



# Intensity and frequency of extreme novel epidemics

Marco Marani<sup>a,b,c,1</sup>, Gabriel G. Katul<sup>b,c</sup>, William K. Pan<sup>b</sup>, and Anthony J. Parolari<sup>d</sup>

<sup>a</sup>Department of Civil and Environmental Engineering, University of Padova, 35122 Padova, Italy; <sup>b</sup>Nicholas School of the Environment, Duke University, Durham, NC 27710; <sup>c</sup>Department of Civil and Environmental Engineering, Duke University, Durham, NC 27708; and <sup>d</sup>Department of Civil, Construction, and Environmental Engineering, Marquette University, Milwaukee, WI 53233

Edited by Simon Asher Levin, Princeton University, Princeton, NJ, and approved July 15, 2021 (received for review March 21, 2021)

**Observational knowledge of the epidemic intensity, defined as the number of deaths divided by global population and epidemic duration, and of the rate of emergence of infectious disease outbreaks is necessary to test theory and models and to inform public health risk assessment by quantifying the probability of extreme pandemics such as COVID-19. Despite its significance, assembling and analyzing a comprehensive global historical record spanning a variety of diseases remains an unexplored task. A global dataset of historical epidemics from 1600 to present is here compiled and examined using novel statistical methods to estimate the yearly probability of occurrence of extreme epidemics. Historical observations covering four orders of magnitude of epidemic intensity follow a common probability distribution with a slowly decaying power-law tail (generalized Pareto distribution, asymptotic exponent = -0.71). The yearly number of epidemics varies ninefold and shows systematic trends. Yearly occurrence probabilities of extreme epidemics,  $P_y$ , vary widely:  $P_y$  of an event with the intensity of the "Spanish influenza" (1918 to 1920) varies between 0.27 and 1.9% from 1600 to present, while its mean recurrence time today is 400 y (95% CI: 332 to 489 y). The slow decay of probability with epidemic intensity implies that extreme epidemics are relatively likely, a property previously undetected due to short observational records and stationary analysis methods. Using recent estimates of the rate of increase in disease emergence from zoonotic reservoirs associated with environmental change, we estimate that the yearly probability of occurrence of extreme epidemics can increase up to threefold in the coming decades.**

epidemics | extremes | infectious diseases

Long-term observations and analysis tools to investigate non-stationary processes are available in several disciplines (1, 2). However, extensive epidemiological information at the global scale remains fragmented and virtually unexplored from this perspective, leading to a lack of analyses attempting to reconcile observations of a heterogeneous past. The objectives of this work are to identify the emergent features of the probability distribution of epidemic intensities and to quantify the probability of occurrence of extreme epidemics by assembling and analyzing a global historical dataset. This long historical record of infectious disease epidemics (1600 to present) was assembled from an extensive literature (3–9) and includes 476 documented infectious disease epidemics (217 epidemics with known occurrence, duration, and number of deaths, 145 known to have caused less than 10,000 deaths, and 114 for which only occurrence and duration are known; see ref. 8). Epidemics are defined according to an independence criterion: 1) Individual epidemics of the same disease may not overlap in time: an epidemic cannot end in the same year that marks the start of a subsequent epidemic of the same disease, irrespective of their occurrence location. Epidemics recorded in the literature that occurred in the same period of time were merged into a single epidemic (e.g., several plague events in Europe in the 17th and 18th centuries). This first independence condition ensures that the analyses only focus on epidemics associated with a new or reemerging pathogen after a previous epidemic has ended in the human population (e.g., due to the reemergence of zoonoses from a natural reservoir).

The composition of the dataset, in terms of the primary reemerging diseases and of disease types, is summarized in *SI Appendix*. We subsequently further selected epidemics to be analyzed by the following additional criteria: 2) epidemics were considered only if they are not currently active (e.g., AIDS/HIV, malaria, and COVID-19 were excluded), and 3) epidemics that were ended by the introduction of vaccines or effective treatments were excluded. This last condition, together with the difficulty of determining how some epidemics were ended at a global scale, led to the exclusion of all epidemics occurring after the end of World War II in 1945. Conditions two and three ensure that the disease dynamics are governed by the properties of the pathogen and by transmission dynamics (susceptible-infected interactions possibly mediated by vectors), unaffected by treatments or interventions. In summary, the 1600 to 1945 dataset includes 182 epidemics with known occurrence, duration, and number of deaths, 108 known to have caused less than 10,000 deaths, and 105 for which only occurrence and duration are recorded, for a total of 395 epidemics.

## Results

**The Probability Distribution of Epidemic Intensity.** The empirical exceedance frequency distribution of epidemic intensity is well described by a generalized Pareto distribution (GPD, Fig. 1) over almost four orders of magnitude of the independent variable. The GPD notably exhibits a power-law tail, which signals the absence of a characteristic epidemic intensity and a slowly decaying probability of intense epidemics (10). The fitted GPD is characterized by a power-law tail exponent  $\alpha = -0.71$  approximately for  $i > 3 \times 10^{-2}$  %o/year (Fig. 1), and is robust with respect to the uncertainty characterizing historical accounts of

## Significance

Estimates of the probability of occurrence of intense epidemics based on the long-observed history of infectious diseases remain lagging or lacking altogether. Here, we assemble and analyze a global dataset of large epidemics spanning four centuries. The rate of occurrence of epidemics varies widely in time, but the probability distribution of epidemic intensity assumes a constant form with a slowly decaying algebraic tail, implying that the probability of extreme epidemics decreases slowly with epidemic intensity. Together with recent estimates of increasing rates of disease emergence from animal reservoirs associated with environmental change, this finding suggests a high probability of observing pandemics similar to COVID-19 (probability of experiencing it in one's lifetime currently about 38%), which may double in coming decades.

Author contributions: M.M., G.G.K., W.K.P., and A.J.P. designed research; M.M. performed research; M.M. analyzed data; and M.M., G.G.K., W.K.P., and A.J.P. wrote the paper.

The authors declare no competing interest.

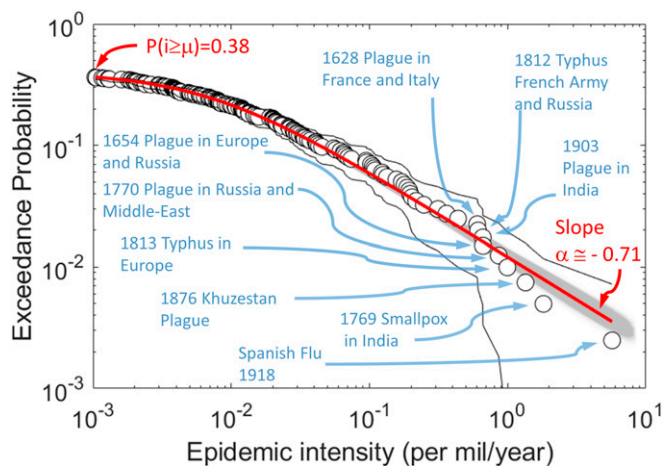
This article is a PNAS Direct Submission.

Published under the PNAS license.

<sup>1</sup>To whom correspondence may be addressed. Email: marco.marani@unipd.it.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2105482118/-DCSupplemental>.

Published August 23, 2021.



**Fig. 1.** Empirical exceedance frequency of epidemic intensity  $i$  (open circles). Black solid lines show the 95% CI around these empirical frequencies (29). The red line is the GPD distribution obtained from maximum likelihood fitting for  $i \geq \mu = 1.000 \times 10^{-3}$  %/year ( $\mu$  being the position parameter, scale parameter  $\sigma = 0.0113$  %/year, and shape parameter  $\xi = 1.40$ ). The value  $P(i \leq \mu) = 0.62$  is determined from the number of observed intensities below  $\mu$  (244 out of 395, including epidemics—105—for which the number of deaths is not available, but historical information suggests  $i \leq \mu$ ). The GPD, for large values of its argument, becomes a power law with exponent  $\alpha = -1/\xi \cong -0.71$ . The value  $|\alpha| < 1$  denotes a fat-tail behavior in which the probability of intense events decreases slowly with event intensity. The gray area results from the overlap of the 10,000 GPD distributions fitted to sample realizations obtained by applying to each observed intensity a random perturbation uniformly distributed in  $[-50\%, +50\%]$  to account for uncertainties in historical records.

epidemic sizes and durations. The collapse of observed epidemic intensities onto a single distribution for the wide diversity of diseases involved and over such a long observational period supports its general validity over time and irrespective of detailed disease dynamics or pathogen characteristics. Hence, this probability distribution of epidemic intensity is assumed here to be time independent, while the rate of disease emergence is allowed to vary to reflect observations (SI Appendix, Fig. S1A).

**The Probability of Occurrence of Extreme Epidemics.** The conventional theory of extremes, as most often applied (11, 12), assumes the process of event occurrence to be stationary: in this interpretation, epidemic event occurrence is governed by a constant rate. Furthermore, it assumes the number of events/year to be “large,” that is, it is asymptotically valid in the limit as the number of events/year  $\rightarrow \infty$ . Neither of these two “mathematical conveniences” is tenable here. The largest number of events in a single year is 12, the variation of the yearly number of events in the 345 y analyzed is ninefold, and the time series exhibits coherent temporal patterns (SI Appendix, Fig. S1). Here, we assume the probability distribution of epidemic intensity to remain the same, as suggested by the analysis in Fig. 1, but we allow the epidemic occurrence process to vary over time through the use of the recent Metastatistical Extreme Value Distribution (MEVD), which relaxes the two above limitations (13). The MEVD expresses the cumulative distribution function of the maximum epidemic intensity occurred within a time interval of 1 y as  $P_T(i) = \langle P(i)^n \rangle$ , where brackets represent ensemble averaging,  $n$  is the number of epidemics that occur during a 1-y period (the values of  $n$  are generated by a nonstationary random process according to observations—SI Appendix, Fig. S1A), and  $P(i) = 1 - H(i)$  is the cumulative probability of epidemic intensity. According to the MEVD, the function  $P_T(i)$  is simply computed by approximating ensemble averaging as a sample mean based on knowledge of  $P(i)$

and of the number of epidemics,  $n_i$ , that occurred in each year:  $P_T(i) \cong (1/L) \cdot \sum_i P(i)^{n_i}$ . In order to determine how  $P_T(i)$  may vary

over time, the sum is extended over all years in time windows of fixed length  $L$ , sliding with no overlap over the time series (here,  $L = 20$  y, a compromise between resolving short time scale variability in epidemic occurrence and robust statistical estimation: results from values of  $L$  in the 10- to 30-y interval are consistent with those obtained with  $L = 20$  y; SI Appendix).

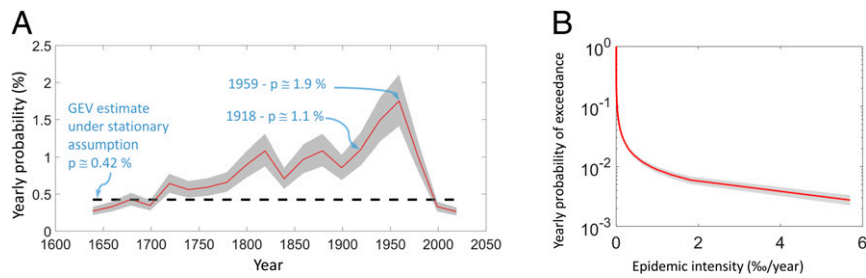
The exceedance probability of the yearly maximum epidemic intensity,  $H_T(i) = 1 - P_T(i)$ , expresses the likelihood that an extreme novel epidemic (irrespective of the specific disease responsible for it), with intensity equal to or greater than  $i$ , occurs anywhere in the world in a given year. As an example, we consider an event with an intensity equivalent to that of the 1918 to 1920 “Spanish flu,” whose yearly probability of exceedance,  $H_T(i = 5.7 \text{ %/year})$ , is plotted in Fig. 2A for nonoverlapping 20-y time periods up to 2019 (for the most recent periods after 1945,  $H_T(i)$  is constructed using the general GPD epidemic intensity distribution, as previously, and the observed number of epidemic occurrences recorded yearly). The values of  $H_T(5.7 \text{ %/year})$  show remarkable temporally coherent variability, which sharply contrasts the constant probability that would be obtained from a conventional approach using a generalized extreme value (GEV) distribution. This wide variability is due to large variations in the rate of occurrence of emerging/reemerging infectious diseases over the course of history and points to the importance of this factor in defining the likelihood of infectious diseases to come. The MEVD can be used to infer the yearly probability distribution of epidemic intensity at a specified time, for example, the present. This probability distribution (Fig. 2B) is necessary to assess expected global losses of lives and economic damages and to motivate global coordination and resource mobilization for public health capacity building (14).

A pandemic of an intensity equal to or greater than that of the Spanish flu, which resulted in 20 to 100 million deaths [32 million being an accredited estimate (4)], is considered for illustrating the use of the average recurrence interval  $T(i)$ . This pandemic yielded  $i = 5.7 \text{ %/year}$  and is estimated here to have occurred when its mean recurrence time was  $T(5.7 \text{ %/year}) = 1/0.011 = 91 \text{ y}$  (95% CI is 85 to 101 y). Based on the observed number of epidemics in our dataset from the most recent 20-y period (2000 to 2019), the mean recurrence time of the same intensity today is  $T(5.7 \text{ %/year}) = 1/0.0025 = 400 \text{ y}$  (95% CI is 332 to 489 y). A naive estimate using a stationary GEV assumption yields a lower and constant  $T(5.7 \text{ %/year}) = 235 \text{ y}$  (SI Appendix, Fig. S5).

In addition to large epidemics, a necessary global health focus is on building capacity for early responses to infectious disease outbreaks of smaller proportions (e.g., see the Global Health Security Agenda, <https://ghsagenda.org/>). The MEVD statistics of extreme epidemic intensity does not provide information as to where an epidemic may emerge; however, it does apply to extreme epidemics of relatively smaller intensities as long as they are greater than the value of the location parameter. This value,  $\mu = 0.001 \text{ %/year}$ , considering the current global population, now corresponds to an epidemic event with absolute intensity of about 8,000 deaths/year (e.g., to be compared with the current absolute intensity of the COVID-19 pandemic of 2.5 million deaths/year, see Discussion).

## Discussion

The empirical distribution of epidemic intensity from about 350 y of data follows closely a GPD over about four orders of magnitude. This finding supports the hypothesis that the epidemic dynamics of emerging/reemerging infectious diseases, when not significantly affected by pharmaceutical interventions, display a general statistical behavior characterized by an exceedance probability with a slowly decaying power-law tail. Mechanisms have



**Fig. 2.** (A) Yearly probability of exceedance,  $H_i$  ( $i = 5.7$  %/year), of an epidemic with the same intensity as the Spanish influenza or greater at different times in history (red). The gray area represents the 95% CI computed from 10,000 realizations obtained by randomly perturbing each historical observation with a perturbation in the range  $[-50\%, +50\%]$  (gray area in Fig. 1). Note that fitting a standard, stationary GEV distribution yields a constant and misleadingly low probability of occurrence. (B) Probability of exceedance of maximum yearly epidemic intensity computed on the basis of the number of epidemic occurrences in the most recent 2000 to 2019 period. Gray area represents the 95% CI as in A.

been proposed to explain the possible emergence of power-law epidemic size distributions in short records of single infectious diseases and in small populations. Such mechanisms are often based on Susceptible–Infected–Recovered (SIR) dynamics (15), with formulations that permit parallels with forest fire models (10, 16). However, to our knowledge, the power-law distribution of size has not been connected to the distribution of epidemic intensities, which would require accounting for the multivariate probability distribution of epidemic size and duration. Hence, the possibility that SIR formulations may explain the power-law features observed in global epidemic intensities remains currently unexplored.

While the probability distribution of epidemic intensity exhibits general, time-independent features in our multicentennial dataset, the likelihood of epidemic occurrence is far from constant in time. This is due to the variability in the rate of emergence of infectious diseases and, in our record, to changes in how epidemics are monitored and reported. These latter differences potentially affect mostly the initial parts of the historical record, while estimates of the MEVD probability of extreme epidemic occurrence in recent times, when disease monitoring has been more systematic, remains unaffected.

Recent analyses of small-scale infectious disease emergence events document a significant increase in the yearly rate of emergence in the period 1940 to 2000 (17). Specific mechanisms of increase in the rate of disease emergence have been identified and connected to anthropogenic environmental change as one of the major drivers (18). These effects of anthropogenic environmental change may carry a high price. Using the MEVD model, we find that a tripling of the rate of disease emergence, an increase consistent with the recorded recent changes, implies an approximate tripling of the probability of extreme epidemics,  $H_I(i)$ , with respect to present values. Such a change would bring, possibly over decadal time scales, the average recurrence interval of a Spanish flu–like event down to 127 y (95% CI 115 to 141 y), comparable to the value it had around 1918 (i.e., 91 y).

Our analysis also quantifies how frequently a COVID-19–like event may occur in the future. Current information (19) indicates that the epidemic progresses at a rate of about 2.5 million deaths/year (3,549,710 in 72 wk), which, normalized by the global population, corresponds to an intensity of the epidemic of 0.33 %/year. Using the number of epidemic occurrences observed in the past 20 y (i.e., 2000 to 2019) in the MEVD model, this intensity corresponds to an average recurrence time of 59 y (95% CI 55 to 64 y). This value is much lower than intuitively expected. However, in many countries, drastic nonpharmaceutical interventions, contact tracing, and quarantine have significantly reduced the number of deaths that could have otherwise occurred. Detailed modeling work suggests that unconstrained epidemic spread would have led to as much as eight times the number of deaths that actually occurred in some countries (20). Assuming

this amplification factor, one obtains an intensity of 2.63 %/year, which corresponds to an average recurrence time of 209 y (95% CI 182 to 244 y). To better appreciate the significance of this value, it may be useful to compute the probability of experiencing an event of this intensity in one’s lifetime (here taken, for simplicity, equal to 100 y), when a constant likelihood is assumed: this probability is

$$P_{Life} = 1 - \left(1 - \frac{1}{209}\right)^{100} = 0.38.$$

Assuming a tripling of the rate of disease emergence, as suggested by the evidence discussed above, this probability may increase to

$$P_{Life} = 1 - \left(1 - \frac{3}{209}\right)^{100} = 0.76.$$

These probability values should be a sufficient warning of the urgency of global preparedness to future pandemic events.

## Materials and Methods

A fundamental characteristic of an epidemic is the number of fatalities per unit time. It is this property that determines how well health care systems cope with epidemics and the socioeconomic damages that are caused by it. For this reason, we define and study the *epidemic intensity*,  $i = s/[d \times S_0(t)]$  (expressed in % fatalities per year), where  $s$  is the total fatalities in an epidemic,  $S_0(t)$  is the size of the global population at the beginning of the epidemic, and  $d$  is the duration of the epidemic. The study of the intensity, rather than of the total number of deaths in an epidemic, is preferred because the former is unbounded. Hence, intensity eliminates the need for ad hoc assumptions near the upper bound of  $s$ ,  $S_0(t)$  (e.g., discussed in ref. 12). To compute  $S_0(t)$  and epidemic intensity, we used reconstructions of the global population history (21–25) (see data in ref. 8). Reliable quantitative information about the total number of deaths is not available for all known historical epidemics. Such quantitative information becomes more frequent with the introduction of public records in many parts of the world, starting in the 17th century. For this reason, we focused on epidemics that occurred in the period 1600 to 1945, a total of 395 events. Quantitative information about duration and the number of fatalities was available for 182 of these epidemics.

The computation of the yearly probability of occurrence of extreme epidemics requires two pieces of information. The first is the probability distribution of epidemic intensity given that an epidemic has indeed occurred. We study this distribution based on the 182 observed epidemic intensities in 1600 to 1945. The second piece of information is the probability distribution of the number of epidemics occurring in a given year. We expect this distribution to vary over time due to varying human–environment interactions and their decisive effect on the rate of emergence of novel epidemics (17, 26–28). The varying number of yearly epidemic emergences is analyzed using all 395 known outbreaks in the 1600 to 1945 record (SI Appendix, Fig. S1).

The probability distribution of the maximum size among  $n$  epidemics occurring in a prespecified fixed time interval ( $w$ ) is considered. When the sizes of these epidemics are independent within each  $w$ -year block and identically distributed according to the same intensity distribution  $P(i)$ , then  $P_w(i, n) = P(i)^n$ . If  $P(i)$  does not change from one  $w$ -year block to another, because  $n$  is also a random variable, the probability that the maximum epidemic size within a  $w$ -year block is smaller than or equal to  $i$  can be written as



$$P_w(i) = \sum_{n=0}^{\infty} P_w(i, n) \cdot g(n) = \sum_{n=0}^{\infty} P(i)^n \cdot g(n),$$

where  $g(n)$  is the probability distribution of  $n$ . The MEVD is defined by substituting the ensemble average above with its corresponding sample average (13), thereby avoiding making restrictive parametric assumptions on the shape of  $g(n)$ . This substitution results in

$$P_w(i) \cong \frac{1}{N_w} \sum_{j=1}^{N_w} P(i)^{n_j},$$

where  $n_j$  is the number of epidemics that occurred in the  $j$ -th  $w$ -year block, and  $N_w$  is the total number of  $w$ -year blocks in the record. Here, we have used  $w = 1$  y and  $N_w = 20$  y.

The GPD is a three-parameter function defined through its cumulative distribution function (11)

$$P(i) = 1 - \left(1 + \frac{\xi(i - \mu)}{\sigma}\right)^{-1/\xi},$$

with location parameter  $\mu$ , scale parameter  $\sigma$ , and shape parameter  $\xi$ . This expression is only valid for  $i \geq \mu$ . Hence, the choice of the location parameter defines the subset of the data to which the GPD is fitted, and, as a consequence, it should not be defined by optimization or data fitting. For this reason, no accepted general method is available to estimate the GPD location parameter. The location parameter must be viewed as a deliberate choice of which part of the data the statistical GPD model should be able to represent. An analysis of the dependence of maximum likelihood-estimated

scale and shape parameters on the choice of  $\mu$  (SI Appendix, Fig. S6) shows that their values remain invariant within  $0.0005\%/\text{year} < \mu < 0.02\%/\text{year}$ . Estimation uncertainty of the aforementioned two parameters, on the other hand, grows rapidly with increasing  $\mu$  as expected because more and more observations are progressively being censored. We thus adopt the value  $\mu = 0.001\%/\text{year}$ , which, for most of the historical global population values, is near the lower “detection” threshold of 10,000 deaths/year characterizing the dataset and allows a reduced uncertainty in comparison to larger values.  $p_0 = P(i \leq \mu) = 0.62$  is the probability that an epidemic intensity is less than  $\mu$  (obtained from the number of events on record [244 out of 395] with an intensity below this threshold or presumed, according to historical sources, to be below this threshold). The probability of exceedance, valid for all values of  $i$  and plotted in Fig. 1, is given by  $H(i) = 1 - [p_0 + (1 - p_0) \cdot P(i)] = (1 - p_0) \cdot \left(1 + \frac{\xi(i - \mu)}{\sigma}\right)^{-1/\xi}$ . For large values of the argument, that is, indicatively for  $(i - \mu) > \sigma/\xi$ , the exceedance probability  $H(i)$  is approximately a power law with exponent  $-1/\xi$ .

**Data Availability.** The historical epidemics dataset generated in the current study and a MATLAB code that analyzes it are available in the Zenodo repository; DOI: <https://doi.org/10.5281/zenodo.4626111>.

**ACKNOWLEDGMENTS.** M.M. acknowledges support within the Venice 2021 research grant funded by Provveditorato for the Public Works of Veneto, Trentino Alto Adige, and Friuli Venezia Giulia, provided through the concessionary of State Consorzio Venezia Nuova and coordinated by CORILA (Consorzio per il coordinamento delle ricerche inerenti al sistema lagunare di Venezia). W.K.P. acknowledges support from NASA (NNX15AP74G).

1. P. C. D. Milly, et al., Stationarity is dead: Whither water management? *Science* **319**, 573–574 (2008).
2. L. Cheng, A. Aghakouchak, E. Gilleland, R. W. Katz, Non-stationary extreme value analysis in a changing climate. *Clim. Change* **127**, 353–369 (2014).
3. F. Fenner, D. A. Henderson, I. Arita, Z. Jezek, I. D. Ladnyi, “Smallpox and its eradication” in *History of International Public Health*, vol. 6 (World Health Organization, Geneva, 1988), pp. 1371–1409.
4. K. D. Patterson, G. F. Pyle, The geography and mortality of the 1918 influenza pandemic. *Bull. Hist. Med.* **65**, 4–21 (1991).
5. W. McNeill, *Plagues and Peoples* (Anchor Books, 1998).
6. G. C. Kohn, *Encyclopedia of Plague & Pestilence* (Wordsworth, 1999).
7. V. Harding, *The Dead and the Living in Paris and London, 1500–1670* (Cambridge University Press, Cambridge, UK, 2007).
8. M. Marani, G. Katul, W. Pan, A. Parolari, A global epidemics dataset (1500–2020) (2021). *Zenodo*. <https://doi.org/10.5281/zenodo.4626111>. Deposited 28 March 2021.
9. C. Socolovschi, D. Raoult, “Typhus fevers and other rickettsial diseases, historical” in *Encyclopedia of Microbiology (Third Edition)*, M. Schaechter, Ed. (Academic Press, 2009), pp. 100–120.
10. D. Sornette, *Critical Phenomena in Natural Sciences* (Springer, ed. 2, 2006).
11. S. Coles, *An Introduction to Statistical Modeling of Extreme Values* (Springer, London, 2001).
12. P. Cirillo, N. N. Taleb, Tail risk of contagious diseases. *Nat. Phys.* **16**, 606–613 (2020).
13. M. Marani, M. Ignaccolo, A metastatistical approach to rainfall extremes. *Adv. Water Resour.* **79**, 121–126 (2015).
14. W. N. Madhav, B. Oppenheim, M. Gallivan, P. Mulembakani, E. Rubin, “Pandemics: Risks, impacts, and mitigation” in *Disease Control Priorities: Improving Health and Reducing Poverty*, D. T. Jamison et al., Eds. (The International Bank for Reconstruction and Development/The World Bank, 3rd ed., 2017), pp. 1–47.
15. V. A. A. Jansen, et al., Measles outbreaks in a population with declining vaccine uptake. *Science* **301**, 804 (2003).
16. M. Roy, R. D. Zinck, M. J. Bouma, M. Pascual, Epidemic cholera spreads like wildfire. *Sci. Rep.* **4**, 3710 (2014).
17. K. E. Jones et al., Global trends in emerging infectious diseases. *Nature* **451**, 990–993 (2008).
18. P. Daszak, A. A. Cunningham, A. D. Hyatt, Anthropogenic environmental change and the emergence of infectious diseases in wildlife. *Acta Trop.* **78**, 103–116 (2001).
19. COVID-19 data repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. GitHub. <https://github.com/CSSEGISandData/COVID-19>. Accessed 12 July 2021.
20. M. Gatto et al., Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10484–10491 (2020).
21. J. D. Durand, C. McEvedy, R. Jones, *Atlas of World Population History* (Popul. Stud. New York, 1979).
22. J. N. Biraben, An essay concerning mankind’s demographic evolution. *J. Hum. Evol.* **9**, 655–663 (1980).
23. J. Tanton, End of the Migration Epoch? *Soc. Contract* **4**, 162–174 (1994).
24. A. Maddison, *The World Economy: Historical Statistics* (Development Centre Studies, OECD Publishing, Paris, France, 2003).
25. K. Klein Goldewijk, A. Beusen, G. Van Drecht, M. De Vos, The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Glob. Ecol. Biogeogr.* **20**, 73–86 (2011).
26. C. K. Johnson et al., Global shifts in mammalian population trends reveal key predictors of virus spillover risk. *Proc. Biol. Sci.* **287**, 20192736 (2020).
27. R. Gibb et al., Zoonotic host diversity increases in human-dominated ecosystems. *Nature* **584**, 398–402 (2020).
28. D. M. Morens, A. S. Fauci, Emerging pandemic diseases: How we got to COVID-19. *Cell* **182**, 1077–1092 (2020).
29. E. B. Wilson, Probable inference, the law of succession, and statistical inference. *J. Am. Stat. Assoc.* **22**, 209–212. (1927).