

Challenges in modeling the emergence of novel pathogens.

Emma E. Glennon¹, Marjolein Bruijning², Justin Lessler³, Ian F. Miller^{2,4}, Benjamin L. Rice^{2,5}, Robin N. Thompson^{6,7}, Konstans Wells⁸, C. Jessica E. Metcalf^{1,9}

¹ Disease Dynamics Unit, Department of Veterinary Medicine, University of Cambridge, Cambridge CB3 0ES, UK

² Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA

³ Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

⁴ Rocky Mountain Biological Laboratory, Crested Butte, CO 81224, USA

⁵ Madagascar Health and Environmental Research (MAHERY), Maroantsetra, Madagascar

⁶ Mathematics Institute, University of Warwick, Warwick, CV4 7AL, UK

⁷ The Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, University of Warwick, Warwick, CV4 7AL, UK

⁸ Department of Biosciences, Swansea University, Swansea, SA28PP, UK

⁹ Princeton School of Public and International Affairs, Princeton University, Princeton, NJ, USA

Abstract

The emergence of infectious agents with pandemic potential present scientific challenges from detection to data interpretation to understanding determinants of risk and forecasts.

Mathematical models could play an essential role in how we prepare for future emergent pathogens. Here, we describe core directions for expansion of the existing tools and knowledge base, including using mathematical models to identify critical directions and paths for strengthening data collection; expanding basic theory to identify infectious agents that present the greatest risks, over both the short and longer term; by strengthening estimation tools that make the most use of the likely range and uncertainties in existing data; and by expanding access to effective, transparent tools to harness modelling for increased public health benefit.

Keywords: immune landscape, genotype to phenotype map, big data, data integration, fundamental theory, health system functioning.

Introduction

In 2020, the emergence of novel pathogens, or the successful spread of a pathogen into a new host environment, sprung to unwanted prominence. The previously unknown coronavirus, SARS-CoV-2 made the jump to humans from a zoonotic reservoir in China, and it rapidly achieved global reach (Kissler et al. 2020). Mathematical models were deployed at every stage of this trajectory and continue to be an important part of scientific and political conversations regarding how to best contain the spread of SARS-CoV-2 and other pathogens with pandemic potential. But with the zoonotic virome diverse and rich with both known and unknown pathogens, key challenges remain in the effort to understand and mitigate the emergence of future zoonotic epidemics. Although there have been striking advances in some of the challenges laid out in previous overviews (Lloyd-Smith et al. 2015), for example in viral genomic epidemiology (Grubaugh et al. 2019) and characterizing host competence (Mollentze et al.

2020), other challenges remain relatively neglected, such as the potential roles of intermediate hosts in spillover into people or “spillback” into wildlife (Glennon et al. 2018, Fagre et al. 2021). In addition to these familiar challenges in mathematical modeling for pathogen emergence, new ones have emerged in the wake of SARS-CoV-2. Here we provide an overview of some of these major remaining challenges.

For a pathogen to emerge within a new host population two steps are involved: introduction and emergence. Although these two steps can be further divided into ecologically distinct substeps (Plowright et al. 2017), these two stages are crucially driven by distinct underlying dynamic processes. First, the infectious agent must be introduced into the new host population from either a zoonotic reservoir or intermediate species such as a domesticated animal (i.e., spillover), or an infected conspecific from another population, noting that this might happen several times, as is the case for MERS-CoV. This process is driven both by the circulation of a pathogen in the reservoir and contact patterns at the human-wildlife interface. Second, the pathogen must become established within the new host population (requiring sustained local transmission between hosts, commonly summarised by a value of R_t , or the number of secondary infections per infected individual, greater than 1; Figure 1). Models may make contributions to advancing our understanding of both steps, and critical challenges remain to effectively modelling both stages.

Here, we consider challenges in the ways in which models can contribute to strengthening data collection in the context of future pandemics; identify major challenges in refining the theoretical framing around pathogen emergence; and detail important questions in estimation of critical pandemic related parameters. While the emergence of a novel pathogen might occur with devastating impact in any host species (evidenced, for example, by devastation of the citrus industry by the virus *Citrus tristeza* (Lee, Baker, and Roche-Peña 1994)), our focus is predominantly on spillover into human populations, and emergence following such spillovers.

Data challenges of emergence of future pandemics

Two of the major questions around data in the context of future pandemics are: What data need to be collected to detect pathogen emergence? And, what can we do with (and what are the limits of) data that already exists? Models may contribute to answering both questions.

Focusing on the first question, what surveillance tools could most effectively be leveraged to detect that pathogen emergence is (or was) occurring? Models could be used to probe what scope (temporal, spatial), types of data (individual measurements, clinical convenience samples, designed cohort studies), and types of measurements (symptoms, antibody responses, sequencing), and sampling intensity are needed to reliably detect anomalies indicative of a pathogen emergence event. For example, a sudden uptick in cross-reactive antibodies to coronaviruses relative to baseline in a convenience sample (e.g., from a blood bank) might have provided a useful early warning of an emergent coronavirus (Mina et al. 2020), but the scale and scope of sampling required for a relevant anomaly to be reliably detected remains an open question, especially in light of recent evidence for cross-reactivity against SARS-CoV-2 in pre-pandemic sera in Southern Sudan, for example (Wiens et al. 2021).

Similarly, genetic sequencing clearly has the potential to identify pathogen emergence events, but whether the required frequency of sampling, or range of reservoir or spillover host individuals sampled, is likely to be tractable is unknown, and should not be oversold (“From PREDICT to Prevention, One Pandemic Later” 2020). Models could be designed to explore this issue, for example exploring the degree to which pathogen life histories might require different sampling designs, or whether existing convenience samples (e.g., blood banks) or unexpected clusters of large case reports might be adequate. For pathogens that are already circulating within a focal population, emergence could take the form of the appearance and establishment of novel variants (e.g., greater transmission, or immune escape, or greater virulence, as for SARS-CoV-2 (McCormick, Jacobs, and Mellors 2021); or features such as artemisinin resistance for malaria (Miotto et al. 2013)). Curtailing the spread of such variants has clear public health benefits. Characterising the degree of surveillance required to detect novel sequences at a fast enough rate to enable containment could be addressed via modeling. For poorly characterized or unknown pathogens, predicting potential geographical ‘hotspots’ for future emergence events could facilitate more efficient and targeted surveillance. Overall, a first challenge for models in addressing the data around emergence events is thus in informing the utility and best designs for curation of wide-scale, regularly collected data on pathogen (or variant) presence via direct detection or immune measurement.

Moving to the second question, a second major set of issues associated with data on pathogen emergence is the variation in data quality and completeness that emerges from global resource inequalities. Ebola outbreaks, for example, can easily evade detection in places where access to care or availability of confirmatory diagnostics are limited (Jephcott, Wood, and Cunningham 2017; Glennon et al. 2019). Data on emerging outbreaks inherently reflects those outbreaks which have successfully been detected and pathogens conclusively identified. Available data therefore reflects extensive but difficult-to-quantify biases toward larger, more syndromically distinct outbreaks in places with well-resourced health systems and effective disease surveillance programs (Glennon et al. 2020). Furthermore, as much pathogen prevalence data is collected in clinical settings, surveillance represents a low priority relative to providing direct care in many situations (Kim, Farmer, and Porter 2013). Models that characterize relevant data biases and feedback loops, as well as models that identify approaches to correct or account for such challenges, will clearly be of value for planning exercises. Awareness of these limitations will be key to effective and just prioritization of disease prevention efforts. A second challenge for modeling is therefore in appropriately addressing these imbalances in the data, e.g., via integration of data collected by different systems (including qualitative study and local/Indigenous knowledge) or quantitative estimation of underreporting and other observation biases.

Relatedly, beyond direct or indirect measures of pathogen presence and abundance, modeling might also be used to ask the question of whether existing non-traditional data-streams (such as contact patterns from mobile devices (Chang et al. 2021; Wesolowski et al. 2016), global connectivity from integrated global data sources (Tatem, Hay, and Rogers 2006), population heterogeneity by integrating satellite images to census data (Worldpop n.d.), digital trend data) could contribute to anticipating the trajectory that will follow pathogen emergence. Suggestive

use cases exist. For example, mobility data from mobile devices contributed to understanding the early phases of spread (and effects of containment) of SARS-CoV-2 (Lai et al. 2020). However, such data does not directly measure transmission events, and, for example, only limited predictability could be obtained from mobile phone data for the spatial spread of a measles outbreak in Pakistan (Wesolowski et al. 2018). In the context of future pandemics, models could evaluate the characteristics of data that would most powerfully refine inference into future pandemic trajectories (temporal scale? spatial scale? demographic features?) to identify a critical set of extensions that could be requested of the technology companies and should be included in data use agreements (bearing in mind ethical and privacy constraints).

Challenges in developing the theoretical framework for understanding pathogen emergence

Detecting pathogen emergence is clearly, in itself, ambitious. An even more ambitious goal would be to identify pathogens within their zoonotic reservoir before they have had the opportunity to spill over into people, as well as to understand and mitigate the risk of emergence after spillover. There are many possible threads by which models could contribute to this.

Sequencing important pathogen reservoirs, such as the ‘virome’ is increasingly tractable on large scales, and has certainly deepened our understanding of the community ecology of viruses (Wille et al. 2019). Nevertheless, vast numbers of species remain undescribed. It has been estimated that more than 600,000 viral species with zoonotic potential remain to be characterised (Carroll et al. 2018). As the vast majority of sequenced viruses will be irrelevant to human health, identifying ways to target sampling to more relevant parts of the biome is an important question, for viruses, but also beyond (bacteria, protozoa), and with potential to contribute to monitoring of zoonotic reservoirs (Lloyd-Smith et al. 2015). This question links to the important theoretical challenge of pinpointing characteristics of pathogens with spillover potential.

A series of theoretical approaches have evaluated the degree to which very general characteristics (e.g., mutation rate, degree of clustering in host contacts, pathogen related mortality) shape risks of pathogen emergence (reviewed in (Gandon et al. 2013)). Another way to tackle this question is to use a comparative approach: models are used to leverage knowledge and trait characteristics of hosts and pathogens that have historically spilled over to identify features of future pandemic pathogens (Olival et al. 2017; Wells et al. 2020; Shaw et al. 2020). Remaining key challenges in this area are to more realistically link species-level predictions to real-world landscapes by accounting for fine-scale species distributions and exposure risk. This area of study may prove particularly important for the interdisciplinary challenge of understanding the effects of ecological and climatic change on disease emergence. A related approach, and which echoes a perennial and persisting challenge in biology is in generating a genotype to phenotype map (Visscher et al. 2017). This might require possibly unfeasible (at this stage) mechanistic understanding of features from cellular tropism to pathogen replication to interactions with existing immunity. Better calibrated and detailed models capturing within-host to between-host dynamics (Ke et al., n.d.) will contribute to this

effort. Arguably, a critical and as yet unresolved component will be a larger understanding of the landscape of immunity for the focal species and across the pathogen community, as cross-reactivity among different pathogen species might constrain pathogen emergence. Engaging with this in turn requires engaging with the important modeling challenge of moving beyond a single pathogen perspective to encompass multiple species, a well recognized and persisting challenge in the study of infectious pathogens (Lipsitch et al. 2009; Wikramaratna et al. 2015; Kucharski, Andreasen, and Gog 2016).

As the focal pathogen spreads, growing immunity within the population will elicit selection pressures on within-human replication and human to human transmission. For SARS-CoV-2, in early 2021, variants of concern feature combinations of traits encompassing receptor avidity, immune escape, transmissibility and virulence (Martin et al. 2021). Knowledge of correlations between these traits and trade-offs (implying a genotype to phenotype map) would open the way to better anticipating everything from immune escape to shifts in virulence, as well as a better understanding of how these evolutionary events might be driven by acquired immunity, therapeutics, or vaccination regimes (Saad-Roy et al. 2021). While the potential for such mechanistic details to refine the predictions of evolutionary models is clear, adequate characterization of mechanism remains an important frontier, to which more nuanced knowledge of molecular mechanisms through to better models of within-host spread could be brought to bear.

An important component in developing models to project the impacts of such pathogen traits and trade-offs (or the genotype to phenotype map) is that they may differ from host to host. Sex differences in immune function, for example, are ubiquitous (Klein and Flanagan 2016); and older individuals generally have less efficient immunity as a result of immunosenescence (Simon, Hollander, and McMichael 2015). The feedback driven nature of immunity may also mean that very small differences may escalate into highly variable outcomes (e.g., along the lines described in ecological terms by ‘alternative stable states’ (Metcalf, Grenfell, and Graham 2020)). Such population heterogeneities may have important consequences for both the prospects for pathogen spillover, initial pathogen spread (i.e., the effective reproductive number (Lloyd-Smith et al. 2005; Metcalf et al. 2015)), and selection on pathogen traits, such as virulence (Miller and Metcalf 2019). However, such heterogeneities remain relatively rarely modeled, in part as they are generally extremely hard to quantify in practice (see next section), especially when one considers that individual variation in important traits such as immunity, behavior, and mobility may all compound to produce highly complex patterns. Nevertheless, efforts to evaluate the purely theoretical impact of such heterogeneities, rooted in broadly known differences (by sex, age) could significantly advance understanding of variation in transmission dynamics and disease outcomes within and between host populations.

Challenges in estimation around pathogen emergence

While many spillover events may rapidly go extinct in human populations, sometimes primary infections (i.e., people infected directly from the reservoir or entering the local population from elsewhere) will go on to infect other people. Branching process models are commonly used to

estimate the risk that initial cases of disease will establish sustained chains of transmission (Althaus et al., n.d.; Guzzetta et al. 2016; Thompson, Gilligan, and Cunniffe 2016; Abdullah et al. 2018; Thompson, Jalava, and Obolski 2019; Thompson, Gilligan, and Cunniffe 2020). These models were applied early in the COVID-19 pandemic, when cases had only been observed in China, to assess the risk the cases exported to other countries would establish local epidemics and to investigate how control measures affect this risk (Hellewell et al. 2020; Thompson 2020). Modeling approaches have also been developed to extract key epidemiological measures such as R_0 from observations of early 'stuttering chains' of transmission (Blumberg and Lloyd-Smith 2013). Expanding this work to explore how models might be used to estimate R_0 and other core parameters such as the incubation period, or proportion symptomatic in the very early phases prior to a spillover establishing using other sources of data (immunological information, etc) might help identify pathogens that might make the leap to establish ongoing transmission. However, stuttering chains can be recurrent but not necessarily transition to a larger threat (as is the case for MERS-CoV, for example), and it is important to note that there are still large questions as to where investigations should be focussed.

For events that follow this phase, the classical modeling toolkit was perhaps first defined during the emergence of HIV (May and Anderson 1987). This work established a roadmap for estimating the range of core parameters required to delineate the trajectory that an emerging pathogen will take from the early growth in cases (R_0 , generation time, incubation period, overdispersion, infection fatality ratio (Metcalf and Lessler 2017)). This toolkit has been expanded by use of genetic data-streams to infer incidence or pathogen population growth (Vaughan et al. 2019), or innovative use of cross-sectional data on viral copy numbers across individuals population to capture whether pathogen populations are growing or shrinking (Hay et al. 2020). There are presumably an array of further innovations in this space that might further enhance this set of approaches; building around elements listed in previous sections, and grappling with key features such as how to address variability (e.g., presence of superspreading) but also uncertainty in critical transmission parameters (e.g., incubation time, quantities such as R_0 in human populations once spillover has occurred) and grounding them mechanistically (e.g., in heterogeneity in contact patterns or distributions of co-morbidities).

Integrating diverse data sources is likely to prove important both in the very early phases when stuttering chains are occurring, and once important spread has occurred. For example, phylodynamic approaches can provide another window onto how a virus moves through space (Bedford et al. 2020; Deng et al. 2020). Such integration will also provide a means to triangulate on core measures such as R_0 (Vaughan et al., n.d.) using a different source of data (noting that this approach may have limited power to resolve the issue of pinning down transmission for a pathogen like SARS-CoV-2 where the genome evolves relatively slowly, time to onset of symptoms, as well as incubation and infection periods are variable, and the rate of asymptomatics is high). Expanding models to estimate critical aspects of within-host dynamics (using data on viral load, immune parameters, etc) but also to translate these estimates into parameters relevant to population scale transmission remains very much in its early phases.

There are also very important required modelling extensions of the classical toolkit to use new data sources. There is arguably more data than ever before on aspects of human contact and risk. Since every model of pathogen emergence arguably involves a contact rate, such data has clear potential. However, there are still gaps in thinking about how best to harness these data. Transmission events remain frustratingly unspecified; they are only 'observed' (and even then, indirectly) in extraordinarily detailed data such as contact tracing (e.g., (Bi et al. 2020)). Thus, we do not know whether the 'medium' or 'big' data sources we have access to at the population scale, be they diary studies of contacts, or mobile phone call data records (Grantz et al. 2020) actually capture transmission relevant contacts. A challenge is finding a principled way to grapple with these issues of model mis-specification that remains tractable, and also sensibly reflects uncertainty. Multi-model comparisons (Reich et al. 2019) may make important contributions here.

Models of early pathogen spread following emergence that better account for logistics and health systems, relevant to both understanding reporting, but also to potential for intervention and containment at early phases is another important gap. Spatial mechanistic models that include population connectivity networks, socioeconomic factors and distribution of potential reservoir species as recently employed to model Ebola emergence (Redding et al. 2019) may provide crucial insights into spatial aspects of disease emergence. Such mechanistic models can be also used to explore a wide range of possible scenarios in order to explore the model behaviour across a large array of combinations of transmission parameters or narrow down intractable parameter values through likelihood-free approximation methods. Challenges to be tackled in this area include sensible choices of model complexity, efficient sampling of large unknown parameter spaces, accounting for transient dynamics, as well as accurate model validation amid the generally sparse empirical evidence available for emerging diseases.

Challenges in harnessing models for public health benefit

Beyond developing methods and theory, there are important challenges in implementing models to effectively contribute to outbreak prevention and response. Such challenges are cross-cutting, affecting every stage and scale of emergence and of model development. Broadly, these challenges include assessment and communication of uncertainty; clear contextualization of short- and long-term disease prevention and control strategies; development of infrastructure for effective and responsive modelling in cooperation with public health bodies; and equitable access to epidemiological tools and insights provided by modelling.

Effectively assessing and communicating the uncertainty involved in model predictions is crucial to meeting the needs of public health decision makers and policy makers, as well as facilitating informed decision-making by the general public in the face of misinformation. Clear communication is especially difficult in the early stages of an epidemic, when basic data such as infection rates, infection fatality ratios, and risk factors may be unavailable or strongly skewed by the chance circumstances of early cases. Careless or overconfident communication of modelling predictions during this stage risks losing the trust and cooperation of public health decision makers and the general public. Nonetheless, the early stages of an outbreak, while

case counts are still low, offer the best opportunity for effective intervention to minimize direct and indirect costs. Waiting for more accurate data risks foregoing the chance to intervene while an outbreak is still manageable. There is therefore an urgent need to develop strategies to balance communication of urgency and uncertainty in presenting modelling results as a disease begins to emerge, as well as to effectively communicate changes in strategy as modelling approaches evolve with new data (Becker et al. 2021).

Similarly, there is an urgent need to improve communication of model assumptions and limitations, including contextualization of short-term and narrowly disease-centered strategies (e.g., vaccination and behavior changes such as social distancing) within local context and long-term or holistic strategies (e.g., broad improvements in sanitation, access to safe housing and work environments, and universalization of care). Models have been used to disentangle the influences of such factors on the dynamics of endemic and historical emerging diseases (e.g., Phillips et al. 2020), and recent efforts have retrospectively estimated the consequences of such fundamental causes of disease as structural racism on the emergence of COVID-19 in the United States (Richardson et al. 2021). Prospectively quantifying the impacts of health infrastructure and broad social change, however, remains a long-term challenge. As modellers work to advance our understanding of the local environment (e.g., climate, health infrastructure, structural inequality, and interactions with other diseases) on disease spread, it remains important to highlight the limitations of acting only on the proximate mechanisms and easily measured parameters most commonly captured by quantitative modelling. In particular, modellers can support systemic public health efforts by emphasising that parameters such as R_0 are emergent and changeable properties of a disease within a human social context, rather than inherent properties of pathogens themselves.

Finally, additional efforts are needed to define and shape the role of modelling within broader public health efforts in disease prevention and outbreak response. More work is needed to build strong, open collaborations between modellers and public health decision makers, as well as to support additional infrastructure for data sharing and to develop accessible tools based on modelling that will enable insights from modelling to be distributed more equitably.

Conclusions

Detection, and even more ambitiously, anticipation of the emergence of an infectious agent into a new host population could importantly contribute to efforts to assess, prevent, control, and contain future pandemic risks. We have outlined here an array of ways that models have the potential to inform these questions, and important challenges ahead. A global health perspective underscores the enormous advantages in terms of speed and transparency for early reporting of potentially important spillover events. Both better fundamental understanding and expectations, alongside global data standards, transparency and sharing, informed by model framing, could contribute to this while also guarding against over-selling the potential of this research agenda.

Acknowledgements

MB was funded by NWO Rubicon (019.192EN.017). IFM was funded by the High Meadows Environmental Institute, Princeton. EEG is funded by a Sir Henry Wellcome Postdoctoral Fellowship [220463/Z/20/Z]. We would like to thank Denis Mollison & Valerie Isham for comments on an earlier draft.

References

- Abdullah, Nurshariza, James T. Kelly, Stephen C. Graham, Jamie Birch, Daniel Gonçalves-Carneiro, Tim Mitchell, Robin N. Thompson, et al. 2018. "Structure-Guided Identification of a Nonhuman Morbillivirus with Zoonotic Potential." *Journal of Virology*. <https://doi.org/10.1128/jvi.01248-18>.
- Althaus, Christian L., Sandro Gsteiger, Emmanuel O. Musa, Faisal Shuaib, and Nicola Low. n.d. "Ebola Virus Disease Outbreak in Nigeria: Transmission Dynamics and Rapid Control." <https://doi.org/10.7287/peerj.preprints.569v2>.
- Becker, Alexander D., Kyra H. Grantz, Sonia T. Hegde, Sophie Bérubé, Derek A. T. Cummings, and Amy Wesolowski. 2021. "Development and Dissemination of Infectious Disease Dynamic Transmission Models during the COVID-19 Pandemic: What Can We Learn from Other Pathogens and How Can We Move Forward?" *The Lancet. Digital Health* 3 (1): e41–50.
- Bedford, Trevor, Alexander L. Greninger, Pavitra Roychoudhury, Lea M. Starita, Michael Famulare, Meei-Li Huang, Arun Nalla, et al. 2020. "Cryptic Transmission of SARS-CoV-2 in Washington State." *Science*, September. <https://doi.org/10.1126/science.abc0523>.
- Bi, Qifang, Yongsheng Wu, Shujiang Mei, Chenfei Ye, Xuan Zou, Zhen Zhang, Xiaojian Liu, et al. 2020. "Epidemiology and Transmission of COVID-19 in 391 Cases and 1286 of Their Close Contacts in Shenzhen, China: A Retrospective Cohort Study." *The Lancet Infectious Diseases* 20 (8): 911–19.
- Blumberg, Seth, and James O. Lloyd-Smith. 2013. "Inference of $R(0)$ and Transmission Heterogeneity from the Size Distribution of Stuttering Chains." *PLoS Computational Biology* 9 (5): e1002993.
- Carroll, Dennis, Peter Daszak, Nathan D. Wolfe, George F. Gao, Carlos M. Morel, Subhash Morzaria, Ariel Pablos-Méndez, Oyewale Tomori, and Jonna A. K. Mazet. 2018. "The Global Virome Project." *Science* 359 (6378): 872–74.
- Chang, Serina, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. 2021. "Mobility Network Models of COVID-19 Explain Inequities and Inform Reopening." *Nature* 589 (7840): 82–87.
- Deng, Xianding, Wei Gu, Scot Federman, Louis du Plessis, Oliver G. Pybus, Nuno R. Faria, Candace Wang, et al. 2020. "Genomic Surveillance Reveals Multiple Introductions of SARS-CoV-2 into Northern California." *Science* 369 (6503): 582–87.
- "From PREDICT to Prevention, One Pandemic Later." 2020. *The Lancet Microbe* 1 (1): e6–7.
- Fagre, Anna C., Lily E. Cohen, Evan A. Eskew, Max Farrell, Emma E. Glennon, Maxwell B. Joseph, Hannah K. Frank, Sadie Ryan, Colin J. Carlson, and Gregory F. Albery. 2021. "Spillover in the Anthropocene: the risk of human to wildlife pathogen transmission for conservation and public health." *EcoEvoRxiv*.
- Gandon, S., M. E. Hochberg, R. D. Holt, and T. Day. 2013. "What Limits the Evolutionary Emergence of Pathogens?" *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 368 (1610): 20120086.
- Glennon, Emma E., Freya L. Jephcott, Alexandra Oti, Colin J. Carlson, Fausto A. Bustos Carillo, C. Reed Hranac, Edyth Parker, James L. N. Wood, and Olivier Restif. 2020. "Syndromic Detectability of Haemorrhagic Fever Outbreaks." *medRxiv*, March, 2020.03.28.20019463.
- Glennon, Emma E., Freya L. Jephcott, Olivier Restif, and James L. N. Wood. 2019. "Estimating

- Undetected Ebola Spillovers." *PLoS Neglected Tropical Diseases* 13 (6): e0007428.
- Glennon, Emma E., O. Restif, S. Sbarbaro, R. Garnier, R. Suu-Ire, R. Osei-Amponsah, A., Cunningham et al. 2018. "Domesticated animals as hosts of henipaviruses and filoviruses: A systematic review." *The Veterinary Journal*, 233 25-34.
- Grantz, Kyra H., Hannah R. Meredith, Derek A. T. Cummings, C. Jessica E. Metcalf, Bryan T. Grenfell, John R. Giles, Shruti Mehta, et al. 2020. "The Use of Mobile Phone Data to Inform Analysis of COVID-19 Pandemic Epidemiology." *Nature Communications* 11 (1): 4961.
- Grubaugh, Nathan D., Jason T. Ladner, Philippe Lemey, Oliver G. Pybus, Andrew Rambaut, Edward C. Holmes, and Kristian G. Andersen. 2019. "Tracking Virus Outbreaks in the Twenty-First Century." *Nature Microbiology* 4 (1): 10–19.
- Guzzetta, Giorgio, Piero Poletti, Fabrizio Montarsi, Frederic Baldacchino, Gioia Capelli, Annapaola Rizzoli, Roberto Rosà, and Stefano Merler. 2016. "Assessing the Potential Risk of Zika Virus Epidemics in Temperate Areas with Established Aedes Albopictus Populations." *Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin* 21 (15). <https://doi.org/10.2807/1560-7917.ES.2016.21.15.30199>.
- Hay, James A., Lee Kennedy-Shaffer, Sanjat Kanjilal, Marc Lipsitch, and Michael J. Mina. 2020. "Estimating Epidemiologic Dynamics from Single Cross-Sectional Viral Load Distributions." *medRxiv*, October, 2020.10.08.20204222.
- Hellewell, Joel, Sam Abbott, Amy Gimma, Nikos I. Bosse, Christopher I. Jarvis, Timothy W. Russell, James D. Munday, et al. 2020. "Feasibility of Controlling COVID-19 Outbreaks by Isolation of Cases and Contacts." *The Lancet. Global Health* 8 (4): e488–96.
- Jephcott, Freya L., James L. N. Wood, and Andrew A. Cunningham. 2017. "Facility-Based Surveillance for Emerging Infectious Diseases; Diagnostic Practices in Rural West African Hospital Settings: Observations from Ghana." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 372 (1725). <https://doi.org/10.1098/rstb.2016.0544>.
- Ke, Ruian, Carolin Zitzmann, Ruy M. Ribeiro, and Alan S. Perelson. n.d. "Kinetics of SARS-CoV-2 Infection in the Human Upper and Lower Respiratory Tracts and Their Relationship with Infectiousness." <https://doi.org/10.1101/2020.09.25.20201772>.
- Kim, Jim Yong, Paul Farmer, and Michael E. Porter. 2013. "Redefining Global Health-Care Delivery." *The Lancet* 382 (9897): 1060–69.
- Kissler, Stephen M., Christine Tedijanto, Edward Goldstein, Yonatan H. Grad, and Marc Lipsitch. 2020. "Projecting the Transmission Dynamics of SARS-CoV-2 through the Postpandemic Period." *Science* 368 (6493): 860–68.
- Klein, Sabra L., and Katie L. Flanagan. 2016. "Sex Differences in Immune Responses." *Nature Reviews. Immunology* 16 (10): 626–38.
- Kucharski, Adam J., Viggo Andreasen, and Julia R. Gog. 2016. "Capturing the Dynamics of Pathogens with Many Strains." *Journal of Mathematical Biology* 72 (1-2): 1–24.
- Lai, Shengjie, Nick W. Ruktanonchai, Liangcai Zhou, Olivia Prosper, Wei Luo, Jessica R. Floyd, Amy Wesolowski, et al. 2020. "Effect of Non-Pharmaceutical Interventions to Contain COVID-19 in China." *Nature* 585 (7825): 410–13.
- Lee, R. F., P. S. Baker, and M. A. Roche-Peña. 1994. *The Citrus Tristeza Virus (CTV): An Introduction to Current Priorities, with Special Reference to the Worsening Situation in Central America and the Caribbean*. C A B International.
- Lipsitch, Marc, Caroline Colijn, Ted Cohen, William P. Hanage, and Christophe Fraser. 2009. "No Coexistence for Free: Neutral Null Models for Multistrain Pathogens." *Epidemics* 1 (1): 2–13.
- Lloyd-Smith, James O., Sebastian Funk, Angela R. McLean, Steven Riley, and James L. N. Wood. 2015. "Nine Challenges in Modelling the Emergence of Novel Pathogens." *Epidemics* 10 (March): 35–39.

- Lloyd-Smith, J. O., S. J. Schreiber, P. E. Kopp, and W. M. Getz. 2005. "Superspreading and the Effect of Individual Variation on Disease Emergence." *Nature* 438 (7066): 355–59.
- Martin, Darren P., Steven Weaver, Houryiah Tegally, Emmanuel James San, Stephen D. Shank, Eduan Wilkinson, Jennifer Giandhari, et al. 2021. "The Emergence and Ongoing Convergent Evolution of the N501Y Lineages Coincides with a Major Global Shift in the SARS-CoV-2 Selective Landscape." *medRxiv : The Preprint Server for Health Sciences*, March. <https://doi.org/10.1101/2021.02.23.21252268>.
- May, R. M., and R. M. Anderson. 1987. "Transmission Dynamics of HIV Infection." *Nature* 326 (6109): 137–42.
- McCormick, Kevin D., Jana L. Jacobs, and John W. Mellors. 2021. "The Emerging Plasticity of SARS-CoV-2." *Science* 371 (6536): 1306–8.
- Metcalf, C. J. E., R. B. Birger, S. Funk, R. D. Kouyos, J. O. Lloyd-Smith, and V. A. A. Jansen. 2015. "Five Challenges in Evolution and Infectious Diseases." *Epidemics* 10 (March): 40–44.
- Metcalf, C. Jessica E., Bryan T. Grenfell, and Andrea L. Graham. 2020. "Disentangling the Dynamical Underpinnings of Differences in SARS-CoV-2 Pathology Using within-Host Ecological Models." *PLoS Pathogens* 16 (12): e1009105.
- Metcalf, C. Jessica E., and Justin Lessler. 2017. "Opportunities and Challenges in Modeling Emerging Infectious Diseases." *Science* 357 (6347): 149–52.
- Miller, Ian F., and C. Jessica Metcalf. 2019. "Vaccine-Driven Virulence Evolution: Consequences of Unbalanced Reductions in Mortality and Transmission and Implications for Pertussis Vaccines." *Journal of the Royal Society, Interface / the Royal Society* 16 (161): 20190642.
- Mina, Michael J., C. Jessica E. Metcalf, Adrian B. McDermott, Daniel C. Douek, Jeremy Farrar, and Bryan T. Grenfell. 2020. "A Global Immunological Observatory to Meet a Time of Pandemics." *eLife*. <https://doi.org/10.7554/elife.58989>.
- Miotto, Olivo, Jacob Almagro-Garcia, Magnus Manske, Bronwyn Macinnis, Susana Campino, Kirk A. Rockett, Chanaki Amaratunga, et al. 2013. "Multiple Populations of Artemisinin-Resistant Plasmodium Falciparum in Cambodia." *Nature Genetics* 45 (6): 648–55.
- Mollentze, Nardus, Daniel G. Streicker, Pablo R. Murcia, Katie Hampson, and Roman Biek. 2020. "Virulence Mismatches in Index Hosts Shape the Outcomes of Cross-Species Transmission." *Proceedings of the National Academy of Sciences of the United States of America* 117 (46): 28859–66.
- Olival, Kevin J., Parvies R. Hosseini, Carlos Zambrana-Torrel, Noam Ross, Tiffany L. Bogich, and Peter Daszak. 2017. "Erratum: Host and Viral Traits Predict Zoonotic Spillover from Mammals." *Nature* 548 (7669): 612.
- Phillips, Maile T., Katharine A. Owers, Bryan T. Grenfell, and Virginia E. Pitzer. 2020. "Changes in Historical Typhoid Transmission across 16 U.S. Cities, 1889-1931: Quantifying the Impact of Investments in Water and Sewer Infrastructures." *PLoS Neglected Tropical Diseases* 14 (3): e0008048.
- Plowright, Raina K., Colin R. Parrish, Hamish McCallum, Peter J. Hudson, Albert I. Ko, Andrea L. Graham, and James O. Lloyd-Smith. 2017. "Pathways to Zoonotic Spillover." *Nature Reviews. Microbiology* 15 (8): 502–10.
- Redding, David W., Peter M. Atkinson, Andrew A. Cunningham, Gianni Lo Iacono, Lina M. Moses, James L. N. Wood, and Kate E. Jones. 2019. "Author Correction: Impacts of Environmental and Socio-Economic Factors on Emergence and Epidemic Potential of Ebola in Africa." *Nature Communications* 10 (1): 5258.
- Reich, Nicholas G., Craig J. McGowan, Teresa K. Yamana, Abhinav Tushar, Evan L. Ray, Dave Osthus, Sasikiran Kandula, et al. 2019. "Accuracy of Real-Time Multi-Model Ensemble Forecasts for Seasonal Influenza in the U.S." *PLoS Computational Biology* 15 (11): e1007486.

- Richardson, Eugene T., Momin M. Malik, William A. Darity Jr, A. Kirsten Mullen, Michelle E. Morse, Maya Malik, Aletha Maybank, et al. 2021. "Reparations for Black American Descendants of Persons Enslaved in the U.S. and Their Potential Impact on SARS-CoV-2 Transmission." *Social Science & Medicine* 276 (May): 113741.
- Saad-Roy, Chadi M., Sinead E. Morris, C. Jessica E. Metcalf, Michael J. Mina, Rachel E. Baker, Jeremy Farrar, Edward C. Holmes, et al. 2021. "Epidemiological and Evolutionary Considerations of SARS-CoV-2 Vaccine Dosing Regimes." *Science*, March. <https://doi.org/10.1126/science.abg8663>.
- Shaw, Liam P., Alethea D. Wang, David Dylus, Magda Meier, Grega Pogacnik, Christophe Dessimoz, and François Balloux. 2020. "The Phylogenetic Range of Bacterial and Viral Pathogens of Vertebrates." *Molecular Ecology* 29 (17): 3361–79.
- Simon, A. Katharina, Georg A. Hollander, and Andrew McMichael. 2015. "Evolution of the Immune System in Humans from Infancy to Old Age." *Proceedings. Biological Sciences / The Royal Society* 282 (1821): 20143085.
- Tatem, Andrew J., Simon I. Hay, and David J. Rogers. 2006. "Global Traffic and Disease Vector Dispersal." *Proceedings of the National Academy of Sciences of the United States of America* 103 (16): 6242–47.
- Thompson, R. N., C. A. Gilligan, and N. J. Cunniffe. 2020. "Will an Outbreak Exceed Available Resources for Control? Estimating the Risk from Invading Pathogens Using Practical Definitions of a Severe Epidemic." *Journal of the Royal Society, Interface / the Royal Society* 17 (172): 20200690.
- Thompson, Robin N. 2020. "Novel Coronavirus Outbreak in Wuhan, China, 2020: Intense Surveillance Is Vital for Preventing Sustained Transmission in New Locations." *Journal of Clinical Medicine Research* 9 (2). <https://doi.org/10.3390/jcm9020498>.
- Thompson, Robin N., Christopher A. Gilligan, and Nik J. Cunniffe. 2016. "Detecting Presymptomatic Infection Is Necessary to Forecast Major Epidemics in the Earliest Stages of Infectious Disease Outbreaks." *PLoS Computational Biology* 12 (4): e1004836.
- Thompson, Robin N., Katri Jalava, and Uri Obolski. 2019. "Sustained Transmission of Ebola in New Locations: More Likely than Previously Thought." *The Lancet Infectious Diseases*.
- Vaughan, Timothy G., Gabriel E. Leventhal, David A. Rasmussen, Alexei J. Drummond, David Welch, and Tanja Stadler. 2019. "Estimating Epidemic Incidence and Prevalence from Genomic Data." *Molecular Biology and Evolution* 36 (8): 1804–16.
- Vaughan, Timothy G., Jérémie Sciré, Sarah A. Nadeau, and Tanja Stadler. n.d. "Estimates of Outbreak-Specific SARS-CoV-2 Epidemiological Parameters from Genomic Data." <https://doi.org/10.1101/2020.09.12.20193284>.
- Visscher, Peter M., Naomi R. Wray, Qian Zhang, Pamela Sklar, Mark I. McCarthy, Matthew A. Brown, and Jian Yang. 2017. "10 Years of GWAS Discovery: Biology, Function, and Translation." *American Journal of Human Genetics* 101 (1): 5–22.
- Wells, Konstans, Serge Morand, Maya Wardeh, and Matthew Baylis. 2020. "Distinct Spread of DNA and RNA Viruses among Mammals amid Prominent Role of Domestic Species." *Global Ecology and Biogeography: A Journal of Macroecology* 29 (3): 470–81.
- Wesolowski, Amy, Caroline O. Buckee, Kenth Engø-Monsen, and C. J. E. Metcalf. 2016. "Connecting Mobility to Infectious Diseases: The Promise and Limits of Mobile Phone Data." *The Journal of Infectious Diseases* 214 (suppl_4): S414–20.
- Wesolowski, Amy, Amy Winter, Andrew J. Tatem, Taimur Qureshi, Kenth Engø-Monsen, Caroline O. Buckee, Derek A. T. Cummings, and C. Jessica E. Metcalf. 2018. "Measles Outbreak Risk in Pakistan: Exploring the Potential of Combining Vaccination Coverage and Incidence Data with Novel Data-Streams to Strengthen Control." *Epidemiology and Infection* 146 (12): 1575–83.
- Wiens, Kirsten E., Pinyi Nyimol Mawien, John Rumunu, Damien Slater, Forrest K. Jones, Serina Moheed, Andrea Caflish, et al. 2021. "Seroprevalence of Anti-SARS-CoV-2 IgG Antibodies

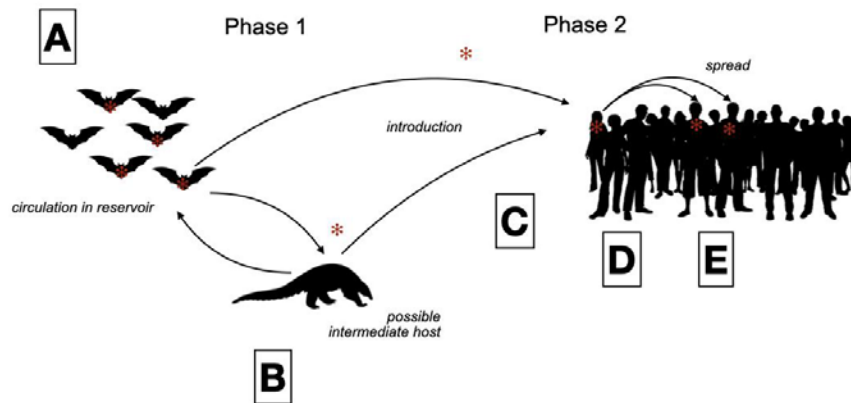
in Juba, South Sudan: A Population-Based Study.” *medRxiv : The Preprint Server for Health Sciences*, March. <https://doi.org/10.1101/2021.03.08.21253009>.

Wikramaratna, Paul S., Adam Kucharski, Sunetra Gupta, Viggo Andreasen, Angela R. McLean, and Julia R. Gog. 2015. “Five Challenges in Modelling Interacting Strain Dynamics.” *Epidemics* 10 (March): 31–34.

Wille, Michelle, Mang Shi, Marcel Klaassen, Aeron C. Hurt, and Edward C. Holmes. 2019. “Virome Heterogeneity and Connectivity in Waterfowl and Shorebird Communities.” *The ISME Journal* 13 (10): 2603–16.

Worldpop. n.d. “WorldPop.” Accessed February 26, 2021. <https://www.worldpop.org/>.

Figure 1: Successful pathogen emergence in a population requires introduction from the reservoir population (here shown as bats, but might be a range of species, including humans where local emergence is being investigated) (phase 1) and then spread within the new host population (phase 2), potentially also via an intermediate host. Data challenges for models around these processes encompass identifying desirable data-streams, and identifying the best ways to leverage existing data. The table below classifies data and stages.



Stage or aspect of zoonotic emergence	Relevant data	Modelling challenges
A Environment and reservoir dynamics	Pathogen sequences from reservoir; pathogen incidence, prevalence, and serology; reservoir host abundance and behavior, including contact mixing and migration	Improving risk assessment in light of diverse and often unknown immunological processes in non-human hosts
B Intermediate species	Farm surveillance data (e.g., pathogen sequences, serological surveys); satellite imagery (density of domesticated animal populations); qualitative data on human-animal (and interspecific animal-animal) interactions across social and ecological contexts	Risk assessment and prioritization of potential mechanisms of spillover
C Spillover	Satellite imagery (human-wildlife interface; human population density); spillover case studies; ecological traits	Risk assessment and prioritization of potential sources of spillover (pathogens and reservoir

	of potential reservoir species	hosts)
D Early-stage emergence (i.e., singletons and stuttering chains)	Contact tracing, outbreak line lists, and other epidemiological data; case studies and histories of detected outbreaks; surveillance data (syndromic, active case-finding or serological surveys, passive reporting, etc)	Improving detection of small and early-stage outbreaks; correcting surveillance biases for risk assessment and prioritization; estimating key parameters from early data
E Secondary transmission and spread	Human mobility and contact mixing (e.g., mobile phones); pathogen sequences; socioeconomic data (e.g., income, racial and other types of marginalization, baseline well-being, access to care); pathogen incidence, prevalence, and serology; clinical data (e.g., signs and symptoms, mortality)	Estimating key parameters for effective risk assessment and forecasting; assessment and prioritization of interventions (including pharmaceutical, non-pharmaceutical, and social-structural)