JOURNAL
OF
THE ROYAL
SOCIETY

Interface

Global rise in human infectious disease outbreaks

Katherine F. Smith, Michael Goldberg, Samantha Rosenthal, Lynn Carlson, Jane Chen, Cici Chen
and Sohini Ramachandran

Supplementary data	"Data Supplement" http://rsif.royalsocietypublishing.org/content/suppl/2014/10/23/rsif.2014.0950.DC1.html
References	**This article cites 14 articles, 3 of which can be accessed free** http://rsif.royalsocietypublishing.org/content/11/101/20140950.full.html#ref-list-1
Subject collections	Articles on similar topics can be found in the following collections biogeography (16 articles) bioinformatics (51 articles) environmental science (110 articles)
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

Report

Author for correspondence:
Katherine F. Smith
e-mail: katherine_smith@brown.edu

†These authors contributed equally to this study.

Royal Society Publishing

Global rise in human infectious disease outbreaks

Katherine F. Smith[1,†], Michael Goldberg[1], Samantha Rosenthal[2], Lynn Carlson[3], Jane Chen[1], Cici Chen[4,†] and Sohini Ramachandran[1,5,†]

[1]Department of Ecology and Evolutionary Biology, Brown University, 80 Waterman St., Box G-W, Providence, RI 02912, USA
[2]Department of Epidemiology, Brown University, Box G-S121-2, 121 South Main St., Providence, RI 02912, USA
[3]Geological Sciences, Brown University, PO Box 1846 Providence, RI 02912, USA
[4]Department of Biostatistics, Brown University, Box G-S121-7, 121 South Main St., Providence, RI 02912, USA
[5]Center for Computational Molecular Biology, Brown University, Providence, RI 02912, USA

To characterize the change in frequency of infectious disease outbreaks over time worldwide, we encoded and analysed a novel 33-year dataset (1980–2013) of 12 102 outbreaks of 215 human infectious diseases, comprising more than 44 million cases occuring in 219 nations. We merged these records with ecological characteristics of the causal pathogens to examine global temporal trends in the total number of outbreaks, disease richness (number of unique diseases), disease diversity (richness and outbreak evenness) and *per capita* cases. Bacteria, viruses, zoonotic diseases (originating in animals) and those caused by pathogens transmitted by vector hosts were responsible for the majority of outbreaks in our dataset. After controlling for disease surveillance, communications, geography and host availability, we find the total number and diversity of outbreaks, and richness of causal diseases increased significantly since 1980 ($p < 0.0001$). When we incorporate Internet usage into the model to control for biased reporting of outbreaks (starting 1990), the overall number of outbreaks and disease richness still increase significantly with time ($p < 0.0001$), but *per capita* cases decrease significantly ($p = 0.005$). Temporal trends in outbreaks differ based on the causal pathogen's taxonomy, host requirements and transmission mode. We discuss our preliminary findings in the context of global disease emergence and surveillance.

1. Introduction

Understanding the spatial and temporal distribution of novel infectious diseases is among the most important and challenging tasks for the coming century [1–5]. To date, generalizations about global disease trends have primarily been made from two forms of data: counts of endemic diseases present within nations at a single time point (denoted here and in previous studies as 'disease richness') [6–8] and records of first-occurrence pathogen emergence events that have occurred globally over time [5]. These past studies have shown that certain pathogens conform to biogeographic trends similar to those exhibited by non-human taxa (e.g. an inverse relationship between disease richness and latitude) [7], that diseases specific to humans (i.e. contagious only between persons) are uniformly distributed around the world, whereas zoonoses (diseases caused by pathogens that spread from animals to humans) are far more localized in their global distribution [6,9,10] and that zoonoses represent the majority of emerging infectious diseases in the human population [5]. However, past studies lack large-scale spatio-temporal data documenting distributions for many pathogens and this has impeded disease biogeographers from fully characterizing the global disease-scape.

Outbreaks occur when the number of cases of disease increases above what would normally be expected in a defined community, geographical area or season [11]. Outbreaks are documented textually and include spatial and temporal attributes, case data, and information about the causal pathogen, and present an
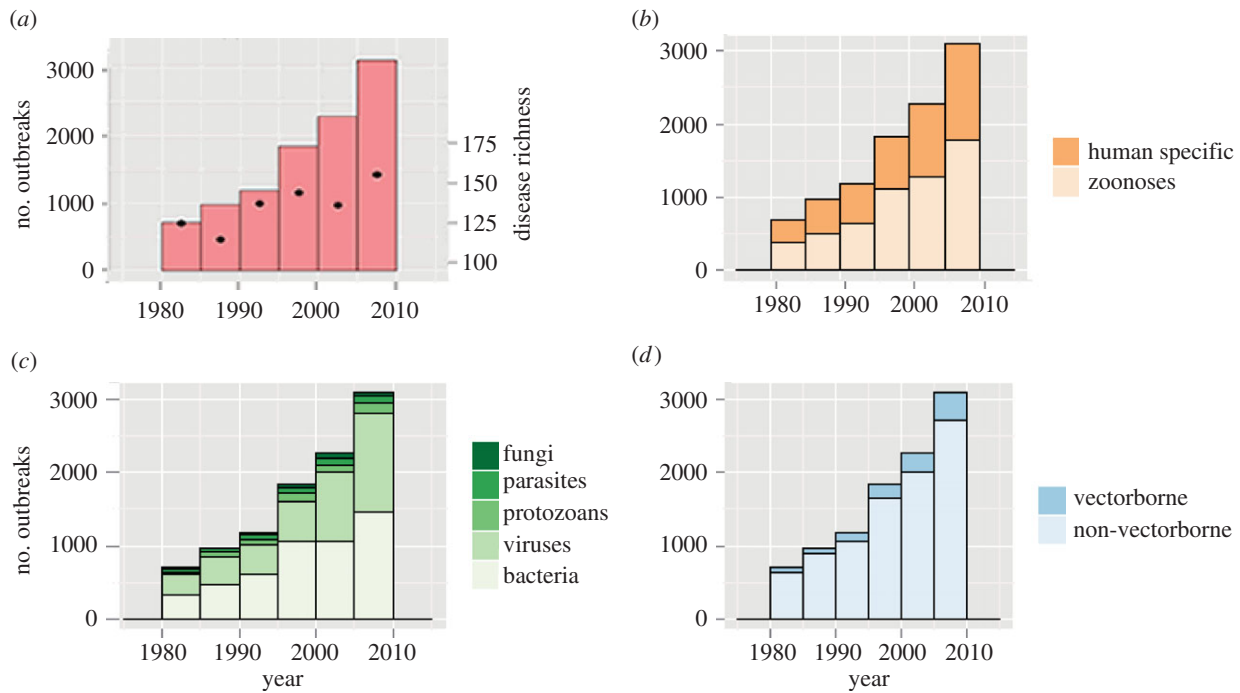
Figure 1. Global number of human infectious disease outbreaks and richness of causal diseases 1980–2010. Outbreak records are plotted with respect to (*a*) total global outbreaks (left axis, bars) and total number of diseases causing outbreaks in each year (right axis, dots), (*b*) host type, (*c*) pathogen taxonomy and (*d*) transmission mode. (Online version in colour.)

opportunity to make new discoveries about global disease trends. However, outbreak records have been largely inaccessible for use in macro-scale analyses as these records are not stored in a manner easily retrievable to the broader research community. The aim of this report is to introduce a new dataset containing data from over 12 000 records of human infectious disease outbreaks and present initial findings from analyses of temporal trends in global outbreaks since the 1980s.

2. Material and methods

We encoded, summarized and analysed a 33-year dataset (1980–2013) of 12 102 outbreaks of 215 human infectious diseases, comprising more than 44 million total cases occurring in 219 nations (table 1). The data are curated as prose records of confirmed outbreaks in the Global Infectious Disease and Epidemiology Online Network (GIDEON) and are accessible via subscription to the site [11]. GIDEON has been used in past macro-scale studies of infectious disease (e.g. [6–8]), but the spatio-temporal data available on outbreaks have not been fully leveraged by researchers because the records are in textual form. We developed a bioinformatics pipeline that automates the parsing and encoding of GIDEON's outbreak records and enables the first macro-scale analyses of global outbreak trends for a suite of unique diseases. We merged these newly encoded records with ecological characteristics of the causal pathogens [6,7] to examine global temporal trends in the total number of outbreaks, richness (total number) of unique causal diseases, diversity (richness in association with outbreak evenness) of causal diseases and *per capita* cases (total cases caused by an outbreak as a proportion of that nation's population in the outbreak year). Because of inherent biases in global disease data (e.g. electronic supplementary material, figure S1, [3,5–7]), we controlled for the effects of six variables previously identified in the literature as confounding the effects of disease occurrence and reporting at large spatial and temporal scales: latitude, GDP, press freedom, Internet usage, population size and population density; all variables were

Table 1. Summary of the human infectious disease outbreak records retrieved by our bioinformatics pipeline.[a]

	total no. diseases (%)	total no. outbreaks (%)
complete dataset	215	12 102
transmission type		
vector transmitted	53 (25)	1518 (13)
non-vector transmitted	159 (74)	10 552 (87)
host type		
zoonotic	139 (65)	6762 (56)
human specific	71 (33)	5300 (44)
taxonomy		
viruses	70 (33)	4901 (40)
bacteria	80 (37)	5848 (48)
parasites	32 (15)	444 (4)
fungi	14 (7)	284 (2)
protozoans	16 (7)	605 (5)

[a]Some diseases could not be assigned to certain sub-categories due to a lack of ecological or epidemiological information [10].

recorded by nation and by year (electronic supplementary material). Using these six confounding variables as the independent variables, we fit quasi-Poisson regression models to identify temporal trends in the number of outbreaks and disease richness. We chose quasi-Poisson models to allow for overdispersion in outbreak occurrences among nation-years. We also used linear regression to model temporal trends in disease diversity and *per capita* cases. The electronic supplementary material fully describes the nature of the GIDEON outbreak records, the pipeline we built to parse and encode them, the ecological characteristics we used to categorize causal pathogens and our statistical methods.
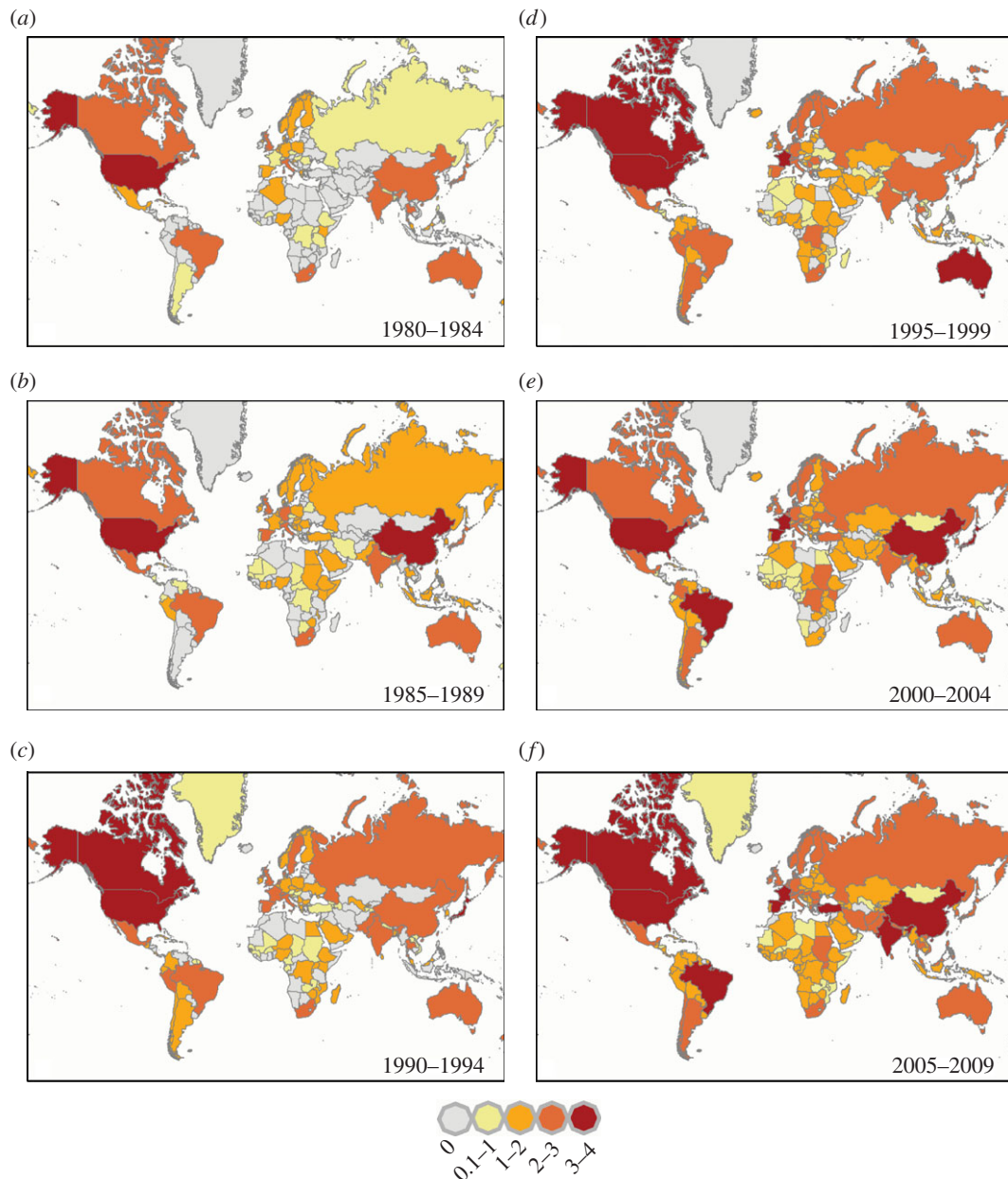
Figure 2. Global outbreak diversity for the nations of the world over time. Diversity is calculated for each nation using Shannon's diversity index (SDI) as described in the methods. Nations with the highest diversity of outbreaks are represented by larger values and darker shading. (Online version in colour.)

3. Results

Our analyses indicate that the total number of outbreaks and richness of causal diseases have each increased globally since 1980 (figure 1a). Bacteria and viruses represented 70% of the 215 diseases in our dataset and caused 88% of outbreaks over time. Sixty-five per cent of diseases in our dataset were zoonoses that collectively caused 56% of outbreaks (compared to 44% of outbreaks caused by human-specific diseases). Non-vector transmitted pathogens were more common (74% of diseases) and caused more outbreaks (87%) than vector transmitted pathogens (table 1). Salmonellosis caused the most outbreaks of any disease in the dataset (855 outbreaks reported since 1980). However, viral gastroenteritis (typically caused by noro-virus) was responsible for the greatest number of recorded cases: more than 15 million globally since 1980.

Previous studies have demonstrated that a nation's like-lihood of experiencing, identifying and reporting an outbreak is influenced by its surveillance capabilities, communication infrastructure, geography and availability of hosts for pathogens

[5,6–9,12–15]. After controlling for these factors using proxies, by nation and year (electronic supplementary material), the number of outbreaks and richness of causal diseases still exhibit a significant increase since 1980 ($p < 0.0001$), as do the number of outbreaks and richness of causal diseases for each sub-category of pathogen taxonomy (bacteria, fungi, parasites, protozoa or viruses), pathogen transmission mode (vector transmitted or non-vector transmitted) and host type (human specific or zoonotic; figure 1b–d; electronic supplementary material, table S1).

The Internet has been shown to significantly improve disease detection and reporting [12–14]. When we add Internet usage as an independent variable to the model (per cent Internet users by nation-year, starting 1990) to control for biased reporting of outbreaks, the overall number of outbreaks and disease rich-ness still increase significantly with time ($p < 0.0001$; electronic supplementary material, table S2). However, the number of protozoan and fungal disease outbreaks, and richness of human-specific, protozoan and fungal diseases do not increase with time in this model (Internet usage included; electronic

Table 2. Top 10 causal diseases in terms of total outbreaks by decade and host type (zoonoses versus human specific).

	1980 – 1990	1990 – 2000	2000 – 2010
zoonoses	total (% by decade)		
anthrax	—	49 (2.8)	169 (5.5)
campylobacterosis	37 (4.2)	44 (2.5)	—
chikungunya	—	—	101 (3.3)
cryptosporidiosis	27 (3.0)	68 (3.8)	—
dengue fever	38 (4.3)	57 (3.2)	150 (4.9)
Escherichia coli diarrhoea	41 (4.6)	180 (10.2)	239 (7.8)
hepatitis A	92 (10.4)	121 (6.8)	178 (5.8)
hepatitis E	28 (3.2)	—	—
influenza A	—	—	209 (6.8)
salmonellosis	102 (11.5)	330 (18.7)	423 (13.7)
shigellosis	64 (7.2)	104 (5.9)	146 (4.7)
trichinosis	39 (4.4)	63 (3.6)	92 (3.0)
tuberculosis	33 (3.7)	127 (7.2)	111 (3.6)
human specific			
adenovirus infection	25 (3.2)	40 (3.2)	—
cholera	41 (5.2)	140 (11.2)	251 (11.0)
enterovirus infection	90 (11.5)	85 (6.8)	175 (7.7)
gastroenteritis (viral)	45 (5.7)	110 (8.8)	381 (16.7)
hepatitis B	—	34 (2.7)	—
legionellosis	58 (7.4)	74 (5.9)	117 (5.1)
malaria	—	39 (3.1)	—
measles	97 (12.4)	124 (9.9)	246 (10.8)
meningitis (bacterial)	40 (5.1)	76 (6.1)	130 (5.7)
mumps	—	—	66 (2.9)
pertussis	—	—	63 (2.8)
rotavirus infection	43 (5.5)	—	67 (2.9)
rubella	25 (3.2)	—	—
typhoid and enteric fever	26 (3.3)	59 (4.7)	106 (4.6)

supplementary material, table S2). Three quarters of the outbreak records reported case data, allowing analysis of global temporal trends in *per capita* cases. After controlling for Internet usage, overall *per capita* cases decrease significantly with time, as do human-specific and protozoan disease outbreaks ($p = 0.005$; electronic supplementary material, table S2).

Measures of disease richness are commonly reported in disease biogeography studies [6–8], but the nature of the outbreak data allows us to quantify, for the first time, global trends in disease diversity. The Shannon diversity index (SDI), common in ecological studies, accounts for both richness or number of unique types (here, unique diseases) and how evenly types are represented in a given dataset (here, across outbreaks) to provide a measure of diversity. Thus, the SDI allows for a new way to examine the global assemblage of infectious diseases (electronic supplementary material). Overall outbreak diversity and the diversity of all sub-categories (by taxonomy, transmission mode and host type) of causal diseases exhibit significant increases since 1980 (figure 2; electronic supplementary material, table S1).

After controlling for Internet usage, this trend disappears ($p = 0.947$), and the diversity of outbreaks caused by human-specific diseases, bacteria, protozoans and fungi exhibit a significant decline since 1990 ($p = 0.023$, 0.034, 0.002, respectively; electronic supplementary material, table S2). By contrast, while diseases caused by pathogens that use non-human hosts to complete their life cycle (e.g. zoonoses and vector transmitted pathogens) exhibit temporal increases in total outbreaks and richness, there is no apparent change in diversity for these diseases (electronic supplementary material, table S2).

Rank-abundance distributions shed some light on this finding. The increasingly long tail of the rank-abundance distribution of outbreaks caused by zoonoses (electronic supplementary material, figure S2) reveals that, while the richness of zoonotic diseases is increasing over time, most of these diseases cause only a small fraction of outbreaks. Indeed, a handful of specific zoonoses appear to cause the majority of outbreaks in each decade: from 1980 to 1990, 80% of all zoonotic disease outbreaks were caused by only

25% of potential zoonoses in the dataset, and only 22% and 21% of zoonoses from 1990 to 2000 and from 2000 to 2010, respectively. Thirteen zoonoses represent the top 10 causal diseases in terms of recorded outbreaks in each of the three decades of the dataset (table 2). The rank-abundance distributions of outbreaks caused by human-specific diseases also reveal dominance by a subset of specific diseases. From 1980 to 1990, 80% of outbreaks caused by human-specific diseases were caused by 31% of all potential human-specific diseases in the dataset (32% and 27% of human-specific diseases from 1990 to 2000 and from 2000 to 2010, respectively). Fifteen human-specific diseases represent the top 10 causal diseases in terms of recorded outbreaks in each of the three decades of the dataset (table 2).

4. Discussion

Here, we analyse human infectious disease outbreaks across the world, spanning multiple decades. Our results provide new descriptions of the global disease-scape and our new dataset, now available for others to use, will help advance the field of disease biogeography.

While outbreaks represent an increase in the number of disease cases beyond expectations for a given population, emerging human infectious diseases are further characterized by novelty: for example, diseases that have undergone recent evolutionary change, entered the human population for the first time, or have been newly discovered [5,9]. The number of outbreaks, like the number of emerging infectious diseases, appears to be increasing with time in the human population both in total number and richness of causal diseases. Although our finding implies that outbreaks are increasing in impact globally, outbreak cases *per capita* appear to be declining over time. Our data suggest that, despite an increase in overall outbreaks, global improvements in prevention, early detection, control and treatment are becoming more effective at reducing the number of people infected [13,16–20].

Temporal trends in outbreaks differ for human-specific diseases versus diseases that rely on non-human hosts. Zoonotic disease outbreaks are increasing globally in both total number and richness but not diversity or *per capita* cases (electronic supplementary material, table S2). Human-specific infectious diseases are also causing an increasing number of outbreaks over time. In contrast to zoonoses, however, human-specific diseases are declining in diversity and in the impact they have through outbreaks (in terms of *per capita* cases). These findings, along with previous work on emerging infectious disease [5], suggest that zoonoses may be increasingly more novel in the global human population when compared with diseases specific to humans. This novelty may be a function of the various ways in which zoonoses occur for the first time in the human population (e.g. spill-over from animals, evolution or discovery) [5,9]. By contrast, human-specific pathogens appear to be less novel (in terms of diversity) and harmful (in terms of *per capita* cases) than in the past. We suspect *per capita* cases for zoonotic outbreaks may indeed be greater than our findings indicate, but this is not detectable due to a lack of communications infrastructure and public health resources in the nations that suffer most from pathogens spilling over to humans from wildlife [5].

The temporal scale of our outbreak dataset allowed us to control for the confounding effects of the Internet (starting in 1990) on the reporting of infectious disease outbreaks. Both the total number of outbreaks and richness of causal diseases increase over time whether we control for Internet usage or not, but temporal trends in diversity and *per capita* cases change direction and significance once Internet usage is controlled for. It is beyond the scope of this report and our current dataset to determine the role the Internet has played in outbreak detection and reporting, but this has been discussed elsewhere by others (e.g. [12–14]). It is becoming increasingly clear that the Internet can improve disease reporting by supplementing formal surveillance with publicly generated digital disease surveillance [12–14]. Because of this, Internet usage and other proxies for national communication infrastructures are important to incorporate into analyses exploring changes in infectious disease over space and/or time.

References

1. Smolinski MS, Hamburg MA, Lederberg J. 2003 *Microbial threats to health: emergence, detection, and response*. Washington, DC: National Academies Press.
2. Cecchine G, Moore M. 2006 *Infectious disease and national security: strategic information needs*. Santa Monica, CA: RAND Corporation.
3. Bonds MH, Dobson AP, Keenan DC. 2012 Disease ecology, biodiversity, and the latitudinal gradient in income. *PLoS Biol.* **10**, e1001456. (doi:10.1371/journal.pbio.1001456)
4. Daszak P. 2012 Anatomy of a pandemic. *Lancet* **380**, 1883–1884. (doi:10.1016/S0140-6736(12)61887-X)
5. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. 2008 Global trends in emerging infectious diseases. *Nature* **451**, 990–993. (doi:10.1038/nature06536)
6. Smith KF, Sax DF, Gaines SD, Guernier V, Guégan J-F. 2007 Globalization of human infectious disease. *Ecology* **88**, 1903–1910. (doi:10.1890/06-1052.1)
7. Guernier V, Hochberg ME, Guégan J-F. 2004 Ecology drives the worldwide distribution of human infectious disease. *PLoS Biol.* **2**, e141. (doi:10.1371/journal.pbio.0020141)
8. Dunn R, Davies JT, Harris NC, Gavin MC. 2010 Global drivers of human pathogen richness and prevalence. *Proc. R. Soc. B* **277**, 2587–2595. (doi:10.1098/rspb.2010.0340)
9. Smith KF, Guégan J-F. 2010 Changing geographic distributions of human pathogens. *Annu. Rev. Ecol. Evol. Syst.* **41**, 231–250. (doi:10.1146/annurev-ecolsys-102209-144634)
10. Wolfe ND, Panosian C, Diamond J. 2007 Origins of major human infectious diseases. *Nature* **447**, 279–283. (doi:10.1038/nature05775)
11. 2013 Global Infectious Disease Epidemiology Network (GIDEON). See http://www.gideononline.com (accessed June 2012–August 2013).
12. Wilson K, Brownstein JS. 2009 Early detection of disease outbreaks using the Internet. *CMAJ* **180**, 829–831. (doi:10.1503/cmaj.1090215)
13. Chan EH *et al*. 2010 Global capacity for emerging infectious disease detection. *Proc. Natl Acad. Sci. USA* **107**, 21 701–21 706. (doi:10.1073/pnas.1006219107)

14. McAlarnen LM. 2012 Internet and free press reduce global lags in outbreak reporting. Professional paper presented in partial completion of MSc degree in Global Health. University of Notre Dame, Department of Biological Sciences.

15. Yang K, LeJeune J, Alsdorf D, Lu B, Shum CK, Liang S. 2012 Global distribution of outbreaks of water-associated infectious diseases. *PLoS Negl. Trop. Dis.* **6**, e1483. (doi:10.1371/journal.pntd.0001483)

16. Morse SS, Mazet JAK, Woolhouse M, Parrish CR, Carroll D, Karesh WB, Zambrana-Torrelio C, Ian Lipkin W, Daszak P. 2012 Prediction and prevention of the next pandemic zoonosis. *Lancet* **380**, 1956–1965. (doi:10.1016/S0140-6736(12)61684-5)

17. 2014 PREDICT: building a global early warning system for emerging diseases that move between wildlife and people. See http://www.vetmed.ucdavis.edu/ohi/predict/index.cfm (accessed 2 January 2014).

18. Madoff LC, Woodall JP. 2005 The Internet and the global monitoring of emerging diseases: lessons from the first 10 years of ProMED-mail. *Arch. Med. Res.* **36**, 724–730. (doi:10.1016/j.arcmed.2005.06.005)

19. Grein TW, Kamara KB, Rodier G, Plant AJ, Bovier P, Ryan MJ, Ohyama T, Heymann DL. 2000 Rumors of disease in the global village: outbreak verification. *Emerg. Infect Dis.* **6**, 97–102. (doi:10.3201/eid0602.000201)

20. World Health Organization. 2005 *International health regulations*, 2nd edn. Geneva, Switzerland: The Organization.