EIGHT CRITICAL QUESTIONS FOR PANDEMIC PREDICTION

Toph Allen, Kris Murray, Kevin J. Olival, and Peter Daszak

Introduction

Like hurricanes or earthquakes, pandemics are rare events that can be extremely devastating, causing substantial mortality and economic damages. However, unlike hurricanes or earthquakes, efforts to identify where pandemics are most likely to originate, and to intervene and preempt their impact, are in their nascence. Here, we review recent advances in disease ecology, virology, and biogeography that move the field towards these goals and pose a series of critical questions that must be addressed to sufficiently improve our predictive capacity. This provides a framework for pandemic prediction that may allow us to better allocate our global resources to mitigate this threat.

Because the majority of recent pandemics are zoonotic in origin, most involving wildlife reservoirs, we consider this group specifically. The emergence of pandemic zoonoses reflects a complex interplay of socioeconomic, ecological, and biological factors and can be thought of as a three-stage process (Morse et al., 2012). Initially, pathogens with pandemic potential exist only in their natural reservoirs. In the first stage, pre-emergence, our encroachment into a reservoir’s natural habitat, often related to changing land use, may bring these pathogens into contact with livestock or humans or otherwise alter the ecological system in which it and its host exist. In the second stage, localized emergence, initial transmission to humans occurs, directly from a wildlife host or via domesticated animals. Some of these events may involve small chains of person-to-person transmission. When a pathogen achieves sustained person-to-person transmission, the right confluence of circumstances can lead to pandemic emergence, ultimately with large outbreaks propelled internationally by the movement of people and disease vectors.

Each of these stages is itself driven by a plethora of socioeconomic, ecological, and biological factors (e.g., change in land use, migration, agricultural intensification) that alter pathogen dynamics and expose human populations to increasing risk of zoonotic disease emergence, amplification, and spread. It follows that to predict and pre-empt pandemics, we must improve our understanding of how these factors drive increased risk of each stage of the pandemic process (Morse et al., 2012). The complexity of these processes is daunting, but the interplay of ecology, demography, virology, and biology provides a wide range of new tools and approaches that can be used in pandemic prediction and prevention.

24 EcoHealth Alliance, 460 West 34th Street, New York, USA.
These include strategies to analyze prior outbreaks, model future trends in pandemic drivers, conduct targeted surveillance in wildlife and human populations, and probe the depth of the zoonotic “pool” from which novel EIDs arise. Here we review some of these by posing eight critical questions for pandemic prediction.

**Eight Critical Questions for Pandemic Prediction**

*Are Emerging Infectious Diseases (EIDs) Really on the Rise?*

The literature on emerging infectious diseases, and concern among policymakers and the public, has grown substantially in recent years (IOM, 1992, 2003). Does this reflect a public health threat that is also growing, or is this trend driven by increased surveillance, or simply better reporting of outbreaks as they occur? To test this, we expanded and updated a database of all known emerging infectious disease, first collated by Mark Woolhouse’s group (Taylor et al., 2001). We focused on “EID events,” which we defined as “the first temporal emergence of a pathogen in a human population . . . related to the increase in distribution, increase in incidence or increase in virulence or other factor which led to that pathogen being classed as an emerging disease” (Jones et al., 2008). For each event, we collected data on location, time, and host and/or vector, as well as on associated ecological, biological, and sociodemographic drivers of disease emergence, and performed a number of temporal and spatial regression analyses.

Our analyses showed that the number of EID events has increased over time, peaking in the 1980–1990 decade. This peak was associated with increased susceptibility to infection due to the HIV/AIDS pandemic. Like Taylor et al. (2001), we found that zoonoses comprised the majority of EID events (60.3 percent), and that almost 71.8 percent of zoonotic EIDs were from wildlife (43.3 percent of all EID events). Furthermore, zoonoses from wildlife were increasing as a proportion of all EID events—in the last decade analyzed (1990–2000), 52.0 percent of EID events were zoonoses with known a wildlife origin. We attempted to correct for increasing infectious disease reporting effort over time by including in our regression model the number of articles published in the *Journal of Infectious Diseases* (which gives a crude measure of research effort for infectious diseases generally, not just EIDs) for each decade as an offset. Controlling for reporting effort gave further support to the conclusions that EID events are becoming more common, that zoonoses comprise the majority of EID events, and that zoonoses are rising significantly faster as a proportion of all EID events.

*Are There Predictable Patterns to Disease Emergence?*

The first step in predicting a biological phenomenon is to look for patterns that underlie previous events. This approach underpins hurricane forecasting and the identification of earthquake zones, and is a logical strategy for pandemic...
prediction. Both hurricane forecasting and the identification of earthquake zones look to the underlying drivers of these phenomena to identify patterns. We used a similar approach for disease emergence, focusing on the hypothesized drivers of zoonotic disease emergence. We assigned geographic coordinates to EID events and, using a logistic regression, tested associations between subsets of EID events and a small selection of hypothesized drivers. We found that drug-resistant and vector-borne pathogens, and zoonoses with wildlife and non-wildlife origins, differed in their global patterning and in their associations with different drivers. In particular, all categories of EID events were strongly associated with human population density, which we have suggested “. . . supports previous hypotheses that disease emergence is largely a product of anthropogenic and demographic changes. . . .” Human population growth—taken as a broad proxy for change in socio-economic factors—predicts zoonoses from non-wildlife and the emergence of drug-resistant pathogens. However, zoonoses from wildlife are alone in their association with wildlife host species richness—patterns of wildlife diversity. The overall predicted risk from different categories was differentially distributed across the globe. For instance, wildlife zoonoses and vector-borne pathogens were more likely to have originated in lower-latitude, developing countries (Jones et al., 2008) (Figure A7-1). Describing these patterns provides the first step towards pandemic prediction: predictive models exist of future trends in socio-economic and demographic drivers, and may be used to derive predictive models of the future trends in disease emergence.

The analyses of Jones et al. (2008) show that EID emergence is driven by socioeconomic as well as biological factors, but they are somewhat preliminary, and substantial gaps remain. For example, what aspects of human population density drive disease emergence? Is it anthropogenic environmental changes (e.g., road building, deforestation, land use change)? Is it increased contact with wildlife, or the perturbation of pathogen transmission dynamics in wildlife? Or do dense human populations simply provide an “amplification zone” that allows more frequent recognition of new EIDs otherwise lost to our analyses? Efforts to tease apart the mechanisms underlying these patterns will involve ecological, virological, and biological disciplines collaborating in exciting new ways (Murray and Daszak, 2013). Finally, it is interesting to note that the Jones et al. (2008) models leave 85 percent of the variation in global patterns of disease emergence unexplained. This emphasizes the magnitude of the problem, sets the bar for future studies, and highlights that efforts to gradually improve the model’s power need to be prioritized if we are to accurately predict the next pandemic.

Where Will the Next Pandemic Originate?

There are significant geopolitical and logistical constraints to pandemic prevention. Newly emerged pathogens often originate in remote areas that are difficult to access, and in resource-constrained countries that cannot afford to
FIGURE A7-1  Map of relative risk of a zoonotic disease of wildlife origin emerging in people. Because almost all prior pandemics, and the majority of emerging infectious diseases, are zoonotic in origin, with the majority of these having a wildlife host, this map acts as a potential basis for future targeted surveillance and the pre-empting of potential pandemics.
systematically identify novel pathogens in their early stages of emergence. Once emerging diseases become pandemic, the large number of cases and wide geographic distribution make response programs costly and complicated by geopolitical issues. Given the finite global capacity for pandemic preparedness, and a limited global budget, can we reconfigure where our resources are spent, based on a scientific understanding of where novel diseases emerge and where our current effort is lacking in relation?

To this end, our previous “EID hotspots” analysis attempts to correct for bias caused by this unequal distribution of surveillance resources and to make recommendations about where surveillance should be increased in response to predicted disease emergence risk. We can draw two conclusions from this work. First, reporting effort significantly influences where we observe EID events. This implies that EID events of a similar scale are occurring, unobserved, in locations with weaker disease reporting infrastructure. Second, reporting infrastructure is stronger in developed countries, in northern latitudes, whereas wildlife zoonoses more commonly emerge in lower latitudes, in countries with weaker reporting effort. The implications are that our resources to rapidly identify novel EIDs poorly match their likely occurrence, and that this can be remedied by improving infrastructure in EID hotspot developing countries to identify pathogens spilling over from wildlife into people.

It is important to note that our analysis, while suggestive, is preliminary. Reporting effort is likely more collinear with population density than a country-level measure can show, and the *Journal of Infectious Diseases* may not be the most accurate measure of where infectious disease reporting is strongest. Constructing higher-fidelity maps of infectious disease reporting effort would allow us to better correct for the lens through which we view disease emergence and identify areas with the greatest need for increased surveillance. Furthermore, the EID hotspot maps are relevant only at large spatial scales. New approaches are needed to identify where, within a region, country, or landscape, the highest risk of a new disease originating exists. One approach is to conduct targeted surveillance efforts at specific wildlife–human interfaces such as people living in remote villages close to forests in EID hotspots, or people engaged in hunting bushmeat, producing livestock, selling live animals in markets, or butchering them in abattoirs or restaurants. Better analysis of the spatial distribution and relative risk of these interfaces is likely to be a productive research line. Finally, with the growing availability of “big data,” increasing ease by which it can be manipulated and analyzed, and new models that predict future trends in the underlying drivers of EIDs, hotspot models will become more rigorous, accurate, and based on concrete hypotheses about biological mechanisms.

*How Many Unknown Pathogens Are There?*

The perfect pandemic prevention program would prevent spillover of pathogens from wildlife to human hosts before they have the opportunity of infecting
people, amplifying their transmission, and becoming pandemic. This approach is theoretically possible. If we target surveillance of wildlife to EID hotspot countries and conduct pathogen discovery in these species, we can identify pathogens with pandemic potential before they emerge and target prevention efforts to block their spillover. This is the basis for a number of new programs, including the USAID Emerging Pandemic Threat (EPT) program (Morse et al., 2012) and research programs that target pathogen discovery in bats and other zoonotic disease reservoirs (Drexler et al., 2012; Marsh et al., 2012; Wacharapluesadee et al., 2013).

However, even when we have narrowed down interfaces and locales of interest, two significant challenges remain. Firstly, the diversity of unknown pathogens may be so high that it is not cost effective to identify them all. Indeed, until recently there was no systematic attempt to predict the unknown viral diversity in any single species, let alone all wildlife. Using samples collected and tested through the USAID EPT PREDICT program, we have recently published the first attempt at a strategy to estimate unknown viral diversity. We did this using incidence-based species richness estimators, which have their origin in the “mark-recapture” modeling approach used by conservation biologists to estimate the density of rare animals in a patch of land. In this method, animals are captured, tagged, and released, and the number of recaptures of tagged individuals relative to the number of untagged individuals gives a way to statistically predict the total number of individuals in a region. For pathogen discovery, we repeatedly sampled a large population of *Pteropus giganteus*, a bat species known to carry zoonotic viruses, collecting high-quality samples from around 2,000 unique individual bats. We then used degenerate viral family-level primers (12,793 separate consensus PCR assays) to discover 55 viruses from nine viral families known to harbor zoonoses (Anthony et al., 2013). We then used statistical approaches to estimate the total viral richness of these nine families in this single species. Our analysis suggests that this bat species harbors 58 viruses (i.e., 3 not yet discovered) in these viral families, and if this is extrapolated simplistically to all 5,517 known mammal species, we estimate that there are at least 320,000 mammalian viruses awaiting discovery in these nine viral families. This is a large number, but using the PREDICT program costs of field and lab work, we estimate the cost to uncover 100 percent of virodiversity in this critical group of wildlife reservoirs to be $6.8 billion, and to uncover 85 percent of virodiversity to be only $1.4 billion, considering the exponentially diminishing returns of continued sampling. The latter figure is less than the cost of a single SARS-scale pandemic and, if spread over a decade, a small portion of current global pandemic prevention spending.

*Which Wildlife Species Harbor the Next Pandemic Pathogen?*

The second challenge to wildlife pathogen discovery as a pandemic prevention strategy is knowing which wildlife are the highest-risk reservoirs (i.e., which species to sample so that we can maximize the discovery of pathogens
with zoonotic, and pandemic, potential). Species differ in the composition of their viral diversity and in the propensity of those pathogens to infect people, but the genetic, behavioral, and ecological rules that underpin these relationships are poorly understood (Bogich et al., 2012b). A recent analysis of the literature found that sampling effort, IUCN threat status, and population genetic structure of bat species were the best predictors of how many viral species they harbored, independent of their phylogenetic relationships (Turmelle and Olival, 2009). Among mammal groups, rodents and bats host a particularly large number of zoonotic pathogens: Rodents have a larger diversity, while bats host more per species (Luis et al., 2013). Within bat and rodent species, those with greater sympatry (range overlap) with other related species host more viral diversity, and bats with smaller litters, greater longevity, and more litters per year tended to host more zoonoses. These are tantalizing glimpses of ecological and evolutionary patterns that likely drive viral speciation and zoonotic risk, and may ultimately inform which species we target for viral discovery. However, there is much more to learn. For example, a logical assumption is that viruses are more able to infect more closely related species, due to the sharing of host cell receptors, for example. Thus, mammals are the source of the majority of zoonotic EIDs (Jones et al., 2008; Taylor et al., 2001) and across all mammal-virus associations, more closely related mammals are more likely to share virus species (Bogich et al., 2012b). However, when two unrelated species have extensive, intimate contact over long periods of time (e.g., humans and domesticated mammals), does this phylogenetic rule still hold? If we continue to expand the wildlife trade, bringing more diverse animals from different regions into close contact with people, will we see pathogens emerging that would normally have difficulty successfully infecting humans?

Can We Predict the Pandemic Potential of a Newly Discovered Pathogen?

With targeted improvements in public health infrastructure and surveillance for pathogen discovery, we can increase our odds of catching a zoonotic outbreak in its nascence or discovering novel pathogens of pandemic potential. But will we be able to identify which ones, out of the hundreds of thousands of new species of virus to be discovered in wildlife, will be able to infect humans? With most of these potential zoonoses being identified by only a short RNA or DNA sequence, is there a logical strategy to identify their potential pandemicicity? Identifying which novel pathogens in a wildlife species are most likely able to infect, replicate in, cause cycles of human-to-human infection, and then amplify into pandemics remains one of the biggest challenges to pandemic prevention.

Morse et al. (2012) reviewed some of the known factors that affect whether a particular virus can infect a species and what gaps remain. In some pathogens, receptor specificity and other biological characteristics may be used to predict host range and potential pathogenicity to humans. However, animal models, human cell cultures, and similar methods cannot empirically validate a pathogen’s
capacity to infect humans. Some characteristics that may yield improvements in our predictive ability include the effects of host relatedness, relatedness of a virus to known human viruses, host range and evolutionary capacity, and predictive capacity of virulence in humans (some pathogens can infect humans but cause no disease, whereas others cause severe illness) (Morse et al., 2012). As we work towards a better understanding of these factors, we can use a few simple heuristics to prioritize certain pathogens. Certainly, if a pathogen exists at a zoonotic interface, and if there has been documented human infection, it should be prioritized. Pathogens that cause small chains of human-to-human infection with a basic reproductive number ($R_0$) approaching or higher than 1 should be considered “prime epidemics in waiting,” as small evolutionary changes could boost their transmissibility and enable them to cause epidemics.

In fact, though none of the models outlined above can tell us exactly how dangerous a pathogen is, they all contribute valuable information to a risk assessment. Whether a pathogen exists at an interface of interest, how closely related it is to known human pathogens, how closely related to humans its reservoir host is, and various viral traits all convey information about a particular pathogen. Future work may involve testing the zoonotic potential of wildlife pathogens by sequencing receptor binding domains, producing pseudo-type viruses with these proteins expressed, and conducting binding assays, in vitro culture assays, and ultimately animal infections with transgenic animals that express human cell surface receptors. This work has already shown the capacity to identify high-priority potential zoonoses for SARS-like viruses in bats, which bind to human, civet, and bat ACE2 (Ge et al., 2013).

**Can We Predict How, and to Where, a New EID Will Spread?**

The emergence of triple reassortant A/H1N1 influenza in 2009 highlighted how rapidly diseases can spread once they have achieved capacity for effective human-to-human transmission. Targeting these diseases may be effective if we can accurately predict their likely pattern of spread out of a region and strategically allocate resources to respond. Analyses of travel and trade data have shown that predicting spread is relatively straightforward and can provide accurate estimates of spread and case numbers when applied to prior outbreaks (e.g., of SARS (Hufnagel et al., 2004) and A/H1N1 influenza (Hosseini et al., 2010)). This approach has been used to analyze recent historical spread of vectors through shipping trade, and their likely routes of spread via air travel (Tatem, 2009; Tatem et al., 2006a,b). It has also been used to predict the spread of ongoing emergence events such as the MERS-CoV outbreak in Saudi Arabia (Khan et al., 2010). It has particular relevance in zoonotic disease spread when patterns of wildlife migration and trade are implicated, and where policy can be rapidly set to prevent importation. This approach has been used to examine the likely cause of past spreading events for A/H5N1 influenza and to predict and set policy for its likely
route of introduction to the New World (Kilpatrick et al., 2006a). Finally, it has been used effectively in Hawaii and the Galapagos Islands to allocate resources to reduce the risk of West Nile virus introduction via the most likely pathway of mosquitoes transported via air travel (Kilpatrick, 2011; Kilpatrick et al., 2004, 2006b). As in all predictive models, their rigor improves as the quality of data on travel and trade pathways and volumes, on biological characteristics of pathogen and host, and on the human contact networks that allow transmission also improves. For example, the capacity and willingness of countries to identify and report outbreaks early are critical to make accurate predictions about spread, once a pandemic has begun. Analyses of the spread of the 2009 H1N1 influenza showed that two key factors influenced the pandemic’s arrival time—a country’s global accessibility through air travel, and the percentage of GDP per capita spent on health care (a proxy for testing and reporting capacity) (Hosseini et al., 2010). Again, gaps in this approach remain, including the need for a better understanding of the role of intra-country human movement in disease spread. Newly available datasets on road infrastructure, migration, and human network connectivity will increasingly illuminate this area.

Can We Eventually Stop Pandemics from Emerging?

The new approaches described above to identifying novel pathogens in emerging disease hotspots, and predicting their pandemic potential and likely spread, have likely improved our global pandemic preparedness. But what progress has been made in using this approach to prevent pandemics? One significant shift is in the way pandemic prevention programs are funded and managed. Traditionally, outbreak threats were dealt with by state and national agencies, the World Health Organization, and field laboratory networks funded through these programs. The emergence of H5N1 influenza via small-scale outbreaks, which suggested chronic persistence in backyard poultry farms, led to calls for a “systems approach” to the pandemic prevention (Bogich et al., 2012a), and a cross-sectorial “One Health” collaboration of animal health, public health, and environmental agencies (FAO et al., 2008; Karesh, 2009; Zinsstag et al., 2011). International development agencies, which had been trending towards specialized programs to target specific infectious diseases, are now actively involved in this systems approach to pandemic prevention. This involves funding for crucial infrastructure investments required for pandemic prevention, and a specific focus on collaborative One Health programs (Bogich et al., 2012a). With most EID events occurring in regions that are under-resourced in public health capacities, disease-based programs for AIDS, malaria, TB, and polio do not address the underlying flaws in public health systems that predispose locations to outbreaks of emerging infectious diseases (Standley and Bogich, 2013).

Standley and Bogich (2013) propose an “ecohealth” approach, addressing destructive land use change and biodiversity loss in places like China, Brazil, and India. This approach defines how we can deal with pandemics as distinct
from dealing with hurricanes or earthquakes: by identifying and mitigating the underlying causes, particularly anthropogenic activities that promote pathogen spillover, amplification, and spread. Strategies include programs that educate and promote alternatives to high pandemic risk behavior like the trading, butchering, and consumption of wild animals, or the comingling of livestock and wildlife on farms. They also include more fundamental approaches that address large-scale anthropogenic changes. For example, 43 percent of past EID events are attributable to land use change and agricultural changes, including extractive industries (timber/logging, oil and gas, mining, and plantations). The economic impact of EIDs from land use change is estimated to be $10–40 billion over the next 10 years, which could be considered a potential liability to extractive industries. Industrialized mining and plantation operations in EID hotspot countries are likely to be on the front line of disease outbreaks, and are often under pressure to improve their environmental impacts. Programs that better quantify the risk of novel pathogens to these industries, and the economic damages they might entail, may become valuable in mitigating their impact on global health, conservation, and the environment (Murray and Daszak, 2013).

The two-fold approach of treating emerging pandemics as targets for international development programs and as byproducts of economic activity is relatively new and suggests that long-term solutions to their emergence can be found. A future without pandemics may be possible, but only with the very best interdisciplinary science, ambitious approaches to risk prediction, and bold strategies taken by industry and development agencies to ensure against them.

References


The Influence of Global Environmental Change on Infectious Disease Dynamics: Workshop Summary

GLOBAL CHANGE AND INFECTIOUS DISEASE DYNAMICS


The last 25 years have seen a renaissance in the use of mathematical models in epidemiology; much of this is largely due to the influence of Anderson and May and their colleagues (Anderson and May, 1992; Grenfell and Dobson, 1994, 1995). The transformation came about as the models they developed were based upon empirical assumptions. This allowed the whole discipline to move from an overt fascination with mathematical elegance, to embrace data and become the pragmatic powerhouse that is at the center of quantitative insight to any modern epidemiological problem. At first glance, this creates problems for the use of these models in studies of emerging diseases, as almost by definition, there will be no data prior to emergence. Nonetheless, all of the recent major studies of disease emergence have quickly led to the almost obligatory use of mathematical models in infectious disease biology. A nice index of this was the chance remark by the editor of one major journal during a recent influenza outbreak, “Half the world is worried about this new pathogen—while we’re facing an epidemic of submitted papers, all claiming to have produced the definitive predictive model for it!”

In this short overview, I will take a brief personal and idiosyncratic review of the key ways in which mathematical models have been used, misused, or could potentially be used to provide insights into the dynamics of emerging pathogens. I will offer no specific recommendations or recipes for the “best way” to use models to understand pathogen emergence. This is partly because different model structures will provide different insights to different pathogens; moreover, each new emergence usually leads to the development of new mathematical tricks, techniques, and approaches that

25 Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544.