fact of human survival as evidence (in a Bayesian sense) to provide a revised estimate about the probability of natural pandemics. Those interested in mitigating extreme tail risk, such as Jebari,⁷ also assert that

natural pandemics are less worrying than other risks, with implications for research agendas and funding targeted at mitigating these extreme risks. By outlining the analysis more exactly, we can see several ways in which this argument is prone to overstatement and provide preliminary work needed for further analysis.

Bayesian analyses are those in which a model relating the probabilities, evidence, and prior beliefs is specified, and then evidence is gathered in order to revise, or “update,” the prior belief. An updated estimate, or “posterior probability,” can then be computed via the use of Bayes’s theorem.

Bostrom’s quote notes that humanity’s survival itself is the evidence, and, given some prior belief about the likelihood of a humanity-extminating natural pandemic event, we can use the evidence to update our estimate of the likelihood. This analysis requires a probability model that includes both the evidence and the various probabilities in

question. As Laskey notes, when there is “uncertainty about a model’s structure,”⁸ an analysis must account for this. Thus, Ord outlines how model uncertainty in analyses of large-scale risks reduces our confidence in the estimate.⁹

Next, an analysis involving risks to humanity must deal with the idea of anthropic bias, introduced by Bostrom,¹⁰ where the fact that someone is asking the question is itself relevant to the probability, because the existence of the observer is a precondition. For example, it might be claimed that evolution of intelligence is likely because it occurred on Earth, the only planet extensively investigated. This ignores that the evidence requires an observer and is therefore biased in a statistical sense. Anthropic shadows exist when an anthropic bias leads to a situation in which “the frequencies of catastrophes that destroy or are otherwise incompatible with the existence of observers are systematically underestimated.”^{11(p0000)}

Bostrom’s analysis concludes that, even after accounting for model uncertainty and anthropic bias, the risk of natural GCBRs is still low. A fuller analysis, one which accounts for the nature of the historical evidence and the type of model uncertainties that exist, shows why this conclusion is premature.

DEFINING THE RISK

To build a Bayesian model as discussed above, it is necessary to identify what qualifies as a GCBR. Yassif suggests GCBRs “could permanently alter the trajectory of human civilization in a way that would undermine its long-term potential or, in the most extreme case, threaten its survival.”^{12(p000)} Our argument requires further restricting the definition to events caused by a naturally occurring pathogen. This definition is sufficiently broad, but because the ways in which a naturally occurring biological event could alter the trajectory of humankind are manifold, it is difficult to capture analytically.

For example, the Spanish flu of 1918 was too small in magnitude to directly alter the course of human civilization. Despite this, it has been argued that the death toll due to the flu, combined with the fact that US President Wilson and others were incapacitated by influenza before and during the 1918 Paris peace conference, led to a peace deal which itself may have contributed to the outbreak of World War II.¹³ Such indirect effects are historically important but outside the scope of an epidemiologic analysis. For this reason, we confine the present model to direct effects of a disease.

According to experts surveyed by Sandberg and Bostrom, natural pandemics were thought to be twice as likely as nuclear war to kill 1 million people, but only half as likely to kill 1 billion.¹⁴ This type of expert elicitation is limited,^{9,15} but only the latter case is even plausibly sufficient to directly affect the course of civilization. Our model will therefore address the question of the probability that a

naturally occurring disease kills or otherwise incapacitates more than 1 billion people. This narrower question is still quite complex, since a natural GCBR, as defined, requires not only a sufficiently virulent disease, but also specific exposure levels and/or response failures.

DECOMPOSING THE RISK TO UPDATE BELIEFS

The probability distribution of diseases that will occur in a given future timeframe includes GCBRs, subcatastrophic events,¹⁶ less problematic new diseases, and the chance of the emergence of no new diseases. The probability of a GCBR can be understood as the probability of a certain subset of the events in this full joint probability distribution. To restate the question in those terms, we wish to estimate the portion of the joint distribution $p(\text{emergence, disease, impact})$ where more than 1 billion people are disabled or killed. (Note that this analysis conceptually subsumes more detailed analyses appropriate for specific diseases that consider likelihoods of outbreak, exposure, and outcome as separate components and implicitly conditions on disease emergence and characteristics.)

This article will not quantitatively evaluate these probabilities, but understanding the way in which evidence would be used formally enables a more principled understanding of the evidence and the uncertainties involved. In starting a formal analysis, Betancourt recommends beginning with just this type of conceptual analysis, then continues with defining prior understanding and relevant uncertainties in order to update based on data.¹⁷ The procedure advocated by Betancourt requires considering the joint probability distribution between the parameters of interest, such as the disease characteristics and the number of people killed, and whatever data might be available. By specifying conditional probability relationships between various classes of data and the parameters of interest, Bayesian methods can estimate the updated distribution of those parameters based on the data.

We will first consider the rate of emergence of diseases with a given level of impact, $p(\text{emergence}|\text{impact, data})$, where the data are known events of this type. This evidence can also inform $p(\text{disease}|\text{emergence, data})$, the relative rate at which disease with various characteristics and etiologies emerge. Lastly, the risk of a GCBR is based on $p(\text{impact}|\text{disease, emergence})$. While few data are available to directly inform this relationship, models and expert understanding can be used to inform priors and relationship to estimate the probability that a given disease that could emerge becomes a GCBR.

Uncertainties

Before outlining how evidence relates to the rate of GCBRs, it is worth reviewing the relevant general questions and arguments that will be presented. After doing so, the following

sections provide an overview of why the identified uncertainties are important, illustrate the considerations discussed, and then consider specific types of evidence that can be used (Table 1).

Motivating Questions

- What is the historical rate of disease with GCBR potential?
- What is the impact of human behavior on the present rate of emergence?
- What etiologies or characteristics make a disease likely to become a GCBR?
- What is the historical distribution of these disease characteristics?
- How do rapid travel and dense, poor populations magnify the impact from such diseases?
- How likely are high-fatality and high-risk events from a given disease?
- How effective are containment and risk-mitigation regimes?
- How likely are the models used for this analysis to be wrong?

TOWARD A FULLER ANALYSIS OF NATURAL GCBRS

The probability of emergence of a disease with a potential for GCBR impacts, $p(\text{emergence}|\text{impact})$, is challenging to estimate for 3 principal reasons: (1) Models for available historical data must account for both observer effects and selection effects. (2) The rate of emergence of diseases depends heavily on human behavior and is nonconstant. This

requires considering how $p(\text{emergence}|\text{disease})$ has changed and will change over time. And (3) $p(\text{impact}|\text{disease}, \text{emergence})$ also depends on human behavior and public health response, requiring consideration of how public health response changes the risk.

Emergence Rates of GCBRs

As mentioned, $p(\text{emergence}|\text{impact})$ can be considered in part on the basis of observable historical events with known impacts. Doing so properly requires careful analysis of both available data and how those data relate to our question. This second consideration requires explaining observer effects and selection effects, before later discussion of potential data sources for estimating the risk.

Observer Effects and Anthropropic Shadows

Cirkovic, Sandberg, and Bostrom¹¹ do not discuss pandemics in their article introducing anthropic shadows, but the model suggested by Sandberg and colleagues¹⁹ suggests that the lack of near misses may (counter-intuitively) be evidence that such an anthropic shadow exists. Cirkovic et al also note that anthropic bias “can be understood as a form of sampling bias,” and in the present case, other forms of sampling bias issues should also be considered.

Selection Effects, Sampling Biases, and Data-Generating Processes

In addition to anthropic biases, incomplete historical records can create sampling biases. Unlike typical sampling biases that can be addressed by weighted sampling and

Table 1. Evidence and Uncertainties for Estimating Natural GCBR

<i>Evidence</i>	<i>Sources of Uncertainty</i>
No human extinction	Anthropic shadow
	Extirpations may be unobserved and not appear in the fossil record.
	Changing ecological and social conditions
Few near misses	Historical extirpations in humans might now have GCBR potential.
	Mammalian extinctions and extirpations may be more common than appreciated. ¹⁸
Limited historic virality, infectivity of events	Population density, travel patterns, and interaction with animals have changed.
	“Optimal virality” generally reduces virality over time, but novel pathogens have not yet faced selection pressure.
GCBR pathogens are relatively rare	Little evidence has been gathered about the distributions of stealth, acute, and robust pathogens.
	Other GCBR pathogens would be found by health systems early enough to be stopped.
Effectiveness of modern health response	Historical effectiveness lower than current systems, which have improved over time.
	Effectiveness depends on pathogen visibility and international cooperation.

other techniques, the recording of historical infectious diseases depends on the observation process in a way that cannot be accounted for by simple weighting. For this reason, accounting for the data-generating process directly is useful.

As an example of the importance of accounting for the data-generating process, consider the 2011 Tohoku earthquake, which was stronger than many estimates of the maximum possible magnitude. As explained by Kagan and Jackson, “maximum earthquake size is often guessed from the available history of earthquakes, a method known for its significant downward bias.”^{20(p0000)} This methodology relies on observations, without accounting for the (in this case, seismological) underlying process generating those observations. The analyses based on historical data cited by Kagan and Jackson provided maximum size estimates of between 7.7 and 8.5, while analyses combining that historical data with models of the geography and seismology of the zone provided a maximum of 9.6. Tohoku was magnitude 9.1.

Returning to GCBRs, we first consider $p(\text{emergence}|\text{impact, data})$. As an indicative example, we consider the plague, *Yersinia pestis*, which caused both the Justinian plague and later the Black Death in Europe, which each killed around half of the population.²¹ Plague has “only” a 50% to 60% case fatality rate when untreated.²² Other contenders for worst-case diseases include smallpox and cholera. The earlier analysis, which matches the type criticized by Kagan and Jackson for earthquakes, implies that we can find the risk by an empirical distribution of historically observed cases over the years elapsed since the dawn of humanity 200,000 years ago. A more careful analysis must consider the reference class, and the set of diseases, more carefully.

Estimating Emergence of GCBRs with Data

There are several issues with the analysis sketched above. First, the 200,000-year reference class is optimistic. Wolfe and colleagues suggest that the relevant infectious diseases (ie, those easily spread between hosts) require some degree of population concentration, which became common only more recently.²³ Even disregarding this hypothesis, only 7,000 years have elapsed since the earliest written records. Where events occurred, outbreaks leave much less obvious evidence than natural disasters, and records may not have been created, or they may have been lost. Lastly, there is a clear cultural and linguistic bias in historically preserved records.

We do not have fundamental probabilistic models of the types of diseases that can occur, which in practice rules out the earlier mentioned approach to address the downward bias in the estimates used in seismology. The naive statistical estimates can still be adjusted on the basis of the true

data-generating process and the nature of observed events.^{24,25} Due to significant uncertainty about the underlying process, this adjustment is less robust than that used in earthquake modeling, but it can at least correct some problematic assumptions of the naive method. Doing so at least requires a (seemingly heretofore unattempted) analysis of the process by which we would fail to observe historical events of relevance. As a simple example, if we expect only 25% of relevant events since the advent of writing to have been recorded, the rate should be 4 times that found in the sample. (Existential risks and anthropic shadows still must be accounted for, since, by definition, no samples exist.)

Estimating the Distribution of Diseases

In addition to informing our estimate of the rate of emergence, historical data can also give us some indication of which classes of disease might emerge. Even if the impacts of diseases are vastly different, factors influencing disease emergence that arise from (as yet) unchanged biological factors are likely to exist. The etiologies and mechanisms of spread from historical diseases are therefore useful evidence about what new diseases are most likely to emerge naturally. In addition to the historical evidence considered later in this article, biological plausibility, evolutionary pressure, and similar constraints would also be critical in understanding anthropogenic sources of risk.

For this class of analysis, we largely defer to Adalja et al’s recent “inductive, microbe-agnostic analysis of the microbial world... of microorganisms that have potential to cause global catastrophe.”^{26(p00)} As they note, a list-based historical approach is unlikely to uncover new threats, but we suggest it might still be useful for finding the historical rate of emergence of such threats. Their analysis focuses on $p(\text{disease}|\text{impact, emergence})$ and finds that viruses, especially RNA respiratory viruses, are the most likely causes of future GCBRs. They then consider several other less worrying possibilities, including bacteria—especially antibiotic-resistant bacteria—prionic diseases, and protozoa. They also include fungi, regarding which other experts have noted²⁷ the worrisome fact that fungal adaptation to heat may be enhanced over the coming decades by selection pressure due to global warming and natural disasters.²⁸⁻³⁰ While unlikely, one nonhuman case of extinction due to fungi is discussed below, and it seems this potential is worth monitoring.

Adalja et al also consider modes of transmission and other factors that affect Casadevall’s “pathogenic potential,”²⁸ which focuses even more narrowly on $p(\text{fatality}|\text{communicability, pathology, disease, emergence})$. For this reason, pathogenic potential cannot be used to estimate the likelihood of a given pathology evolving or becoming adapted to humans. Another issue is that the work of Adalja

and colleagues addresses the most likely candidates for a GCBR, corresponding to the maxima of $p(\text{disease}|\text{impact})$ and focused on policy and recommendations. For this reason, the health system response in this analysis is an uncertain factor that must be conditioned on, rather than an independent variable to manipulate. The subset of diseases that are likely to have GCBR impacts even after accounting for the responses of affected populations and public health officials are therefore much more critical in this analysis.

WILL DISEASES CAUSE GCBRS?

The set of events that were observed historically is only indirectly related to what would constitute a GCBR now. For example, the plague is quickly identifiable by modern diagnostics, treatable with antibiotics if identified early, and preventable via pest control. For these reasons, it does not present a significant risk of leading to a GCBR as defined earlier. On the other hand, there are reasons to think the rate of emergence and the severity of diseases have been worsened by other recent changes in the social, physical, and biological environments.

Heightened Risk

Inglesby³¹ notes 3 reasons the risk may be higher than historical evidence implies: (1) global travel, which allows more rapid pathogen spread; (2) increased population density, poverty, and the growth of megacities; and (3) more close contact with animal populations due to both densely packed factory-farms and expansion into uninhabited areas lead to higher rates of emergence.³¹ There are also factors that reduce this risk, such as modern sanitation, public health response, and modern treatment. Because the conditions Inglesby mentions have changed and are still changing, any historical estimate only partially informs our estimate of present risk. Specifically, the first and second of Inglesby's triad are reason to expect that pathogens that would normally emerge in small or isolated populations earlier in human history or in animals would have larger or even global impacts for modern humans, while the second and third are reasons that disease emergence is more likely than was true historically.

Extending the analysis to the future, if most new human pathogens are due to encroaching on new territories and encountering extant diseases that have never before affected humans, we could expect the trend toward more diseases to eventually slow and reverse as humanity fills the ecosphere. Similarly, if the critical factor is newly dense concentrations of people in poverty with poor sanitary conditions, as suggested by Moore et al.,³² we would similarly expect this trend to slow and reverse if global poverty continues to drop. If, however, population density and travel frequency are more critical, we may see the emergence of new diseases

accelerate over time, since the trend toward urbanization and interconnection seems likely to continue accelerating.³³ The balance of these impacts is both unclear and critically important.

Public Health

Humanity's ability to identify, respond to, manage, and contain novel pathogens has gotten better over the previous few decades and continues to improve as further lessons are learned. Modern quarantine and isolation methods and safety protocols greatly limit spread, and most diseases have greatly reduced fatality rates when treated.

Because of modern advances, there is a strong reason to think that only unusual cases (discussed below) would pose risk of a GCBR. But there is a twofold caveat: This assumes the system can detect and respond in time, and that the system is effective and geographically sufficiently well distributed. To the extent that broad biosurveillance, rather than disease-specific surveillance, is funded, the system becomes more likely to detect pathogens in time to mitigate GCBR potential. To the extent that international aid for pandemic response is available and response planning is effective, outbreaks can be contained early and effectively.³⁴

Evidence for the effectiveness of global public health response is the record of halting the spread of disease, even if not immediately. For example, the 2017 pneumonic and bubonic plague outbreak in Madagascar infected slightly more than 2,000 people and killed less than 10% of those infected, and the spread was stopped within a few months.³⁵ Lipsitch noted as a counterexample the failure to identify the 2014-15 Ebola epidemic in time to prevent tens of thousands of cases.¹⁶ This failure was tragic, but, if anything, serves as evidence that a well-understood nascent GCBR could not spread without effective containment even after initial response fails. For example, smallpox vaccines are available, and postexposure inoculation is effective at lessening severity.³⁶

One might assume that even if novel, highly virulent diseases were historically common, the risk of GCBRs is now low. If, however, the disease is not identified, or containment is either not attempted or fails, a virulent enough novel pathogen could infect a large portion of humanity well before vaccines become available. Further, medical care that drastically reduces the impact of disease could quickly become unavailable as health systems are overwhelmed with cases. The analysis must therefore consider both what biosurveillance will exist and how well it is integrated and used.³⁷ This suggests a multi-stage model: first, a model of whether health systems will identify a nascent GCBR, then whether political, economic, and health concerns prompt a response sufficient to stop a GCBR once it is identified. This model will form a critical input to any assessment of how risk differs from the historical rate.

Limitations of Public Health

Given the efficacy of modern public health measures and medical interventions, unless there were a significant slacking in the funding of international public health, or a failure to respond appropriately and quickly, we expect them to be successful. In addition to the various clinically severe factors identified by Casadevall²⁸ and by Adalja et al,²⁶ several conditions are necessary to allow a pandemic event to occur without public health response missteps.

The ways in which diseases might become GCBRs despite an exemplary response can be characterized into 1 of 3 main etiologies: (1) robust events, where the pathogen is able to spread despite the fact that large public health and medical responses are attempted; (2) acute events, where the spread is too rapid to be noticed in time to contain the pathogen's spread; or (3) stealth events, where the pathogen is not noticed for long periods after the disease begins spreading, too late to be easily contained. We have seen some events of each type occur, and if we can estimate the rate at which events occur with given characteristics, the categorization may help us begin to characterize the probability of GCBR events based on historical data.

DATA FOR ESTIMATING DISEASE EMERGENCE RATES

Given the above discussion of how data could be used to estimate the probabilities of interest, we can specify what data can be used. We can consider various reference classes on which to base such an estimate: human diseases, diseases affecting mammals, and diseases affecting vertebrates. The historical record for each is both limited and uncertain, and each has a different relationship with the question of interest. Because of reference class problems and data availability, following Wallmann and Williamson, multiple estimates on the basis of different reference classes are appropriate, and they should be combined via a better understanding of mechanisms.^{38,39}

While the most directly relevant reference class is human diseases, looking more widely at disease of mammals, or vertebrates generally, would give a much larger dataset from which to infer disease emergence rates. The difference in how well these estimates compare to the desired class, of course, is critical, but there are reference class problems with inferring current likelihood of disease with historical likelihood as well, since the rate is changing. To anticipate the later discussion, the way in which the emergence of disease relates to characteristics of the population like density and travel can also be better understood when using such comparative data sets. For this reason, our qualitative overview of the types of data available will note these relationships.

Modern Human Diseases

Humanity has records of diseases since the development of modern medicine. The list of emerging diseases of note

includes AIDS and Ebola in the mid- to late 20th century, Nipah virus in 1999, and SARS and MERS-CoV in the 2000s. Given the focus on GCBRs, rather than epidemics generally, we exclude diseases such as Marburg that are unlikely to have pandemic potential—in that case due to the relative rarity and limited geographic footprint of the specific fruit bats that serve as host.

Focusing even more closely on the most recent diseases, Woolhouse and Gaunt's literature review noted 1,399 species of human pathogen and investigated the 87 that have emerged since 1980.⁴⁰ This overview shows an acceleration in identified human diseases, but there is significant uncertainty about how much of this change is due to intensive biosurveillance. In addition, these diseases also all fall far short of GCBR-level impacts, albeit in part due to human response.

Ancient Records and Density

Analyses of modern human diseases, their etiologies, and their impacts are less susceptible to some biases due to the comprehensive data collection, but they are limited by the small sample size. A larger historical sample would include records since the beginning of classical antiquity. This dataset may inform both disease emergence rates and the relationship between changing social and economic factors and disease emergence. Diseases in this historical record would appear if they spread widely in the western world, and provide additional evidence, but this dataset is more susceptible to the observational biases noted above.

Critical for this analysis, these data are also used to claim that population density and agriculture led to a significant increase in infectious disease risk. The records approximately coincide with the emergence of agriculture and population concentration, and the bias created by this coincidence partly undermines the data used to claim that pandemics required these high population densities.^{41,42} Morris suggests that Rome was the first city with a population of more than 1 million people, early in the Common Era,⁴³ and the spread of the Roman empire led to increased travel. Some of the earliest recorded plagues—for example, the Antonine Plague and the Plague of Cyprian—occurred in Rome. However, while smallpox seems to have been the cause of these plagues, its presence in humans significantly predates this,⁴⁴ and it is therefore not clearly a data point in the reference class.

No records of bubonic plague predate the 541 AD Plague of Justinian, and though it is unclear when it first emerged, it may be an example of initial pandemic emergence within the historical period. In that case, it struck Constantinople several centuries later, following a similar population boom, lending credence to the idea that population concentrations are a critical enabling factor.

There are also some non-Western sources of data about severe early human diseases. For example, cholera, which first became a pandemic in the 1800s, was seemingly identified in the Sushruta Samhita⁴⁵ about 600 BCE. There

is also evidence in the Charaka Samhita,⁴⁶ dating from early in the Common Era, of earlier epidemics in India of other types. Shembavnekar's claims of significant population density in early India, matching claims by Herodotus, again support the link between population concentration and disease emergence.⁴⁷

Unavailable Evidence

Given the emergence of several diseases in the historical record as soon as population densities reached a certain level, even if the rate of emergence of diseases was unchanged, new conditions made these diseases more likely to spread. Thus, the lack of diseases in early historical records is even less convincing evidence of their rarity. The "missing disease" hypothesis accords with Roberts's sociological model of ancient diseases,⁴⁸ and, if correct, it changes the implications drawn from not finding such events in the historical record. The critical issue is the percentage of diseases that emerged that we expect to not have seen. This splits into 2 cases: extinct pathogens and former pathogens.

Extinct pathogens would have emerged earlier in history and failed to spread widely due to lack of sufficient population density and travel, likely causing extirpations. Roberts's model implies these events could have been more common in prehistory than recorded, since extirpated populations left no records, and it would be difficult to identify such events based on archaeological evidence. Mythological claims of plagues also lend weak support to the hypothesis that extinct pathogens emerged in local human populations and perhaps died off with the hosts.

Former pathogens would have emerged as virulent diseases, likely to damage the hosts, then become endemic before written human history. In this case, they have been under selection pressure for millennia toward optimal virulence, where pathogens that kill rapidly are out-competed by less virulent strains, as discussed by Messenger and colleagues.¹⁸ Some evidence of this exists in the form of normal or transient symbiotic microbiota and relatively benign parasites. Humans have co-evolved with these pathogens,⁴⁹ and modern humans are therefore adapted to surviving the modern versions of perhaps once–highly lethal pathogens. These emergence events could, it seems, have led to GCBR events in a different cultural or sociological environment. Some evidence for this thesis, and the absence of the diseases from historical records, is the lack of historical reference to several hepatitides. Genetic evidence is suggestive of fairly recent origins in humans, with unclear initial virulence. If it did emerge within written history, no records remain. This also reinforces the impression that we should not expect to find early historical records of non-catastrophic diseases.

Because novel diseases by definition infect naive populations, the fatality rate is plausibly higher than for diseases that co-evolved with humans. This would imply both an underestimation of the rate of historical events due to ex-

tinct and former pathogens and a heightened risk in the future. Armelegos and Dewey suggested that pathogens, which have short generations, undergo more selection pressure than humans. At the same time, they note that host changes are significant as well, since people unaffected by certain diseases are much more likely to reproduce.⁵⁰ For example, the introduction of new diseases to the Americas with lower virality in European cultures were much worse in the epidemiologically naive Native populations. The selection is partially genetic, but Robertson argues that cultural changes in norms around visiting the sick were in populations exposed to many infectious diseases.⁵¹

As noted by Adalja and colleagues, much of the potential for GCBRs comes from novel diseases.²⁶ The relationship between records of diseases and their emergence is a feature of the data-generating process that is significant but difficult to quantify. Any missing records, however, will lead to underestimation of the frequency of such events. If unobserved extirpation occurs, the visible record will also overestimate the relative proportion of near-miss extinction events. Several approaches for further investigation suggest themselves, including comparative analysis to animals and a clearer understanding of the relationship among population density, exposure to animals that may serve as reservoirs of new human diseases, and disease emergence.

Mammal and Vertebrate Diseases as Evidence

The historical record for human diseases is limited, but extant historical records for human diseases are still much more comprehensive than those for nonhuman animals. Despite this, as Ray notes, "Animals provide a valuable reference for human extinction."⁵² This is true for several reasons. First, despite the limited historical length, it includes many more species and is therefore an important cross-sectional sample. Second, animal disease data may allow a less biased understanding of the distribution of disease types, as it is not limited by the extirpation-selection bias. Lastly, it mitigates biases of anthropic shadows, at least to the extent cross-species diseases are uncommon.

For diseases that affect mammals, we have sporadic records of animal infections in the Western world going back perhaps as far as 3,000 years.⁵³ We more reliably expect to know of most such events in the past half century, even those affecting local populations. Some of these events may have relationships to population density, hygiene, and other factors that are partly due to human interaction, and the data may be useful for understanding that relationship as well.

The historical record for vertebrates more generally is confined to even more modern events, but a wealth of information is available, including etiologies, how the diseases emerged, how other animals served as reservoirs, and how the diseases were spread. Ray's analysis and partial catalogue of vertebrate extinctions also found several cases

where significant epidemics and extinctions have been caused by diseases that emerged and/or were spread due to human interaction, which lends even further credence to the concern that changes in the biosphere and travel is a critical factor.⁵² Also, 1 case that Ray discusses, chytridiomycosis, has resulted in the extinction of a large variety of amphibian species. This provides a reason to be concerned that anthropic shadows may, in fact, have biased the sample for mammals. It is also worth noting that this multispecies disease is fungal, rather than viral.

WHAT DO HISTORICAL DATA (NOT) TELL US?

While the historical record and comparisons to other mammals are valuable sources of information on the naive emergence rate, there is some evidence that this rate is lower than the modern one, and fundamental reasons can be identified for why we should expect this to be true. As mentioned earlier, Woolhouse and Gaunt found in 2007 that over 6% of all human pathogens have emerged since 1980.⁴⁰ The implications, however, are more ambiguous, because the risk depends on evolutionarily recent changes in the social, physical, and biological environments discussed earlier. These changes are continuing and have overlapping and hard to disentangle effects.

Despite the usefulness of historical evidence, there remains significant uncertainty about the relationship between historical patterns and present risk. In light of these model uncertainties, a model for quantifying these factors should incorporate expert opinion of the relationship among the historical risk, the factors identified by Inglesby, and the classes of risk identified as most problematic. These may be supplemented by evidence provided by historical changes, such as Woolhouse and Gaunt's noted acceleration of disease occurrence.

The resulting model would provide probability distributions over diseases and characteristics and use the multistage model mentioned earlier to find the likelihood a disease would cause a GCBR, either because it remained undetected or was not stopped by modern public health systems. Models and expert opinion about how these situations would resolve are not data in the earlier sense, but they enable constraining the distribution in a Bayesian model. Though these questions are complex, they can be addressed using well understood epidemiologic models of disease progression and pathogenicity, particularly Casadevall's model of pathogenic potential discussed earlier²⁸ and various other models and non-model tools.⁵⁴

CONCLUSION

This article has shown that the observed historical rate is a useful guide to the present risk of natural pandemics only after

adjusting for several difficult to quantify factors. The historical record is limited because of both anthropic factors and other observational selection biases. Historical data do provide a baseline for assessment, and data about nonhuman diseases can provide evidence about both the overall risk of disease emergence and the degree to which selection and anthropic concerns bias the historical evidence for humans.

Still, any assessment of current risk using historical evidence must account for the changing rates of pathogen emergence due to increased mobility and other human factors. While these may promote disease emergence, there are also modern public health technologies that make such events less likely, and the tension between these factors is worth further investigation.

Earlier analyses are likely correct that the risk of natural GCBRs is low, especially when compared to growing risks from anthropogenic sources like biological warfare. Despite this, the uncertainties in those estimates due to observational biases, uncertainty about the changing nature of the risk, and the relative likelihood of especially dangerous etiologies are worth considering seriously because of the enormous potential negative impact of a GCBR. As the global health system becomes more capable of responding to outbreaks and pandemics, a better understanding of what events are most likely to occur will also be valuable. Reducing uncertainty around which types of threatening pathogens are more or less likely is therefore potentially significant, and this discussion of relevant uncertainties and evidence complements ongoing work on better preparing for and responding to potential crises.

REFERENCES

1. Bostrom N. Existential risk prevention as global priority. *Glob Policy* 2013;4(1):15-31.
2. Schoch-Spana M, Cicero A, Adalja A, et al. Global catastrophic biological risks: toward a working definition. *Health Secur* 2017;15(4):323-328.
3. Tonn B, Stiefel D. Evaluating methods for estimating existential risks. *Risk Anal* 2013;33(10):1772-1787.
4. Inglesby TV, Editor. Special Feature: Global catastrophic biological risks. *Health Secur* 2017;15(4).
5. Millett P, Snyder-Beattie A. Human agency and global catastrophic biorisks. *Health Secur* 2017;15(4):335-336.
6. Kambouris ME, Manoussopoulos Y, Kantzanou M, et al. Rebooting bioresilience: a multi-omics approach to tackle global catastrophic biological risks and next-generation biothreats. *OMICS* 2018;22(1):35-51.
7. Jebbari K. Existential risks: exploring a robust risk reduction strategy. *Sci Eng Ethics* 2015;21(3):541-554.
8. Laskey K. Model uncertainty: theory and practical implications. *IEEE Trans Syst Man Cybern A Syst Hum* 1996; 26(3):340-348.
9. Ord T, Hillerbrand R, Sandberg A. Probing the improbable: methodological challenges for risks with low probabilities and high stakes. *J Risk Res* 2010;13:191-205.
10. Bostrom N. *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. Abingdon, UK: Routledge; 2010.

11. Cirkovic MM, Sandberg A, Bostrom N. Anthropoc shadow: observation selection effects and human extinction risks. *Risk Anal* 2010;30(10):1495-1506.
12. Yassif J. Reducing global catastrophic biological risks. *Health Secur* 2017;15(4):329-330.
13. Crosby AW. *America's Forgotten Pandemic: The Influenza of 1918*. Cambridge: Cambridge University Press; 2003.
14. Sandberg A, Bostrom N. *Global Catastrophic Risks Survey*. Technical Report #2008-1, Future of Humanity Institute, Oxford University; 2008:1-5.
15. Tetlock PE, Gardner D. *Superforecasting: The Art and Science of Prediction*. New York: Crown Publishing; 2015.
16. Lipsitch M. If a global catastrophic biological risk materializes, at what stage will we recognize it? *Health Secur* 2017; 15(4):331-334.
17. Betancourt M. A principled Bayesian workflow. June 2018. https://betanalpha.github.io/assets/case_studies/principled_bayesian_workflow.html. Accessed October 31, 2018.
18. Messenger SL, Molineux IJ, Bull JJ. Virulence evolution in a virus obeys a trade-off. *Proc Biol Sci* 1999;266(1417):397-404.
19. Sandberg A, Snyder-Beattie A, Armstrong S. Nuclear war near misses and anthropic shadows. Manuscript in preparation, 2018.
20. Kagan YY, Jackson DD. Tohoku earthquake: a surprise? *Bulletin of the Seismological Society of America* 2013;103(2B): 1181-1194.
21. Christakos G, Olea RA, Yu HL. Recent results on the spatiotemporal modelling and comparative analysis of Black Death and bubonic plague epidemics. *Public Health* 2007; 121(9):700-720.
22. World Health Organization. Plague. <http://www.who.int/ith/diseases/plague/en/> Accessed October 31, 2018.
23. Wolfe ND, Dunavan CP, Diamond J. Origins of major human infectious diseases. *Nature* 2007;447(7142):279-283.
24. Domowitz I, White H. Misspecified models with dependent observations. *J Econom* 1982;20(1):35-58.
25. Becker WE Jr. *An Online Handbook for the Use of Contemporary Econometrics in Economic Education Research*. Module One, Part One: Data Generating Processes. October 2009. <https://www.aeaweb.org/about-aea/committees/economic-education/econometrics-training-modules>. Accessed October 31, 2018.
26. Adalja A, Watson M, Toner E, Cicero A, Inglesby T. *The Characteristics of Pandemic Pathogens*. Baltimore, MD: Johns Hopkins Center for Health Security; 2018. <http://www.centerforhealthsecurity.org/our-work/publications/the-characteristics-of-pandemic-pathogens>. Accessed October 31, 2018.
27. Garcia-Solache MA, Casadevall A. Global warming will bring new fungal diseases for mammals. *MBio* 2010;1(1):e00061-10.
28. Casadevall A. The pathogenic potential of a microbe. *mSphere* 2017;2(1):e00015-17.
29. Manheim D. Eliciting evaluations of existential risk from infectious disease. Unpublished; 2018. <https://www.openphilanthropy.org/focus/global-catastrophic-risks/biosecurity/david-manheim-research-existential-risk>. Accessed October 31, 2018.
30. Benedict K, Park BJ. Invasive fungal infections after natural disasters. *Emerg Infect Dis* 2014;20(3):349-355.
31. Wiblin R. The careers and policies that can prevent global catastrophic biological risks, according to world-leading health security expert Dr. Inglesby. Audio blog post. The 80,000 Hours Podcast with Robert Wiblin. April 18, 2018.
32. Moore M, Gould P, Keary BS. Global urbanization and impact on health. *Int J Hyg Environ Health* 2003;206(4-5): 269-278.
33. United Nations Environment Programme. GEO: Global Environment Outlook-5: Environment for the future we want. Malta: Progress Press; 2012.
34. Ross AG, Crowe SM, Tyndall MW. Planning for the next global pandemic. *Int J Infect Dis* 2015;38:89-94.
35. World Health Organization. Plague – Madagascar. November 27, 2017. <http://www.who.int/csr/don/27-november-2017-plague-madagascar/en/>. Accessed October 31, 2018.
36. Barbera J, Macintyre A, Gostin L, et al. Large-scale quarantine following biological terrorism in the United States: scientific examination, logistic and legal limits, and possible consequences. *JAMA* 2001;286(21):2711-2717.
37. Moore M, Fisher G, Stevens C. Toward integrated DoD biosurveillance: assessment and opportunities. *Rand Health Q* 2014;3(4):9.
38. Hajek A. The reference class problem is your problem too. *Synthese* 2007;156(3):563-585.
39. Wallmann C, Williamson J. 2017. Four approaches to the reference class problem. In: Hofer-Szabo G, Wronski L, eds. *Making It Formally Explicit: Probability, Causality and Indeterminism*. Springer; 2017:61-81.
40. Woolhouse M, Gaunt E. Ecological origins of novel human pathogens. *Crit Rev Microbiol* 2007;33(4):231-242.
41. Tatem AJ, Rogers DJ, Hay SI. Global transport networks and infectious disease spread. *Adv Parasitol* 2006;62:293-343.
42. Morse SS. Factors in the emergence of infectious diseases. In: Price-Smith AT, ed. *Plagues and Politics: Infectious Disease and International Policy*. London: Palgrave Macmillan; 2001:8-26.
43. Morris I. Social Development. Stanford University, October 2010.
44. Hopkins DR. *Princes and Peasants: Smallpox in History*. Chicago: University of Chicago Press; 1985.
45. Sushruta Samhita. c. 7th or 6th century BCE. English Translation, Bhashagratna, Kunja Lal, ed. *An English Translation of the Sushruta Samhita: Based on Original Sanskrit Text, with a Full and Comprehensive Introd., Additional Texts, Different Readings, Notes, Comparative Views, Index, Glossary and Plates*. Vol. 30. Chowkhamba Sanskrit Series Office; 1963
46. Charaka-Samhita. Early Common Era (translated into English). *JAMA* 1894;XXII(16):600-601. <https://jamanetwork.com/journals/jama/fullarticle/445670?resultClick=1>. Accessed October 31, 2018.
47. Shembavnekar KM. The population of ancient India (500 B.C. to 100 A.D.). *Annals of the Bhandarkar Oriental Research Institute* 1952;33(1/4):83-96.
48. Roberts A. The emergence of disease in early world-systems: a theoretical model of world-system and pathogen evolution. PEWS Roundtable on the Evolution of World-Systems; American Sociological Association; August 15, 2010; Atlanta, Georgia. <http://irows.ucr.edu/papers/irows62/irows62.htm>. Accessed October 31, 2018.
49. Bliven KA, Maurelli AT. Evolution of bacterial pathogens within the human host. *Microbiol Spectr* 2016;4(1).
50. Armelagos GJ, Dewey JR. Evolutionary response to human infectious disease. In: Logan MH, Hunt EE, eds. *Health and the Human Condition*. North Scituate, MA: Duxbury Press; 1970:101-106.

51. Robertson RG. *Rotting Face Smallpox and the American Indian*. New York: Caxton; 2001.
52. Ray G. Animal extinctions from disease and implications for human existential risk. Manuscript in preparation; 2018.
53. Fleming G. *Animal Plagues: Their History, Nature, and Prevention. Vol. 1*. London: Chapman and Hall; 1871.
54. Manheim D, Chamberlin M, Osoba OA, Vardavas R, Moore M. *Improving Decision Support for Infectious Disease Prevention and Control*. Santa Monica, CA: Rand Corporation; 2016.
- Manuscript received May 16, 2018;
revisions returned August 8, 2018, and October 8, 2018;
accepted for publication October 14, 2018.*

Address correspondence to:
David Manheim, PhD
8734 First Avenue
Silver Spring, MD 20902

E-mail: davidmanheim@gmail.com