

Productive disruption: opportunities and challenges for innovation in infectious disease surveillance

Caroline O. Buckee,¹ Maria I E Cardenas,² June Corpuz,³ Arpita Ghosh,⁴ Farhana Haque,⁵ Jahirul Karim,^{6,7} Ayesha S. Mahmud,¹ Richard J Maude,^{1,7,8} Keitly Mensah,⁹ Nkengafac Villyen Motaze,¹⁰ Maria Nabaggala,¹¹ Charlotte Jessica Eland Metcalf,⁹ Sedera Aurélien Mioramalala,¹² Frank Mubiru,¹¹ Corey M. Peak,¹ Santanu Pramanik,⁴ Jean Marius Rakotondramanga,¹³ Eric Remera,¹⁴ Ipsita Sinha,^{7,8} Siv Sovannaroeth,¹⁵ Andrew J Tatem,¹⁶ Win Zaw⁷

To cite: Buckee CO., Cardenas MIE, Corpuz J, *et al*. Productive disruption: opportunities and challenges for innovation in infectious disease surveillance. *BMJ Glob Health* 2018;**3**:e000538. doi:10.1136/bmjgh-2017-000538

Handling editor Alberto L Garcia-Basteiro

Received 25 August 2017
Revised 16 November 2017
Accepted 17 November 2017

BACKGROUND

Infectious diseases place an unacceptable and disproportionate social and economic burden on low-income countries. National disease control programmes have the difficult task of allocating limited budgets for interventions across regions of their countries, based on often disparate datasets of varying quality from a range of sources including clinics, hospitals, village health workers, the private sector and non-governmental organisations (NGOs). Every stage of the data collection and analysis pipeline for surveillance systems may be affected by a lack of capacity as well as by biases and misaligned incentives for reporting and managing data. Addressing these issues will be essential for effective reduction in the burden of endemic infectious diseases globally as well as to preparing for emerging epidemic threats.

Meanwhile, academic researchers—often in high-income settings—are developing increasingly sophisticated methods to collect and analyse data to improve spatial estimates of disease burden using new Big Data sources, mobile-Health or m-Health approaches or mechanistic and statistical modelling techniques. While these advances leap ahead, however, many remain most useful for estimating global disease distribution,¹ rather than for national control programme prioritisation. Translating these new techniques to inform policy in endemic settings remains challenging. The pronounced disconnect between health systems and academia may limit the utility of new approaches. The high burden of work placed on healthcare workers

Summary box

- ▶ New innovations that could transform infectious disease surveillance and control, including the use of Big Data, mobile health approaches and cutting edge quantitative methods, offer hope for disrupting traditional health systems and improving health worldwide.
- ▶ Much has been made of their potential, but very few have been translated successfully into policy or scaled up to a population level.
- ▶ We argue that there is currently a lack of integration of new approaches, making them unsustainable or unrealistic for most national control programmes and that the gulf between academia and policy makers remains a major barrier to their implementation.
- ▶ We propose that these innovations must be designed with direct input from national control programmes and embedded within already existing health systems.

in low-income settings further limits their scope and time available for engagement with methodological developments.

Despite ongoing challenges to implementation, however, there are promising analytical approaches that can leverage even patchy and low-quality data and diverse new data streams that can be productively harnessed to strengthen strategies for resource allocation when integrated with existing surveillance systems. We detail the data and analysis challenges faced by national disease control programmes, outline possible solutions offered by analytical approaches and new data-streams and conclude by outlining barriers to implementation.



For numbered affiliations see end of article.

Correspondence to

Dr Caroline O. Buckee;
cbuckee@hsph.harvard.edu

CHALLENGES ASSOCIATED WITH INFECTIOUS DISEASE SURVEILLANCE SYSTEMS

Generally, epidemiological data about patients are reported by healthcare practitioners via passive surveillance systems to a central database, which is used to determine trends over time in and map the geographic distribution of burden of disease in different regions as well as the extent and efficacy of interventions (active surveillance via sentinel sites may also inform these efforts). These regional data in turn serve as an important basis for resource allocation decisions. Figure 1 illustrates the flow of data and potential hurdles faced by national control programmes and the ways in which new approaches may be used in parallel to traditional systems.

Data quality is often perceived as a major barrier to using passive surveillance data to guide resource prioritisation. Data may be delayed or simply missing from core settings and, even where available, is frequently under-reported, potentially in spatially and temporally variable ways.² For example, a problematic reporting pattern that emerges repeatedly is an apparent increase in disease incidence that is actually caused by increased surveillance efforts and/or diagnostic capacity. Over-reporting is also a potential hazard when local regions are financially incentivised to exaggerate their needs. Further,

since many local health centres lack diagnostic capabilities, much of the large-scale data rely on syndromic surveillance (influenza-like-illness, diarrhoea, fever) with low specificity. Additionally, estimates of the catchment population of health facility or district may be flawed, since data may be based on a single census from up to a decade ago, thus potentially both out of date³ and a poor reflection of seasonal fluctuations in population numbers.² Collection of more detailed, highly curated data at sentinel sites seems promising, but may add little to national decision-making because such high-quality sites are likely to be limited in their spatial scale. All these issues mean that where case numbers exist, policy makers may have little confidence in them.

LEVERAGING FRAGILE DATA USING STATISTICAL AND MATHEMATICAL PRINCIPLES

The first question is what can be done with existing infrastructure and data. With the right statistical or theoretical tools, even low-quality data can potentially be leveraged to help inform strategic deployment of control efforts. Thoughtful deployment of interpolation or geostatistical tools can be used to create smooth maps of burden or intervention efforts across space, also

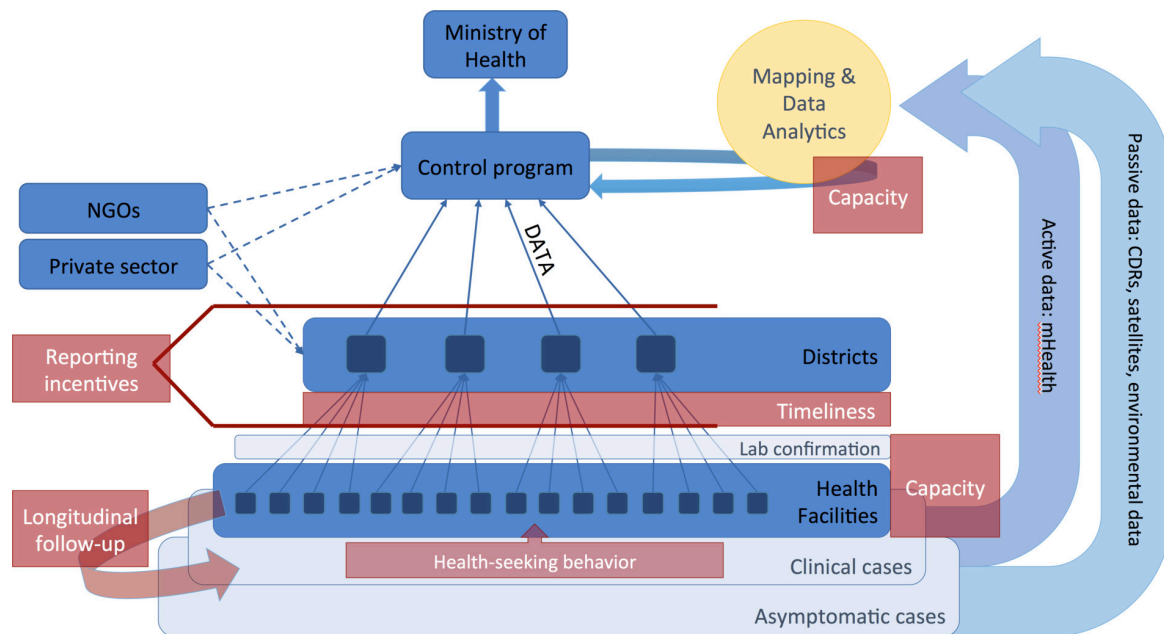


Figure 1 Data flows through health systems (blue) and major challenges faced by control programmes (red). A subset of clinical cases, which often represent only a subset of total infections both asymptomatic and clinical, are first detected by local health workers, most typically in health facilities and hospitals. Local health workers are also responsible for following up individuals with chronic infections requiring multiple treatments over months or years. Some fraction of clinical cases are lab confirmed, depending on capacity, and reported to regional or district centres, which in turn report to national control programmes. Data are often aggregated before being reported centrally. NGOs and the private sector may also produce a significant amount of epidemiological data. National control programmes aggregate and analyse data to map the distribution of disease burden, intervention efficacy and so on. New direct mHealth approaches (eg, participatory surveillance) and passively collected data (eg, from mobile phones via Call Data Records (CDRs); and satellites) may be used directly by control programmes to map underlying risks and population distributions. At every level, capacity remains an enormous issue for routine surveillance, and training for new approaches will be challenging for most control programmes. At different levels of the health system, incentives for reporting accurately may be misaligned, and timeliness of reporting may be particularly problematic for emerging threats. NGOs, non-governmental organisations.

allowing extrapolation to unmeasured contexts.¹⁴⁵ Auto-correlation models are also powerful tools, building on surveillance data to guide predictions about outbreaks of dengue, for example.⁶ Moving from statistical to mechanistic approaches, even if incidence reporting is erratic, dynamical signatures of the infectious process might still be detectable if additional data on features of cases are available, such as age, geographic location and gender. Age is a powerful covariate for infectious disease dynamics, as age of infection is linked to the magnitude of transmission. High-transmission pathogens often have a low average age of infection, as they move quickly through immunologically naïve populations.⁷ Conversely, implementation of control efforts is likely to increase the average age of infection,⁸ so intervention efficacy may be measured using shifting age structure of cases.

Where data are consistent through time, but not space, basic principles from infectious disease dynamics open the way to estimating characteristics of pathogens. For example, the growth rate of an epidemic can be extracted from incidence, allowing estimates of the net reproduction number, or R_0 , which captures the degree to which an outbreak is expected to grow ($R_0 > 1$) or shrink ($R_0 < 1$) (eg, deployed during the recent Ebola outbreak despite variation in reporting rates).⁹ Allocation of resources towards ‘source’ populations, where $R_0 > 1$ then becomes possible—although maps of the locations and densities of rural populations (ie, denominator challenges) are also

necessary. Infectious disease models can allow characteristics of the surveillance system, such as the magnitude of under-reporting, to be estimated where the susceptible population can be inferred (eg, via susceptible reconstruction).¹⁰ Where only syndromic surveillance is available, it may be possible to correct for background rates of focal syndromes to pull out the dynamics associated with a particular infection. This strategy has been used for influenza¹¹ (influenza-like-illness data are frequently available but data on influenza are rare), enabling investigation of signatures of climate effects on the burden of infection,¹² for example, which has potential to contribute to planning efforts. Alternatively, simulation tools based on known epidemiological parameters of particular pathogens can establish the degree to which the data are reliable and forecast outbreaks and emergence events and/or the impact of interventions like vaccination (eg, roll-out of cholera vaccination to contain an epidemic).¹³

COMPLEMENTING EXISTING EPIDEMIOLOGICAL INFORMATION WITH NEW DATA SOURCES

While these analytical strategies can compensate for the limitations of different data quality issues, a range of promising new data streams are also available. Rapid technological advances make these increasingly affordable, offering additional or new data layers to include in epidemiological analyses.¹¹⁴¹⁵ We focus on three new data types

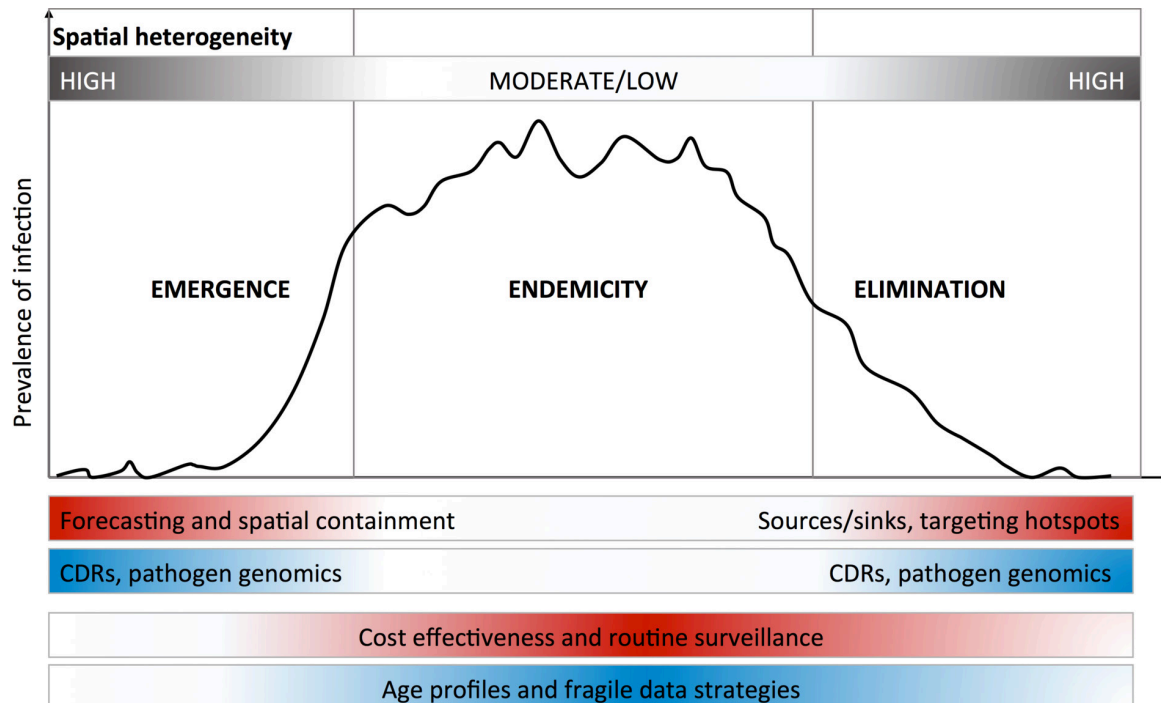


Figure 2 Optimal use of new approaches depends on epidemiological context. Different phases of epidemiological containment and control lend themselves to different analytical approaches and data sources. Here, we have highlighted the spatial dimensions of this issue, with emergence and elimination phases exhibiting high spatial heterogeneity. In these cases, pronounced heterogeneity produces signals in data that can be leveraged to model the spread of infection between populations. For endemic infections where prevalence is distributed throughout the country and controlling disease burden is the primary purpose of interventions, the use of age profiles of exposure and other analytical approaches may be used to enhance or make use of patchy or poor quality data.

that are tractable in the context of control programme capacity: geospatial data, passively collected mobile phone records and pathogen genomic data. Note that we do not discuss the many mHealth approaches to actively engaging with populations directly, for surveillance or for interventions like health education,¹⁶ but these also provide data that do not rely on traditional surveillance systems. In general, different needs dominate at different phases across an epidemiological spectrum from emergence to elimination (figure 2), which will determine which data are useful and how they should be analysed.

Geospatial data are increasingly of high resolution and encompasses settlements and transport networks, indices of vegetation coverage, land use, land surface temperature and wind speed. These data can be combined with other geospatially referenced data such as meteorological data from weather stations, population densities or road networks to generate comprehensive estimates of environmental variables or indicators like remoteness or urbanicity relevant to communicable disease transmission. This can then be combined with point pattern data on vector presence via machine learning algorithms to determine features such as the likely range and local transmission intensity of vector borne infections like Zika virus.¹⁷

Mobile phone data—routinely collected by operators and providing information about the location and movements of subscribers in real time—offer tremendous promise for control programmes to measure disease spread, if appropriately anonymised and aggregated to protect subscriber privacy.¹⁸ Indeed, the development of appropriate anonymisation and aggregation protocols remains an important priority for academics and public health practitioners and will require carefully balancing the ethical risks of reidentifying individuals with the benefits of predicting disease spread and identifying targets for intervention. Integration of this information into risk mapping by control programmes offers particular promise where disease incidence is heterogeneous in time and space and mobility drives both changing burden of infection and the type of intervention needed. For example, during an emerging epidemic (eg, Ebola), spatial containment of the disease and proactive surveillance (in the correct locations) are essential, requiring specific spatial targets and estimates of how people will move the infection to new regions.¹⁹ At the other end of the spectrum, control programmes aiming for elimination (eg, malaria, measles) require accurate maps of remaining foci of transmission and an understanding of the relative importance of local versus imported cases of disease.¹⁸ These data can also be leveraged to address issues in estimation of the denominator or population at risk.³

Genome sequencing costs have declined strikingly in the last decades, making it increasingly feasible for control programmes to integrate pathogen sequencing and molecular epidemiology into their data collection and analysis strategies. Unlike mobile phone data, which

offers an external view of mobility that can be used to model disease spread, analysis of pathogen genetic data provides complementary insights into transmission chains and pathogen gene flow between locations. During outbreaks, sequencing can provide insights into the place and time of the outbreak's origin and the pace of its spatial spread.²⁰ For elimination planning, pathogen sequencing can identify the regional or national origin of a particular isolate and estimate the rate of migration between populations where this information is difficult to measure by other means. For endemic pathogens like TB, where low incidence and relatively stable prevalence in many places makes analytical inferences and spatial data from mobile phones or satellites less tractable, genomics can provide key insights into the spread of drug resistance and the connectivity between different populations.

ENGAGING COMMUNITY HEALTH WORKERS AND RESEARCHERS WHO COLLECT AND ANALYSE DATA IS KEY

While our focus is on the potential of existing data despite perceived inadequacies and in use of novel data streams, it is important to note that country experience points to many low-cost adjustments to practice that could improve data availability. In multiple settings, there are important opportunities to share data across diseases, breaking down silos between disease surveillance and/or control programmes. This effort would be enhanced by electronic rather than paper-based reporting systems, allowing standardised data collection, aggregation and sharing. Perhaps most importantly, however, effectively engaging community health workers and others who collect and report patient data will be key to strengthening surveillance. For example, in Cambodia, malaria case reporting was improved when healthcare workers received feedback about the incidence of malaria in their area using colour-coded stickers; the effective use of mobile phones and tablets for reporting cases improved when health workers could use them for personal use and where internet coverage was restricted, the flexible reporting using mobile phones via SMS and later smartphones greatly improved timeliness of data collection.

It will be essential to sustain resources for engaging with health workers involved in generating and reporting surveillance data and not to impose technological advances at the expense of motivating and retaining qualified, reliable workers who form the foundation of an effective health system. Barriers to implementing approaches that use novel types of data or new analytical tools are rooted in human capacity (figure 1) at every level within control programmes, from local healthcare centres up to the ministry of health. While a considerable strength of novel data-streams such as satellite or mobile phone data are that they bypass many of the levels of reporting where data may be lost and incentives are misaligned, the insights they provide will be limited for setting public health agendas if they are not integrated

with traditional reporting systems. For academics, one of the most important contributions in this regard will be providing training, so that public health officers have the technical capacity to understand the benefits of, and oversee the implementation of, new methods. Sustainable integrating of these approaches will require sustained commitment from Ministries of Health, donor agencies and academics.

Author affiliations

¹Department of Epidemiology, Center for Communicable Disease Dynamics, Harvard School of Public Health, Boston, Massachusetts, USA

²Centro de Atención y Diagnóstico de Enfermedades Infecciosas, Universidad Industrial de Santander, Bucaramanga, Colombia

³Epidemiology Bureau, Department of Health of the Philippines, Manila, Philippines

⁴Public Health Foundation of India, Vasant Kunj, New Delhi, India

⁵Programme for Emerging Infections (PEI), Infectious Diseases Division (IDD), ICDDR, B, Dhaka, Bangladesh

⁶Disease Control Department, Directorate General of Health Services, Dhaka, Bangladesh

⁷Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand

⁸Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK

⁹Dept of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA

¹⁰Division of Epidemiology and Biostatistics, Stellenbosch University, Stellenbosch, South Africa

¹¹Infectious Disease Institute, College of Health Sciences, Makerere University, Uganda

¹²M&E Service, Data & Malaria Survey Division, National Malaria Control Program, Antananarivo, Madagascar

¹³Epidemiology Unit, Institut Pasteur, Madagascar

¹⁴Rwanda Biomedical Center, Kigali, Rwanda

¹⁵Technical Bureau, National Malaria Control Program, Phnom Penh, Cambodia

¹⁶WorldPop, Department of Geography and Environment, University of Southampton, Southampton, UK

Contributors All authors conceived, planned and contributed intellectually to this work. COB and CJEM wrote the manuscript, and all authors edited the manuscript.

Funding This work was the result of the Harvard-Wellcome Trust Workshop on Infectious Disease Modeling for Control Programs in 2017. It was supported by Wellcome Trust Fund 106866-Sustaining Health Award and by Award Number U54GM088558 from the National Institute of General Medical Sciences.

Disclaimer The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use,

provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- Hay SI, George DB, Moyes CL, *et al*. Big data opportunities for global infectious disease surveillance. *PLoS Med* 2013;10:e1001413.
- Buckee CO, Tatem AJ, Metcalf CJ. Seasonal population movements and the surveillance and control of infectious diseases. *Trends Parasitol* 2017;33:10–20.
- Tatem AJ. Mapping the denominator: spatial demography in the measurement of progress. *Int Health* 2014;6:153–5.
- Kraemer MU, Hay SI, Pigott DM, *et al*. Progress and challenges in infectious disease cartography. *Trends Parasitol* 2016;32:19–29.
- Takahashi S, Metcalf CJE, Ferrari MJ, *et al*. The geography of measles vaccination in the African Great Lakes region. *Nat Commun* 2017;8:15585.
- Johansson MA, Reich NG, Hota A, *et al*. Evaluating the performance of infectious disease forecasts: A comparison of climate-driven and seasonal dengue forecasts for Mexico. *Sci Rep* 2016;6:33707.
- Anderson RM, May RM. *Infectious diseases of humans*. Oxford: Oxford University Press, 1991.
- Ferrari MJ, Grenfell BT, Strebel PM. Think globally, act locally: the role of local demographics and vaccination coverage in the dynamic response of measles infection to control. *Philos Trans R Soc Lond B Biol Sci* 2013;368:20120141.
- Shaman J, Yang W, Kandula S. Inference and forecast of the current west african ebola outbreak in Guinea, sierra leone and liberia. *PLoS Curr* 2014;6.
- Bjørnstad ON, Finkenstadt B, Grenfell BT. Endemic and epidemic dynamics of measles: estimating epidemiological scaling with a time series SIR model. *Ecological Monographs* 2002;72:169–84.
- Charu V, Zeger S, Gog J, *et al*. Human mobility and the spatial transmission of influenza in the United States. *PLoS Comput Biol* 2017;13:e1005382.
- Fisman DN, Tuite AR, Brown KA. Impact of el niño southern oscillation on infectious disease hospitalization risk in the United States. *Proc Natl Acad Sci U S A* 2016;113:14589–94.
- Azman AS, Luquero FJ, Rodrigues A, *et al*. Urban cholera transmission hotspots and their implications for reactive vaccination: evidence from Bissau city, Guinea bissau. *PLoS Negl Trop Dis* 2012;6:e1901.
- Tatem AJ. Mapping population and pathogen movements. *Int Health* 2014;6:5–11.
- Haque F, Ball RL, Khatun S, *et al*. Evaluation of a smartphone decision-support tool for diarrheal disease management in a resource-limited setting. *PLoS Negl Trop Dis* 2017;11:e0005290.
- Tomlinson M, Rotheram-Borus MJ, Swartz L, *et al*. Scaling up mHealth: where is the evidence? *PLoS Med* 2013;10:e1001382.
- Messina JP, Kraemer MU, Brady OJ, *et al*. Mapping global environmental suitability for Zika virus. *Elife* 2016;5:e15272.
- Wesolowski A, Eagle N, Tatem AJ, *et al*. Quantifying the impact of human mobility on malaria. *Science* 2012;338:267–70.
- Salje H, Lessler J, Endy TP, *et al*. Revealing the microscale spatial signature of dengue transmission and immunity in an urban population. *Proc Natl Acad Sci U S A* 2012;109:9535–8.
- Arias A, Watson SJ, Asogun D, *et al*. Rapid outbreak sequencing of Ebola virus in Sierra Leone identifies transmission chains linked to sporadic cases. *Virus Evol* 2016;2:vew016.