INFECTIOUS DISEASE

Lessons from Ebola: Improving infectious disease surveillance to inform outbreak management

Mark E. J. Woolhouse,[1]* Andrew Rambaut,[1,2] Paul Kellam[3,4]

The current Ebola virus disease outbreak in West Africa has revealed serious shortcomings in national and international capacity to detect, monitor, and respond to infectious disease outbreaks as they occur. Recent advances in diagnostics, risk mapping, mathematical modeling, pathogen genome sequencing, phylogenetics, and phylogeography have the potential to improve substantially the quantity and quality of information available to guide the public health response to outbreaks of all kinds.

INTRODUCTION

The Ebola virus disease (EVD) epidemic in West Africa exemplifies how gaps in capacity for early detection of and rapid response to an infectious disease outbreak can contribute to a public health crisis. Overcoming these gaps is a global public good with benefits that accrue beyond the boundaries of the country first affected (*1*). Surveillance—defined in the 2005 International Health regulations (*2*) as "the systematic, ongoing collection, collation and analysis of data for public health purposes and the timely dissemination of public health information for assessment and public health response as necessary"—is a critical component of outbreak management. Technological advances in diagnostic tools, genome sequencing, computing power, and communications devices can augment traditional surveillance methods to acquire and disseminate information in real time, offering the possibility of better outbreak management and thereby saving lives. Lessons from Ebola as well as other infectious diseases, such as influenza and Middle East respiratory syndrome (MERS), may guide the integration of these technologies for successful disease surveillance (Table 1).

DETECTION AND MONITORING

Most infectious disease outbreaks are first detected through clinical investigation by vigilant frontline health care workers. However, clinical surveillance can be an unreliable tool for outbreak detection and monitoring for a number of reasons. Inadequate surveillance and/or reporting systems, a major issue for EVD in West Africa (*3*), may lead to delayed detection and substantial underreporting. Misdiagnosis, such as the misdiagnosis of sleeping sickness as malaria, may have fatal consequences for the patients concerned (*4*). Also, mild or subclinical cases may not be detected and/or reported to the health services. Such cases accounted for the great majority of infections with pandemic H1N1 influenza in 2009. Indeed, an entirely clinically orientated view can massively underestimate the burden of infection, leading to inaccurate empirical estimates of the scale and trajectory of an outbreak and compromising outbreak management.

One solution to these problems is the development and deployment of rapid, point-of-care (POC) diagnostic tests, linked to modern information technology (*5*). For acute infections, improving detection times by as little as 24 hours, or even less, can make a critical difference to our ability to contain an outbreak (*6*). The focus of POC testing is to generate rapid results that meet WHO "ASSURED" criteria (Affordable, Sensitive, Specific, User-friendly, Rapid & Robust, Equipment-free, and Delivered—to which we would add "Connected"). Suitable platforms are already available for application to a range of viral, bacterial, and protozoal infections. These include nucleic acid amplification techniques that encompass thermal, polymerase chain reaction (PCR)–based tests, isothermal methods (perhaps more suitable for field epidemiology), and enzyme immunoassays and immunochromatographic tests (*5*). Although rapid POC testing has yet to play a substantial role during any major infectious disease epidemic, it is currently being evaluated for dengue and influenza virus and is likely to be increasingly important in the future.

There are also technologies that are of limited use for clinical care but of great value for epidemiological surveillance, providing estimates of cumulative exposure at the population level. One such approach is serosurveillance, the use of serological tests for screening an at-risk population. Serosurveillance has been used to estimate levels of exposure to H5N1 influenza A (*7*) and MERS coronavirus (MERS-CoV) (*8*). Serosurveillance during the H1N1 influenza pandemic gave estimates of 30 to 40% population exposure in many countries (*9*), far higher than clinical surveillance indicated. Protocols for the rapid development and deployment of serological tests have been proposed for influenza (*10*) and could, in principle, be designed for other infections.

Monitoring indirect markers of disease activity, such as Internet use and activity on social media, may also contribute to epidemiological surveillance. However, an early warning system to detect influenza outbreaks (Google Flu Trends) did not detect the arrival of pandemic H1N1 in the United States in 2009, and the challenge for Internet- and social media–based surveillance systems is to develop methods good enough to be used as surrogates for clinical data (*11*). However, other new technologies, such as real-time sequencing and mathematical modeling, may be ready for integrating into surveillance systems.

REAL-TIME SEQUENCE DATA

Probably the most important addition to the arsenal of tools for outbreak investigation and guiding public health interventions is the production and use of time-resolved and geolocated pathogen genome data. Over the past decade, not only has a deep understanding and a detailed evolutionary framework been developed for, in particular, virus genetics (*12*), but powerful computational tools and high-throughput methods for producing virus genomes are now available.

[1]Centre for Immunity, Infection and Evolution, University of Edinburgh, Edinburgh EH9 3FL, UK. [2]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA. [3]Wellcome Trust Sanger Institute, Cambridge CB10 1RQ, UK. [4]Division of Infection & Immunity, University College London, London WC1E 6BT, UK.
*Corresponding author. E-mail: mark.woolhouse@ed.ac.uk

Table 1. Major emerging infectious disease outbreaks in the 21st century.

Pathogen	Disease	Time span	Geographic range	Scale
SARS coronavirus	Severe acute respiratory syndrome	November 2002 to July 2003	29 countries: Australia, Brazil, Canada, China, Colombia, Finland, France, Germany, India, Indonesia, Italy, Kuwait, Malaysia, Mongolia, New Zealand, Philippines, Ireland, Romania, Russia, Singapore, South Africa, South Korea, Spain, Sweden, Switzerland, Thailand, UK, United States, Viet Nam	>8000 cases, 916 deaths
Chikungunya virus	Mosquito-borne viral disease	February 2005 to August 2006	8 countries: Comoros Islands, India, Kenya, Madagascar, Mauritius, Mayotte, Reunion, Seychelles	>1,700,000 cases
H1N1 influenza A virus*	Influenza	April 2009 to August 2010	214 countries worldwide reported laboratory-confirmed cases	Estimated 150,000 to 600,000 deaths, 18,500 laboratory-confirmed
E. coli O104:H4	Enterobacterial infection (can lead to hemolytic uremic syndrome)	May 2011 to July 2011	14 countries: Austria, Canada, Czech Republic, Denmark, France, Germany, Greece, Luxembourg, Netherlands, Norway, Poland, Spain, Sweden, Switzerland	4075 cases, 50 deaths
MERS coronavirus*	Viral respiratory illness	September 2012 to present	23 countries: Algeria, Austria, Egypt, France, Germany, Greece, Iran, Italy, Jordan, Kuwait, Lebanon, Malaysia, Netherlands, Oman, Philippines, Qatar, Saudi Arabia, Turkey, Tunisia, UK, United Arab Emirates, United States, Yemen	1368 laboratory-confirmed cases, 487 deaths (as of 7 July 2015)
H7N9 influenza A virus	Influenza	March 2013 to present	3 countries: China, plus travel-associated cases in Malaysia, Canada	486 laboratory-confirmed cases, 184 deaths (as of 8 February 2015)
Chikungunya virus	Mosquito-borne viral disease	December 2013 to present	47 countries: Widespread in Caribbean and Central America plus Argentina, Bolivia, Brazil, Canada, Chile, Colombia, Guyana, Mexico, Paraguay, Peru, United States, Venezuela	>1,000,000 cases, 176 deaths (as of January 2015)
Zaire ebolavirus*	Hemorrhagic fever	March 2014 to present	9 countries: Guinea, Liberia, Mali, Nigeria, Senegal, Sierra Leone, Spain, UK, United States	28,251 cases, 11,294 deaths (as of 16 September 2015)

*Designated as "public health event of international concern" (PHEIC) according to the International Health Regulations.

Large-scale sequencing has been extensively used as a research tool, especially in the fields of HIV and influenza. HIV sequences for parts of the virus genome conferring drug resistance have been routinely determined as part of clinical patient management for nearly two decades, with peripheral blood samples being taken for virus genome load as well as for determining HIV protease and reverse transcriptase sequences for prediction of likely drug sensitivity or resistance [13]. When organized nationally, such sequences can be linked under appropriate data governance and ethics to other clinical and demographic data. From this, the sequences can inform transmission network analysis [14] and HIV infection dynamics [15]. For influenza viruses, large-scale sequencing of virus isolates, linked to geolocation, provides a rich and detailed insight into global influenza virus transmission both in humans and in animal species [16].

HIV and influenza virus both illustrate that access to and analysis of large numbers of samples (typically hundreds or thousands) is essential. These samples need to be collected without additional sampling of the patient or specialist processing of samples where they are obtained. Fortunately, clinical samples are processed into virus nucleic acid either manually or on robot systems with as little as 20% of the virus nucleic acid used in the diagnostic PCR. It is at this point, when all the costs and logistics associated with diagnosis have been met, that virus genomes can be retrieved from the sample. In short, residual clinical diagnostic nucleic acid should never be discarded before the option of converting to a pathogen genome has been considered.

In practice, full-length virus genomes are not always required. Partial virus genomes that are not "finished" may provide all the information required for molecular epidemiology. A range of genome criteria should be considered in producing high-value or actionable virus genomes [17]. The important addition here is a set of criteria for assessing the quality of the assembled genomes and the desire to limit these criteria to the majority/consensus pathogen genome, rather than the requiring accurate reporting of minority sequence variants in the sample. This strategy has been shown to work in practice for both MERS-CoV and ebolavirus.

ANALYSIS OF SEQUENCE DATA

In recent years, there has been a profusion of methods that link virus gene sequences with other information to reveal the evolutionary and epidemiological dynamics of the virus. One critical set of data is the

dates of sampling of the virus, which transform a phylogeny from a classification procedure into an epidemiological tool. With a time axis, the branching events represent transmissions between hosts and thus the times between these events can be used, in a mathematical model, to learn about the key parameters for the outbreak.

For many infectious disease outbreaks, estimates of the sampling proportion (the proportion of the epidemic the sample viruses represent) may be the most crucial inferences to be made, revealing the extent of the hidden epidemic due to subclinical cases or otherwise unreported cases. Another key motivation for the collection of virus sequence data is to understand the relationship between human cases and an animal reservoir. An important example is MERS-CoV (18), where phylogenetic analysis of virus sequences obtained from camels, particularly camels with no link to a human case, suggests the directionality of transmission from camel to human (19). Virus sequence analysis can also model how the virus spreads through space and time, using individual locations of sampled individuals, at the level of map coordinates and assuming that the movement of the virus is through a process of diffusion (20) or by treating geography as a limited set of discrete locations (for example, cities) and interpreting movements as jumps between them that occur at particular rates (21). These approaches can be equally used to investigate the evolution of phenotypic traits of viruses, such as host switching (19) or virulence, resistance, or the antigenic evolution of influenza (22).

The 2009 H1N1 influenza A pandemic was remarkable for being the first serious outbreak to be tracked in real time by virus genetic data, using the data provided by the U.S. Centers for Disease Control and Prevention (CDC) within days of samples being taken from suspected cases. These data were shared as part of the Global Initiative on Sharing All Influenza Data (GISAID), which had been set up a few years before to encourage the exchange of influenza data. However, no similar initiatives exist for other viruses with epidemic potential.

Virus genome sequencing of the earliest EVD cases from Guinea attributed the outbreak in West Africa to the species ebolavirus within weeks of the first cases being diagnosed (23). The genetic similarity to viruses that had previously caused human outbreaks in Central Africa provided an expectation of the epidemiological and pathological properties of the virus: The Zaire species of ebolavirus had caused 14 documented outbreaks of no more than a few hundred cases but with a case fatality rate of up to 90%. However, even though this was a known virus, the outbreak occurred in an unexpected geographical area and in a population that had a very different demography from previous outbreaks (24).

In June 2014, the Broad Institute in collaboration with partners at the Kenema Government Hospital in Sierra Leone shared 78 virus genome sequences from patients that had presented with EVD in the preceding weeks. These provided information on the rate of evolution and revealed no evidence of virus adaptation to humans (25), a major concern at the time. Another

important finding was that the epidemic was not being driven by multiple zoonotic transfers from an animal reservoir. These sequences provided crucial insights into the virus just at a time when the outbreak was rapidly growing (Fig. 1). The publication of these sequences (25) inspired a series of analytical papers extracting additional inferences about the outbreak including estimates of epidemiological parameters such as the case reproduction rate, infectious period and sampling fraction (26), and the identification of lineages of potential epidemiological significance (27). Estimates of the case reproduction rate (similar to R_0 during the early phase of an outbreak) were broadly in line with epidemiological estimates, providing helpful confirmatory evidence that overcame concerns over the reliability of case reporting data. However, the epidemic had grown to the point where hundreds of cases per week were being reported from the three affected countries by the time these studies were published (in October and November 2014), and the results had limited practical value.

MATHEMATICAL MODELING

Mathematical modeling is an established tool in infectious disease epidemiology (28). Real-time projections of case numbers using mathematical models have been provided during many epidemics in the past three decades, including EVD (29, 30). At a minimum, actionable projections require (i) an appropriate model framework that captures heterogeneities in risks of infection and rates of transmission, (ii) appropriate methods for model parameterization, and (iii) rapid access to infection and disease data. Recent applications of mathematical and statistical models to project the course of the 2014 EVD epidemic provide in-
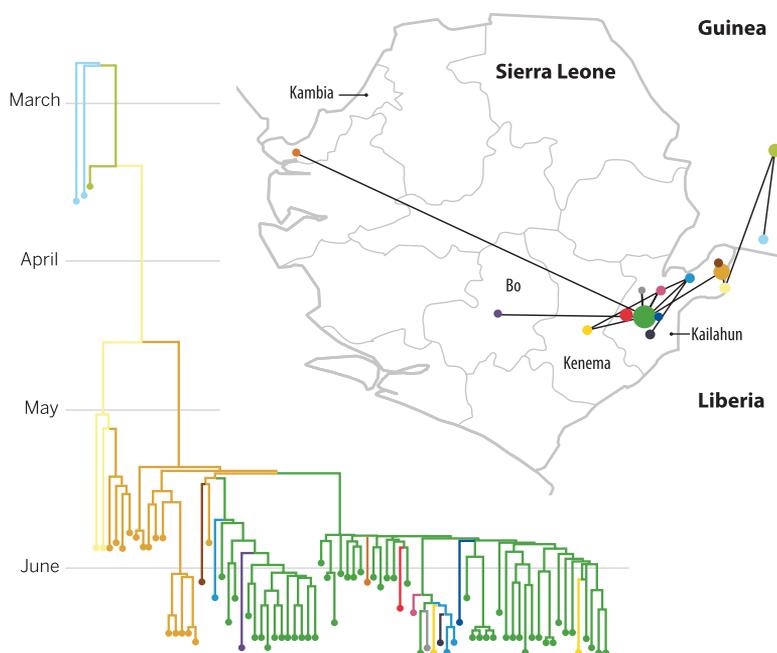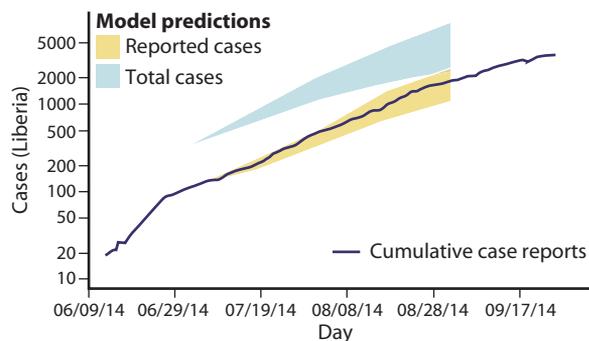
Fig. 1. Time-scaled phylogenetic tree based on ebolavirus sequences [from (25)] from Kenema Government Hospital, Sierra Leone, May and June 2014, plus early samples from Guinea [from (23)]. Branch colors represent probable location of infection with the corresponding locations shown in the inset map. In the map, the radius of the circles denotes the number of sampled sequences, and the lines represent the phylogenetic tree projected onto the map.

structive examples. Two studies (*29*, *30*) were based on the standard compartment model framework (*28*), extended to allow for heterogeneous transmission related to clinical disease, hospitalization, and funerals, and calibrated against early case data. Another study (*31*) fitted both regression and branching process models to clinical case data. A variant of the latter approach incorporated separate probability distributions for different transmission routes, resulting in a multitype branching process model (Fig. 2) (*32*). Together, these make for a set of very different modeling approaches, but all are essentially extrapolations, implicitly assuming near-exponential growth of the epidemic.

Accurate projections depend to a large degree on accurate parameter estimation, not least because exponential processes are highly sensitive to exact parameter values. Two key parameters are R_0 (the average number of secondary cases generated by a single primary case introduced into a previously unexposed population) and the generation time (the average time between initial infection of a case and of cases it gives rise to) (*28*). Together, R_0 and the generation time determine the doubling time of an outbreak during the early, exponential phase. However, exponential growth during the early stages of an outbreak is not expected in all circumstances (*33*), such as when $R_0 \approx 1$, a realistic scenario for which large outbreaks (hundreds of cases or more) are entirely possible (*34*) or when there are multiple introductions separated in time and space, some of which die out due simply to demographic stochasticity. This was the case with pandemic H1N1 influenza A in Scotland in 2009, as indicated by the analysis of virus sequence data (*35*). The history of ebolavirus in Liberia in 2014 may also have involved multiple introductions (*32*), but again, this can only be confirmed from virus genome sequence data.

Although short-term projections are feasible for many outbreaks, extrapolation methods are much less useful in the longer term. This is partly because the confidence intervals on the projections quickly become very wide (Fig. 2), but more fundamentally, because the exponential growth assumption breaks down as the epidemic progresses (often with the introduction of control measures). This underlines the need for clear communication of how model outputs—particularly "worst case" scenarios—should be interpreted (*3*).

Several initiatives [for example, (*36*)] aim to increase the availability of open access tools kits for epidemiological modeling—both for model parameterization and development. Indeed, formal, robust,

and rapid model fitting procedures, generally based on maximum likelihood or Markov chain Monte Carlo (McMC) methods (*37*), are being developed to replace the ad hoc approaches—"calibration" or "tuning"—which are still often used in practice (*33*).

One potentially useful approach is pattern-oriented modeling (POM) (*38*). POM is a technique used originally in ecological modeling both to distinguish between possible model structures and to reduce parameter uncertainty. POM identifies models that reproduce a set of preselected patterns observed in the data—whether qualitative or quantitative. The ability to consider multiple patterns and different kinds of data simultaneously greatly increases both discriminatory power and flexibility. It also addresses a legitimate reluctance to apply very precise model fitting procedures to poor-quality disease data. POM has rarely been applied to infectious diseases (*38*), but a very similar approach has been used to parameterize a model of EVD cases, generating encouragingly precise estimates of a set of seven different parameters (*32*).

RISK MAPPING

Risk mapping has been applied to a range of diseases, including EVD in Africa. The EVD risk map (*24*) incorporated a set of predictors including elevation, an index of vegetation cover, other environmental variables, and estimated composite distribution data for three bat species suspected to be reservoirs of Ebola virus. The output (Fig. 3) suggests that several countries, notably Nigeria and Cameroon, are at risk of EVD but lie outside its currently reported range.

Spatial risk analyses are restricted to predictors for which spatial data are available. In resource-poor settings, this often equates to data available via remote sensing. Analyses are also limited by the quantity and quality of the disease data used to calibrate the models, in particular the issue of ascertainment bias (for example, "pseudo-absence" at locations

Fig. 2. Projected numbers of cases of EVD in Liberia in 2014 obtained using a branching process model with an ensemble of plausible parameter values. The 95% prediction intervals from 4 July 2014 (yellow shading) are compared with the observed cumulative case numbers (logarithmic scale) over the following 2 months (blue line). The 95% prediction intervals for a model that incorporates estimated levels of underreporting are also shown (blue shading). Reproduced with authors' permission from (*32*).
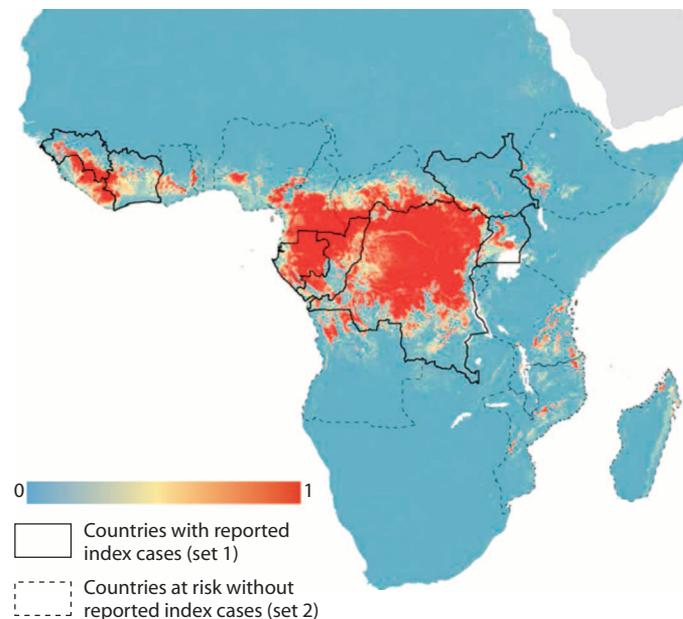

Fig. 3. Predicted probability distribution of zoonotic EVD cases in Africa based on a risk mapping analysis and highlighting at-risk countries with and without index cases reported up to 2014. Blue, low probability; red, high probability. Reproduced with authors' permission from (*24*).

where health reporting is unreliable). The utility of the models is largely determined by how well they deal with this issue. However, even with these limitations, risk maps provide information that helps direct national and international surveillance efforts and contributes to planning and preparedness between outbreaks.

APPLICATIONS OF MODELING

Outbreak size distribution analysis has been successfully used to monitor the epidemiology of measles in the UK after a fall-off in childhood vaccination rates in the late 1990s, charting the approaching loss of herd immunity through shifts in the size and frequency of small outbreaks (39). It has also been applied to monkeypox (40), anticipating a possible increase in monkeypox transmissibility as the fraction of the population immunized against smallpox dwindles. It has recently been applied, using McMC techniques, to EVD (34), confirming that before 2013, R_0 for ebolavirus in humans was close to, or possibly above, 1, indicative of a high risk of major epidemics.

Modeling can aid in predicting the impact of so-called reactive control measures (6) on the course of an infectious disease outbreak. For example, the expected impact of case isolation and/or quarantine of at-risk individuals on the course of an outbreak is determined, inter alia, by the relative timings of a case becoming infectious (and thus potentially transmitting infection to others) and being detected, typically after the appearance of clinical signs (6). Surprisingly, such basic information on the time course of an infection is often lacking, even for well-studied infections such as influenza, and there is a need for greater investment in experimental studies to fill this gap. This example illustrates a wider concern: Many public health interventions are designed to reduce pathogen transmission rates, and neither their intended nor actual impact can be quantified without reference to changes in transmission rates; however, research on pathogen transmission consumes a miniscule fraction of research effort expended on infectious diseases, the bulk of which is aimed at understanding and preventing infection and pathology.

Another consideration, all too often ignored until an outbreak occurs, is the logistic capacity of the affected health system to respond. For the West African EVD epidemic, a key issue was the capacity to roll out isolation units fast enough to "catch up" the epidemic curve (3). However, similar arguments apply more generally to the capacity to administer drugs, vaccines, or any other reactive measures that contribute to reducing the net rate of transmission. In this context, models can help quantify an "effective" response. For EVD, models indicated that hospital capacity and individual behavior (particularly social distancing) were particularly important (32).

Parameterizing the variables that capture both the intended and the actual impact of interventions can be extremely difficult (33). There is a need, first, to monitor the implementation of interventions (noting that targets set by policy-makers do not always correspond to events on the ground) and, second, to analyze these data in real time to evaluate their impact. These activities require resources and are often neglected. Moreover, many of the measures that may be taken have effects, particularly on the rate of transmission, that are difficult to quantify. Examples include the wearing of face masks and social distancing (reducing the risk of infection by changing patterns of contact with the rest of the population, whether in response to public health warnings or through individual initiative).

An important general principle that emerges from the infectious disease modeling literature is that there are substantial benefits arising from the implementation of reactive control measures as early as possible (6). This is a straightforward consequence of the expectation that the absolute numbers of cases will increase exponentially during the early stages of an outbreak. Indeed, during this phase, the costs of delay also increase over time; for an acute infection such as ebolavirus, each week's delay permits a greater number of extra cases than the previous week (3).

PRACTICAL STEPS

The call for better surveillance systems has been made repeatedly in the past decade (41), but there has been too little effective change on the ground (42). One of the most important barriers to the modernization of infectious disease surveillance systems is that nontraditional approaches are all too often seen as an unnecessary distraction from immediate health needs, particularly during an emergency when resources are likely to be severely stretched. This can be exacerbated by real or perceived gaps in technical capacity and expertise (a health emergency is not the best time to be learning new techniques) and by those involved in collecting samples and data (sometimes in extremely challenging circumstances) being disconnected from the subsequent work that depends on their efforts. The best way to remove such barriers to adoption may be to promote a wider appreciation of what is possible, how it can be achieved, and the immediate benefit to public health.

It has been argued that improving global surveillance for emerging infectious diseases is feasible and cost-effective (2), but substantial investment in infrastructure, technology, training, and organization is required. Ultimately, improved global surveillance will emerge from strengthening and connecting national surveillance systems. Similar kinds of investment are needed to strengthen national and international capacity to respond effectively to infectious disease events, and there is an ongoing discussion in the light of the current EVD epidemic as to whether that should include an international rapid response force (1). In addition, there is a need for a greater investment in health policy and systems research, an underfunded and unappreciated field that has a central role to play in meeting the challenge of achieving effective infectious disease surveillance and outbreak management on a global scale.

Any response to an infectious disease outbreak, and especially a coordinated international effort, is contingent not just on the presence of functional national surveillance systems but also on the rapid sharing of information between countries and with international agencies. The revolution in information and communications technology that has occurred over the past 20 to 30 years has removed virtually all technological barriers to this process, even in remote, resource-poor settings. Moreover, as several of the above examples illustrate, it is now routine to integrate and analyze data from multiple sources, such as public health, demographic, location (for example, with global positioning system), movement, geographic, animal distribution, remote sensing, and genome sequence data.

Arguably the biggest remaining barrier to real-time data sharing is cultural, reflecting a reluctance to report disease events. This can be for a number of reasons, not least fear of the imposition of restrictions on freedom of movement or trade or of adverse effects on tourism and

investment (2). The 2005 International Health Regulations provide a framework for disease reporting but do not directly address the question of disincentives, and their implementation has been very patchy to date (42). An obvious solution is to balance the negative consequences of reporting with the promise of effective assistance.

For maximum benefit, data sharing should be as rapid and as open as possible. Again, there are few, if any, technological barriers to this: Data and information sharing platforms such as GenBank, Dryad, and ArXiv have been available for many years. However, although lines of reporting from front-line health officials to international agencies are fairly well set out (2), there is no agreement on responsibility for data sharing, and all too often, this is left to individual or institutional preference. One approach is to penalize countries that do not implement and report from an adequate surveillance system, as was required for participation in the international cattle trade during the bovine spongiform encephalopathy (BSE) epidemic.

One possible consequence of data sharing is a proliferation of analyses of those data, as was seen during the 2009 H1N1 influenza A pandemic and during the current EVD epidemic. Although we regard this as a positive development, it can have perceived disadvantages, notably a loss of control by national or international agencies, and as creating uncertainty over which analyses should be trusted. These issues are not insurmountable and should not be regarded as obstacles to data sharing. In other fields, notably climate change, an ensemble approach to data analysis, interpretation, and projection has been the norm for many years (43). Although this is challenging for many infectious diseases, if only because of the much shorter time scales involved, suitable systems are already in place and there has been, for example, real-time evaluation of multiple models of pandemic influenza (44).

Many of the practical aspects of preparedness for an infectious disease outbreak can and should be addressed in advance of a crisis. These include contingency planning and coordination; developing and stockpiling diagnostics, drugs, and vaccines; setting up sequencing pipelines; designing data sharing protocols; constructing, verifying, and validating mathematical models; agreeing reporting and communication pathways; and anticipating public engagement and ethical issues. One approach to this is to set up sentinel cohorts. This ensures that data collection and reporting (including self-reporting) systems are all in place and tested in advance of an outbreak. It would also cover ethical requirements. Ethical considerations both delayed and limited surveillance in the UK during the 2009 H1N1 influenza A pandemic (45) and can be difficult to deal with rapidly even during a major emergency, as recent experience with trials for ebolavirus vaccines illustrates. The surveillance systems that are set up also need to be flexible and responsive. The infectious disease threat is diverse and dynamic, and periodically presents "out of the blue" challenges

such as BSE/vCJD (variant Creutzfeldt-Jakob disease) in the 1980s or SARS in the 2000s.

To facilitate the provision of virus genomes, we would propose an approach similar to the WHO's "Pandemic Influenza Preparedness Framework for the sharing of influenza viruses and access to vaccines and other benefits" (or PIP Framework), an international arrangement that brings together key stakeholders to strengthen preparedness for the next influenza pandemic. This has now been extended to address sequencing data through a Technical Expert Working Group, with the overall PIP framework encouraging collaborative, transnational working under a framework of a more structured, efficient, and equitable system.

We also need to recognize that managing infectious disease of all kinds is a multidisciplinary problem and, if it is to be done as effectively as possible, requires input from beyond traditional clinical medicine and public health. An integrated, global infectious disease surveillance system needs to take a One Health approach and embrace livestock and wildlife health, as well as geography and environmental sciences, sociology, economics and anthropology, informatics, communications science, and health technology.

CONCLUSIONS

We can readily identify the components of a surveillance system that would enable the collection of infectious disease surveillance data from multiple sources for use as inputs into state-of-the-art epidemiological analysis (Fig. 4). Advances in diagnostics, sequencing platforms, communications technology, and computing and informatics over the past
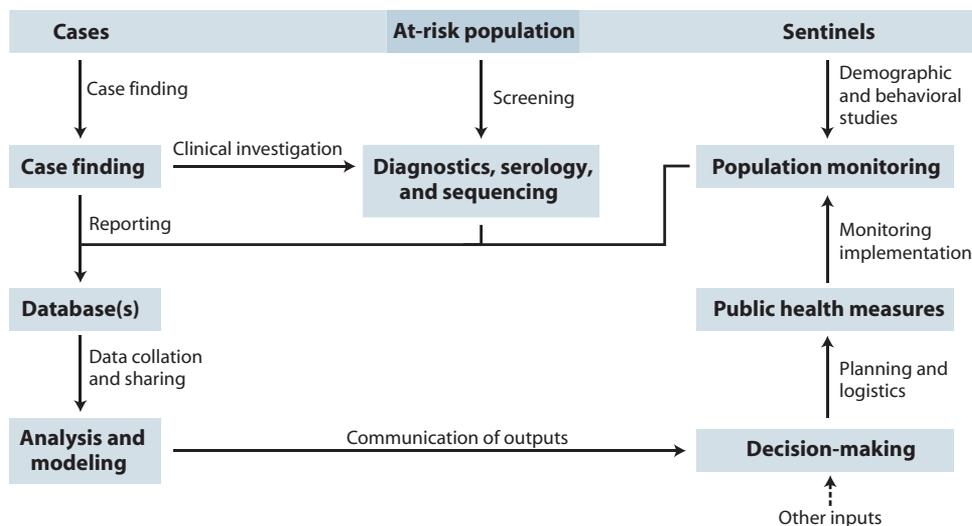
Fig. 4. Key elements of data capture and information flows for real-time quantitative analysis to inform outbreak management. The at-risk population encompasses cases and, where available, a sentinel subpopulation. Three types of data capture activities are identified: case finding, including associated epidemiological investigations such as contact tracing; diagnostic information on individual patients, including serological testing and pathogen sequencing; and so-called denominator studies on the population at risk, including demography, behavior (such as social media activity), and the impact of health measures. Information flows involve communication between data gatherers, data analysts and modelers, policymakers, and public health authorities. We note, however, that decision-making never relies solely on the outputs of real-time epidemiological analyses.

5 to 10 years mean that such analyses can now make an effective contribution to outbreak management in real time. This is a highly significant new capability that we should fully exploit to improve the public health response to future infectious disease outbreaks. A cultural shift is required among health care workers such that these activities come to be regarded as a valuable complement to the clinical care of individual patients, and not as unwelcome competition for resources, time, and effort.

Strengthening surveillance and response capacity around the world would require investment estimated at tens of billions of dollars per annum, but is likely to be cost-effective. Moreover, capacity strengthening should not be the sole responsibility of individual countries; we emphasize that infectious disease surveillance is a global good and should be financed on that basis. We suggest that not all elements of a state-of-the-art surveillance system need to be replicated at a national level; it will often be much more efficient to integrate local activities into an international network. However, this would require considerably more proactive leadership of global surveillance efforts than exists at present. Ultimately, there will be little progress without strong and trusted international governance systems.

REFERENCES AND NOTES

1. M. E. J. Woolhouse, P. Drury, C. Dye, Zero infection. *Science* **346**, 1271 (2014).
2. S. E. Davies, J. R. Youde, *The Politics of Surveillance and Response to Disease Outbreaks: The New Frontier for States and Non-State Actors*, S. E. Davies, J. R. Youde, Eds. (Ashgate Publishing, Surrey, 2015), pp. 9–23
3. C. J. M. Whitty, J. Farrar, N. Ferguson, W. J. Edmunds, P. Piot, M. Leach, S. C. Davies, Infectious disease: Tough choices to reduce Ebola transmission. *Nature* **515**, 192–194 (2014).
4. M. Odiit, P. G. Coleman, W.-C. Liu, J. J. McDermott, E. M. Fèvre, S. C. Welburn, M. E. J. Woolhouse, Quantifying the level of under-detection of *Trypanosoma brucei rhodesiense* sleeping sickness cases. *Trop. Med. Int. Health* **10**, 840–849 (2005).
5. O. Clerc, G. Greub, Routine use of point-of-care tests: Usefulness and application in clinical microbiology. *Clin. Microbiol. Infect.* **16**, 1054–1061 (2010).
6. B. Charleston, B. M. Bankowski, S. Gubbins, M. E. Chase-Topping, D. Schley, R. Howey, P. V. Barnett, D. Gibson, N. D. Juleff, M. E. J. Woolhouse, Relationship between clinical signs and transmission of an infectious disease and the implications for control. *Science* **332**, 726–729 (2011).
7. T. T. Wang, M. K. Parides, P. Palese, Seroevidence for H5N1 influenza infections in humans: Meta-analysis. *Science* **335**, 1463 (2012).
8. R. A. Perera, P. Wang, M. R. Gomaa, R. El-Shesheny, A. Kandeil, O. Bagato, L. Y. Siu, M. M. Shehata, A. S. Kayed, Y. Moatasim, M. Li, L. L. Poon, Y. Guan, R. J. Webby, M. A. Ali, J. S. Peiris, G. Kayali, Seroepidemiology for MERS coronavirus using microneutralisation and pseudoparticle virus neutralisation assays reveal a high prevalence of antibody in dromedary camels in Egypt, June 2013. *Euro Surveill.* **18**, pii=20574 (2013).
9. H. Kelly, H. A. Peck, K. L. Laurie, P. Wu, H. Nishiura, B. J. Cowling, The age-specific cumulative incidence of infection with pandemic influenza H1N1 2009 was similar in various countries prior to vaccination. *PLOS One* **6**, e21828 (2011).
10. K. L. Laurie, P. Huston, S. Riley, J. M. Katz, D. J. Willison, J. S. Tam, A. W. Mounts, K. Hoschler, E. Miller, K. Vandemaele, E. Broberg, M. D. Van Kerkhove, A. Nicoll, Influenza serological studies to inform public health action: Best practices to optimise timing, quality and reporting. *Influenza Other Resp. Viruses* **7**, 211–224 (2013).
11. D. R. Olson, J. J. Konty, M. Paladini, C. Viboud, L. Simonsen, Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: A comparative epidemiological study at three geographic scales. *PLOS Comput. Biol.* **9**, e1003256 (2013).
12. E. C. Holmes, B. T. Grenfell, Discovering the phylodynamics of RNA viruses. *PLOS Comput. Biol.* **5**, e1000505 (2009).
13. S.-Y. Rhee, M. J. Gonzales, R. Kantor, B. J. Betts, J. Ravela, R. W. Shafer, Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* **31**, 298–303 (2003).
14. S. Hué, A. E. Brown, M. Ragonnet-Cronin, S. J. Lycett, D. T. Dunn, E. Fearnhill, D. I. Dolling, A. Pozniak, D. Pillay, V. C. Delpech, A. J. Leigh Brown; UK Collaboration on HIV Drug Resistance and the Collaborative HIV, Anti-HIV Drug Resistance Network (CHAIN), Phylogenetic analyses reveal HIV-1 infections between men misclassified as heterosexual transmissions. *AIDS* **28**, 1967–1975 (2014).
15. A. J. Leigh Brown, S. J. Lycett, L. Weinert, G. J. Hughes, E. Fearnhill, D. T. Dunn, Transmission network parameters estimated from HIV sequences for a nationwide epidemic. *J. Infect. Dis.* **204**, 1463–1469 (2011).
16. G. J. D. Smith, D. Vijaykrishna, J. Bahl, S. J. Lycett, M. Worobey, O. G. Pybus, S. K. Ma, C. L. Cheung, J. Raghwani, S. Bhatt, J. S. M. Peiris, Y. Guan, A. Rambaut, Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* **459**, 1122–1125 (2009).
17. J. T. Ladner, B. Beitzel, P. S. G. Chain, M. G. Davenport, E. Donaldson, M. Frieman, J. Kugelman, J. H. Kuhn, J. O'Rear, P. C. Sabeti, D. E. Wentworth, M. R. Wiley, G.-Y. Yu, The Threat Characterization Consortium, S. Sozhamannan, C. Bradburne, G. Palacios, Standards for sequencing viral genomes in the era of high-throughput sequencing. *MBio* **5**, e01360-14 (2014).
18. Z. A. Memish, M. Cotten, B. Meyer, S. J. Watson, A. J. Alsahafi, A. A. Al Rabeeah, V. M. Corman, A. Sieberg, H. Q. Makhdoom, A. Assiri, M. A. Masri, S. Aldabbagh, B.-J. Bosch, M. Beer, M. A. Müller, P. Kellam, C. Drosten, Human infection with MERS coronavirus after exposure to infected camels, Saudi Arabia, 2013. *Emerg. Infect. Dis.* **20**, 1012–1015 (2014).
19. N. R. Faria, M. A. Suchard, A. Rambaut, D. G. Streicker, P. Lemey, Simultaneously reconstructing viral cross-species transmission history and identifying the underlying constraints. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120196 (2013).
20. P. Lemey, A. Rambaut, J. J. Welch, M. A. Suchard, Phylogeography takes a relaxed random walk in continuous space and time. *Mol. Biol. Evol.* **27**, 1877–1885 (2010).
21. P. Lemey, A. Rambaut, T. Bedford, N. Faria, F. Bielejec, G. Baele, C. A. Russell, D. J. Smith, O. G. Pybus, D. Brockmann, M. A. Suchard, Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLOS Pathog.* **10**, e1003932 (2014).
22. T. Bedford, M. A. Suchard, P. Lemey, G. Dudas, V. Gregory, A. J. Hay, J. W. McCauley, C. A. Russell, D. J. Smith, A. Rambaut, Integrating influenza antigenic dynamics with molecular evolution. *eLife* **3**, e01914 (2014).
23. S. Baize, D. Pannetier, L. Oestereich, T. Rieger, L. Koivogui, N. Magassouba, B. Soropogui, M. S. Sow, S. Keïta, H. De Clerck, A. Tiffany, G. Dominguez, M. Loua, A. Traoré, M. Kolié, E. R. Malano, E. Heleze, A. Bocquin, S. Mély, H. Raoul, V. Caro, D. Cadar, M. Gabriel, M. Pahlmann, D. Tappe, J. Schmidt-Chanasit, B. Impouma, A. K. Diallo, P. Formenty, M. Van Herp, S. Günther, Emergence of Zaire Ebola virus disease in Guinea. *N. Engl. J. Med.* **371**, 1418–1425 (2014).
24. D. M. Pigott, N. Golding, A. Mylne, Z. Huang, A. J. Henry, D. J. Weiss, O. J. Brady, M. U. G. Kraemer, D. L. Smith, C. L. Moyes, S. Bhatt, P. W. Gething, P. W. Horby, I. I. Bogoch, J. S. Brownstein, S. R. Mekaru, A. J. Tatem, K. Khan, S. I. Hay, Mapping the zoonotic niche of Ebola virus disease in Africa. *eLife* **3**, e04395 (2014).
25. S. K. Gire, A. Goba, K. G. Andersen, R. S. G. Sealfon, D. J. Park, L. Kanneh, S. Jalloh, M. Momoh, M. Fullah, G. Dudas, S. Wohl, L. M. Moses, N. L. Yozwiak, S. Winnicki, C. B. Matranga, C. M. Malboeuf, J. Qu, A. D. Gladden, S. F. Schaffner, X. Yang, P.-P. Jiang, M. Nekoui, A. Colubri, M. R. Coomber, M. Fonnie, A. Moigboi, M. Gbakie, F. K. Kamara, V. Tucker, T. Konuwa, S. Saffa, J. Sellu, A. A. Jalloh, A. Kovoma, J. Koninga, I. Mustapha, K. Kargbo, M. Foday, M. Yillah, F. Kanneh, W. Robert, J. L. B. Massally, S. B. Chapman, J. Bochicchio, C. Murphy, C. Nusbaum, S. Young, B. W. Birren, D. S. Grant, J. S. Scheiffelin, E. S. Lander, C. Happi, S. M. Gevao, A. Gnirke, A. Rambaut, R. F. Garry, S. H. Khan, P. C. Sabeti, Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345**, 1369–1372 (2014).
26. T. Stadler, D. Kühnert, D. A. Rasmussen, L. du Plessis, Insights into the early epidemic spread of Ebola in Sierra Leone provided by viral sequence data. *PLOS Curr.* **6**, ecurrents. outbreaks.02bc6d927ecee7bbd33532ec8ba6a25f (2014).
27. M. Łuksza, T. Bedford, M. Lässig, Epidemiology and evolutionary analysis of the 2014 Ebola virus outbreak. arXiv preprint available at http://arxiv.org/abs/1411.1722 (2014).
28. R. M. Anderson, R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford Univ. Press, New York, 1991).
29. A. Pandey, K. E. Atkins, J. Medlock, N. Wenzel, J. P. Townsend, J. E. Childs, T. G. Nyenswah, M. L. Ndeffo-Mbah, A. P. Galvani, Strategies for containing Ebola in West Africa. *Science* **346**, 991–995 (2014).
30. M. F. C. Gomes, A. Pastore y Piontti, N. Perra, N. Samay, Q. Zhang, A. Vespignani, L. Rossi, M. Quaggiotto, A. Panisson, M. Delfino, C. Cattuto, D. Chao, M. E. Halloran, I. M. Longini Jr., C. Poletto, L. Bioglio, V. Colizza, S. Merler, M. Ajelli, L. Fumanelli, Assessing the international spreading risk associated with the 2014 West African Ebola outbreak. *PLOS Curr.* **6**, ecurrents. outbreaks.cd818f63d40e24aef769dda7df9e0da5 (2014).
31. WHO Ebola Response Team, Ebola virus disease in West Africa—The first 9 months of the epidemic and forward projections. *N. Engl. J. Med.* **371**, 1481–1495 (2014).
32. J. M. Drake, R. B. Kaul, L. W. Alexander, S. M. O'Regan, A. M. Kramer, J. T. Pulliam, M. J. Ferrari, A. W. Park, Ebola cases and health system demand in Liberia. *PLOS Biol.* **13**, e1002056 (2015).
33. G. Chowell, H. Nishiura, Characterizing the transmission dynamics and control of Ebola virus disease. *PLOS Biol.* **13**, e1002057 (2015).
34. T. House, Epidemiological dynamics of Ebola outbreaks. *eLife* **3**, e03908 (2014).
35. S. Lycett, N. J. McLeish, C. Robertson, W. Carman, G. Baillie, J. McMenamin, A. Rambaut, P. Simmonds, M. E. J. Woolhouse, A. J. Leigh Brown, Origin and fate of A/H1N1 influenza in Scotland during 2009. *J. Gen. Virol.* **93**, 1253–1260 (2012).
36. J. J. Grefenstette, S. T. Brown, R. Rosenfeld, J. DePasse, N. T. B. Stone, P. C. Cooley, W. D. Wheaton, A. Fyshe, D. D. Galloway, A. Sriram, H. Guclu, T. Abraham, D. S. Burke, FRED (A Framework for

Reconstructing Epidemic Dynamics): An open-source software system for modeling infectious diseases and control strategies using census-based populations. *BMC Publ. Health* **13**, 940 (2013).

37. T. J. McKinley, J. V. Ross, R. Deardon, A. R. Cook, Simulation-based Bayesian inference for epidemic models. *Comput. Stat. Data Anal.* **71**, 434–447 (2014).

38. K. M. Mitchell, F. Mutapi, N. J. Savill, M. E. J. Woolhouse, Explaining observed infection and antibody age-profiles in populations with urogenital schistosomiasis. *PLOS Comput. Biol.* **7**, e1002237 (2011).

39. V. A. A. Jansen, N. Stollenwerk, H. J. Jensen, M. E. Ramsay, W. J. Edmunds, C. J. Rhodes, Measles outbreaks in a population with declining vaccine uptake. *Science* **301**, 804 (2003).

40. S. Blumberg, J. O. Lloyd-Smith, Inference of R_0 and transmission heterogeneity from the size distribution of stuttering chains. *PLOS Comput. Biol.* **9**, e1002993 (2013).

41. S. S. Morse, J. A. K. Mazet, M. E. J. Woolhouse, C. R. Parrish, D. Carroll, W. B. Karesh, C. Zambrana-Torrelio, W. I. Lipkin, P. Daszak, Prediction and prevention of the next pandemic zoonosis. *Lancet* **380**, 1956–1965 (2012).

42. C. R. Braden, S. F. Dowell, D. B. Jernigan, J. M. Hughes, Progress in global surveillance and response capacity 10 years after severe acute respiratory syndrome. *Emerg. Infect. Dis.* **19**, 864–869 (2013).

43. *Climate Change 2007: Synthesis Report* (Intergovernmental Panel on Climate Change, Geneva, Switzerland, 2007).

44. M. D. Van Kerkhove, T. Asikainen, N. G. Becker, S. Bjorge, J.-C. Desenclos, T. dos Santos, C. Fraser, G. M. Leung, M. Lipsitch, I. M. Longini Jr., E. S. McBryde, C. E. Roth, D. K. Shay, D. J. Smith, J. Wallinga, P. J. White, N. M. Ferguson, S. Riley; WHO Informal Network for Mathematical Modelling for Pandemic Influenza H1N1 2009 (Working Group on Data Needs), Studies needed to address public health challenges of the 2009 H1N1 influenza pandemic: Insights from modeling. *PLOS Med.* **7**, e1000275 (2010).

45. N. J. McLeish, P. Simmonds, C. Robertson, I. Handel, M. McGilchrist, B. K. Singh, S. Kerr, M. E. Chase-Topping, K. Sinka, M. Bronsvoort, D. J. Porteous, W. Carman, J. McMenamin, A. J. Leigh-Brown, M. E. J. Woolhouse, Sero-prevalence and incidence of A/H1N1 2009 influenza infection in Scotland in winter 2009–2010. *PLOS One* **6**, e20358 (2011).

Science Translational Medicine

Lessons from Ebola: Improving infectious disease surveillance to inform outbreak management

Mark E. J. Woolhouse, Andrew Rambaut and Paul Kellam

ARTICLE TOOLS	http://stm.sciencemag.org/content/7/307/307rv5
RELATED CONTENT	http://stm.sciencemag.org/content/scitransmed/7/286/286ps11.full http://stm.sciencemag.org/content/scitransmed/4/138/138ra81.full http://science.sciencemag.org/content/sci/350/6267/1461.full http://science.sciencemag.org/content/sci/351/6268/12.full http://science.sciencemag.org/content/sci/351/6275/815.full http://stm.sciencemag.org/content/scitransmed/8/329/329ra33.full http://science.sciencemag.org/content/sci/351/6277/1078.full http://science.sciencemag.org/content/sci/351/6279/1339.full http://science.sciencemag.org/content/sci/351/6279/1343.full http://stm.sciencemag.org/content/scitransmed/8/339/339ed7.full http://stm.sciencemag.org/content/scitransmed/8/341/341ps12.full http://science.sciencemag.org/content/sci/353/6304/1073.full http://science.sciencemag.org/content/sci/354/6310/284.full http://science.sciencemag.org/content/sci/354/6310/350.full http://science.sciencemag.org/content/sci/355/6324/463.full http://science.sciencemag.org/content/sci/353/6300/658.2.full http://science.sciencemag.org/content/sci/353/6300/658.1.full http://stm.sciencemag.org/content/scitransmed/12/554/eabc1126.full
REFERENCES	This article cites 39 articles, 7 of which you can access for free http://stm.sciencemag.org/content/7/307/307rv5#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions