Downloaded from http://rstb.royalsocietypublishing.org/ on June 6, 2017

PHILOSOPHICAL TRANSACTIONS B

rstb.royalsocietypublishing.org

Opinion piece

Cite this article: Cori A *et al*. 2017 Key data for outbreak evaluation: building on the Ebola experience. *Phil. Trans. R. Soc. B* **372**: 20160371.
http://dx.doi.org/10.1098/rstb.2016.0371

Accepted: 11 November 2016

One contribution of 17 to a theme issue 'The 2013–2016 West African Ebola epidemic: data, decision-making and disease control'.

Subject Areas:
health and disease and epidemiology

Keywords:
West African Ebola epidemic, epidemic, mathematical modelling, outbreak response, data, public health

Author for correspondence:
Harriet L. Mills
e-mail: harriet.l.mills@gmail.com

Key data for outbreak evaluation: building on the Ebola experience

Anne Cori[1,†], Christl A. Donnelly[1], Ilaria Dorigatti[1], Neil M. Ferguson[1], Christophe Fraser[2], Tini Garske[1], Thibaut Jombart[1], Gemma Nedjati-Gilani[1], Pierre Nouvellet[1], Steven Riley[1], Maria D. Van Kerkhove[3], Harriet L. Mills[1,4,5,†] and Isobel M. Blake[1,†]

[1]Medical Research Council Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, London W2 1PG, UK
[2]Oxford Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, UK
[3]Centre for Global Health, Institut Pasteur, 25-28 Rue du Dr Roux, 75015 Paris, France
[4]MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol BS8 2BN, UK
[5]School of Veterinary Sciences, University of Bristol, Bristol BS40 5DU, UK

AC, 0000-0002-8443-9162; CAD, 0000-0002-0195-2463; TG, 0000-0002-8952-4710; GN-G, 0000-0001-5723-5028; IMB, 0000-0002-3977-1318

Following the detection of an infectious disease outbreak, rapid epidemiological assessment is critical for guiding an effective public health response. To understand the transmission dynamics and potential impact of an outbreak, several types of data are necessary. Here we build on experience gained in the West African Ebola epidemic and prior emerging infectious disease outbreaks to set out a checklist of data needed to: (1) quantify severity and transmissibility; (2) characterize heterogeneities in transmission and their determinants; and (3) assess the effectiveness of different interventions. We differentiate data needs into individual-level data (e.g. a detailed list of reported cases), exposure data (e.g. identifying where/how cases may have been infected) and population-level data (e.g. size/demographics of the population(s) affected and when/where interventions were implemented). A remarkable amount of individual-level and exposure data was collected during the West African Ebola epidemic, which allowed the assessment of (1) and (2). However, gaps in population-level data (particularly around which interventions were applied when and where) posed challenges to the assessment of (3). Here we highlight recurrent data issues, give practical suggestions for addressing these issues and discuss priorities for improvements in data collection in future outbreaks.

This article is part of the themed issue 'The 2013–2016 West African Ebola epidemic: data, decision-making and disease control'.

1. Introduction

Detection of a new infectious disease outbreak requires rapid assessment of both the clinical severity and the pattern of transmission to plan appropriate response activities. Following the subsequent roll-out of interventions, continued evaluation is necessary to detect reductions in transmission and assess the relative impact of different interventions. Surveillance data are crucial for informing these analyses, and directly determine the extent to which they can be performed.

Despite the unprecedented scale of the 2013–2016 West African Ebola epidemic [1,2], detailed data were collected during the outbreak, which proved invaluable in guiding the response [3–10]. Multiple studies have already considered the lessons to be learned from the Ebola experience with respect to

[†]These authors contributed equally to this study.

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.3705733.

© 2017 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

THE ROYAL SOCIETY PUBLISHING

Downloaded from http://rstb.royalsocietypublishing.org/ on June 6, 2017

Box 1. Recommendations for collecting and using data for outbreak response.

1. *Collecting relevant data*

Data need to be collected at each of the three levels:

- *individual level*: detailed information about cases
- *exposure level*: information about exposure events that may have led to transmission
- *population level*: characteristics of the population(s) in which the outbreak is spreading and the interventions carried out in the population(s)

Although some data will be context-specific, others, in particular at the population level, will be useful for a wide range of epidemics, and should be routinely and centrally collected in preparation for the next outbreak.

2. *Optimizing data quality*

Having a general framework ready in advance of the next outbreak will facilitate:

- *quality and timeliness* of data
- *centralization and harmonization* of data at all three levels
- *preparedness to deploy* training material, personnel and logistical resources

There is substantial room for improvement in the quality of data collected in an epidemic context, particularly for emerging pathogens.

3. *Ensuring adequate data availability*

Data need to be shared ensuring a balance between the following considerations:

- *ethical*: protecting anonymity while ensuring data are sufficiently detailed to be useful
- *scientific*: wide data access is desirable to promote independent analyses; however, mechanisms must be in place for systematic comparison of results
- *practical*: deciding on a data format for sharing, on who will be responsible for data cleaning and on how various roles will be acknowledged

Discussions and decisions relating to data sharing remain ongoing and guidelines should be agreed on in advance of the next outbreak.

4. *Analysing data and reporting results in an appropriate manner*

The scientific community should agree on guidelines for epidemic modelling and analyses (such as those in place for reporting experimental studies), such as:

- *assumptions* underlying analyses should be clearly stated
- *sensitivity* to these assumptions should be tested
- *uncertainty* in results should be adequately explored and reported

These are particularly relevant for reporting epidemic projections.

coordinating international responses to health crises, strengthening local health systems and improving clinical care and surveillance tools [11–23]. Here we discuss what can be learned to improve real-time epidemiological assessment in future outbreaks via improved data collection and analyses, building on similar contributions after other epidemics [24–26]. We focus on efforts to reduce and interrupt transmission. First, we outline analyses that are essential to inform response activities during different stages of an epidemic. Second, we detail the various types of data needed to perform these analyses, with examples from the Ebola experience. Third, we summarize the successes and challenges of data collection experienced during this outbreak, and the implications this had for answering key public health questions. Fourth, we suggest improvements that could be implemented in future outbreaks, again drawing from the Ebola experience. Finally, we discuss issues related to availability of data and analyses (box 1).

2. Key public health questions and corresponding analyses

Key public health questions for any emerging infectious disease outbreak are the following: (i) What is the likely public health impact of the outbreak? (ii) How feasible is controlling the outbreak and what interventions would be appropriate? (iii) Are current interventions effective and could they be improved? Here, we describe statistical and mathematical analyses that facilitate epidemic response planning, focusing on these questions (figure 1). In this section, we take a general view, as these questions are recurring during most, if not all, outbreaks. We provide examples from the West African Ebola epidemic in subsequent sections.

(a) What is the likely public health impact of the outbreak?

A key issue in the early phase of an epidemic is to determine the potential impact of the outbreak in terms of clinical severity and the likely total number of cases over different time periods.

(i) Severity

The severity of a pathogen is often characterized by the case fatality ratio (CFR), the proportion of cases who die as a result of their infection. Estimating the CFR during an outbreak can be challenging due to inconsistent case definitions, incomplete case reporting and right-censoring of data [27–29]. In particular, it is critical to know the proportion of cases for whom clinical outcome is unknown or has not been recorded, which is typically easier to assess using detailed case data rather than aggregated case counts [27]. The CFR may differ across populations (e.g. age, space, treatment); quantifying

Downloaded from http://rstb.royalsocietypublishing.org/ on June 6, 2017

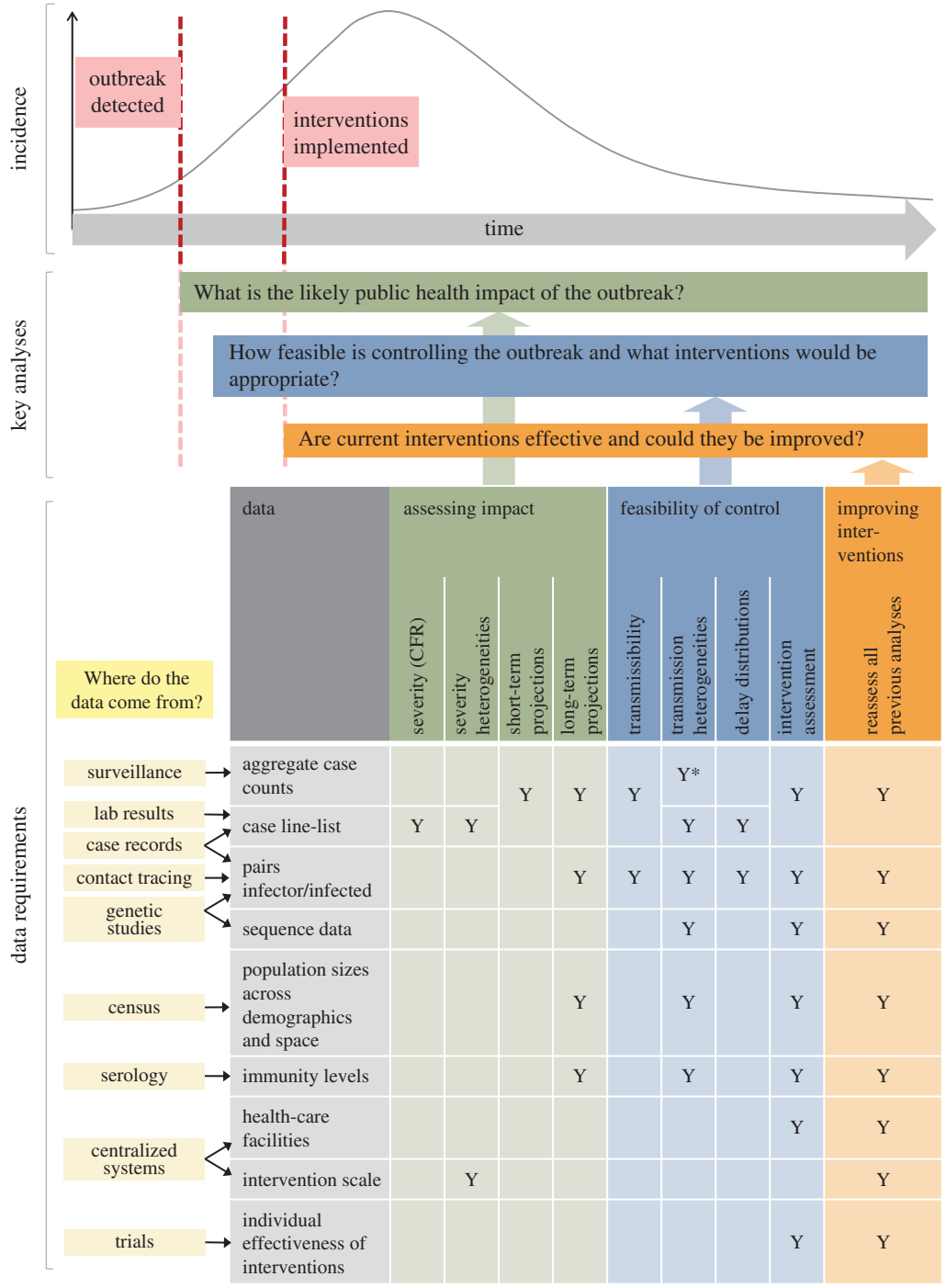


Figure 1. Schematic illustrating the data needed to answer questions at different stages of the epidemic to inform the response. Asterisk indicates that analysis is only possible if aggregate counts are stratified. Footnote: Where two cells in a column are merged, either or both types of data may be used. CFR, case fatality ratio.

these heterogeneities can help target resources appropriately and compare different care regimens. For less severe emerging pathogens, the case definition typically only encompasses a small fraction of all infected individuals, and hence the infection fatality rate (i.e. the proportion of infected individuals who die, rather than the proportion of cases who die—as per the case definition, which may not be equivalent to infection) may be a more useful measure of severity [30].

(ii) Short- and long-term incidence projections

Short-term impact of an outbreak can be assessed by predicting the number of cases that will arise in the next few days or weeks. This is particularly relevant for evaluation of immediate health-care capacity needs. Projections of future incidence and estimates of the doubling time (the time taken for the incidence to double) can be obtained by extrapolating the early time series of reported cases either obtained from aggregated surveillance data or calculated from individual records [31]. These projections typically rely on the assumption that incidence initially grows exponentially [31]. They are subject to uncertainty, which increases the further one looks into the future. Quantifying and appropriately reporting such uncertainty and underlying assumptions are crucial [22,25,32,33]. Overstating uncertainty can lead to inappropriately pessimistic projections, which may in turn be detrimental to the control of the outbreak [34]. On the other hand, understating uncertainty prevents policymakers from making decisions based on the whole spectrum of possible impacts. Some studies have already discussed how to find the balance

Downloaded from http://rstb.royalsocietypublishing.org/ on June 6, 2017

between these two extremes [35–37]. Here, we propose two simple rules of thumb for projecting case numbers. First, projections should not be made for more than two or three generations of cases into the future. Second, central projections should be shown together with lower and upper bounds. In the future, the modelling community should agree on guidelines for reporting epidemic projections, as are already in place for reporting experimental studies [38].

A number of factors can lead to incidence not growing exponentially. In particular, this happens once herd immunity accumulates, if population behaviour changes or as a result of the implementation of interventions. Dynamic transmission models, which account for saturation effects, can be used to assess the long-term impact of the outbreak such as predicting the timing and magnitude of the epidemic peak or the attack rate (final proportion of population infected) [39,40]. However, these models are hard to parametrize as they require information on population size and immunity, interventions (if any) and potential behavioural changes over time, all of which may be subject to uncertainty [41,42]. We discuss these issues in more detail in §3. Projecting longer term is likely to be associated with a large degree of uncertainty and these projections may be more useful for evaluating qualitative trends and evaluating intervention choices rather than predicting exact case numbers.

(b) How feasible is controlling the outbreak and what interventions would be appropriate?

Interventions to reduce transmission can include community mobilization, quarantine, isolation, treatment or vaccination. The potential success of these interventions is determined by general characteristics of the disease such as overall transmissibility and how this varies across populations [43]. Furthermore, certain types of interventions may be more or less appropriate depending on the natural history of the disease and the context of the epidemic.

(i) Overall transmissibility

The transmissibility of a pathogen determines the intensity of interventions needed to achieve epidemic control [44]. The parameter most often used to quantify transmissibility is the reproduction number (R), the mean number of secondary cases infected by a single individual. This parameter has an intuitive interpretation: if $R > 1$, then the epidemic is likely to grow, whereas if $R < 1$, the epidemic will decline [44,45]. The final attack rate (proportion of the population infected) of an epidemic also depends on the value of R at the start of an epidemic (termed R_0 if the population has no immunity). R can be estimated from the incidence of reported cases, given knowledge of the serial interval distribution (i.e. distribution of time between symptom onset in a case and symptom onset in his/her infector; see §2b(iii)) [46–49].

(ii) Transmission heterogeneities

Heterogeneity in the number of secondary cases generated by each infected individual affects epidemic establishment and the ease of control. Greater heterogeneity reduces the chance of an outbreak emerging from a single case [50]. However, this heterogeneity can make an established outbreak hard to control using mass interventions, as a single uncontrolled case can generate a large number of secondary cases [50]. Conversely, heterogeneity in transmission may provide opportunities for targeting interventions if the individuals who contribute more to transmission (because of environmental, behavioural and/or biological factors [51–53]) share socio-demographic or geographical characteristics that can be identified [50,54]. Reconstruction of transmission trees (who infects whom) can provide an understanding of who contributes more to transmission. This can be done using detailed case investigations and/or using genetic data [55–58].

Environmental, behavioural and biological factors may also lead to groups of individuals being disproportionately more likely to acquire infection (e.g. children during influenza outbreaks [53,59] or health-care workers (HCWs) during Ebola outbreaks [60,61]). To identify whether such groups exist and target them appropriately, the proportion infected in each group must be estimated. This requires population size estimates for the different groups, which may be difficult to obtain, as we highlight §3c.

Spatial heterogeneity in transmission is particularly interesting to assess as it can inform the targeting of surveillance and interventions to the geographical areas most at risk. Phylo-geographical studies based on genetic data can improve understanding of the geographical origins of the outbreak, identify and characterize sub-outbreaks and quantify whether transmission is very local or travels large distances [22,62,63]. Results of such analyses can be used to determine the appropriate spatial scale of control measures. Spatially explicit epidemic models can also be used to quantify the risk of exportation of the infection from one place to another. This can help public health officials to tailor prevention and control resources to the level of risk likely to be experienced by a given area. Such models typically require detailed data on mobility patterns and immunity levels of the populations in the areas of interest [64–66].

(iii) Delay distributions

Disease natural history fundamentally affects outbreak dynamics. The generation time distribution (i.e. distribution of time between infection of an index case and infection of its secondary cases) determines—with the reproduction number—the growth rate of an epidemic [67]. Most commonly, the generation time distribution is estimated from data on the serial interval distribution of an infection—the delay between symptom onset in a case and symptom onset in his/her infector. Other delays between events in the natural history of infection (e.g. exposure, onset of symptoms, hospitalization and recovery or death) also affect disease transmission or have implications for control [43,67]. Delays from symptom onset to recovery (or death) will determine the required duration of health-care and case isolation. The incubation period (the delay between infection and symptom onset of a case) and the extent to which infectiousness precedes symptom onset will determine the feasibility and effectiveness of contact tracing or prophylaxis [43]. Estimating these delay distributions requires detailed data on individual cases and their exposure, e.g. through transmission pairs identified in household studies [43].

(iv) Intervention choice

Other analyses can also help refine the type of interventions that should be considered. Ecological associations between transmissibility (measured by R) at a fine spatio-temporal scale and any covariate measured at the same scale, may be

Downloaded from http://rstb.royalsocietypublishing.org/ on June 6, 2017

of interest. For instance, analyses of the West African Ebola epidemic showed that districts with lower reported funeral attendance and faster hospitalization experienced lower transmissibility, highlighting the effectiveness of promoting safe burials and early hospitalization [10].

However, interpreting the results of such analyses can be challenging, as they might be prone to bias and confounding. Efficacy (which measures the impact of an intervention under ideal and controlled circumstances) and effectiveness (which measures the impact of an intervention under real-world conditions) of an intervention (e.g. vaccine) are best measured in a trial setting (either individual- or cluster-randomized [68–70]). However, performing trials to evaluate the comparative impact of different multi-intervention packages is impractical. Dynamic epidemic models, where the interventions of interest can be explicitly incorporated, allow the impact of such intervention packages to be predicted [71]. However, outputs of such models are strongly determined by the underlying assumptions and parameter values. Hence they require careful parametrization, supported by data such as intervention efficacy and the size, infectivity, susceptibility and mixing of different groups [72,73]. These parameters may not be straightforward to estimate, as we discuss in §3 using examples from the West African Ebola epidemic.

Another factor determining the appropriate choice of interventions is their cost, combinations of interventions with higher effectiveness at lower cost (i.e. higher cost-effectiveness) being preferable. Economic analyses combined with mathematical models can help to evaluate the optimal resource allocation among both current available interventions and potential new interventions, accounting for development and testing costs for the latter. Indirect costs, e.g. those associated with a restricted workforce following school closures [74], or trade limitations from air-travel restrictions [75], also need to be considered. While economic analyses have played an important role in designing optimal intervention packages for endemic diseases such as HIV and malaria [76,77], such analyses are more difficult to perform during an epidemic, when cost data might be unavailable and uncertain, costs may vary rapidly over time and ethical considerations suggest interventions should be implemented immediately.

(c) Are current interventions effective and could they be improved?

Tracking changes in estimates of key epidemiological parameters over the course of an outbreak enhances situational awareness. It also allows the impact of interventions to be assessed as they are implemented, although disentangling the effects of different interventions carried out simultaneously may be challenging.

3. Data requirements

Obtaining reliable estimates of the epidemiological parameters detailed above requires a wide range of data, such as incidence time series and detailed case information (figure 1). Here, we explain how these can be obtained from various sources, with the objective to help improve data collection systems in preparation for future outbreaks. We use Ebola as a specific example throughout this section, commenting on the strengths and limitations of the data collected during the West African epidemic.

We distinguish data needs at the individual level, the exposure level and the population level.

(a) Individual-level data: case line-list

Simple analyses can be performed solely using incidence time series, from surveillance designed to capture aggregate case counts. However, individual case reports provide much richer information, essential to estimate many of the key parameters outlined above (e.g. characterization of delay distributions). Such data are typically stored in a case database or 'line-list'—a table with one line containing individual data for each case. The more data recorded on each case, the more detailed the analyses can be. In the Ebola epidemic, demographic characteristics, spatial location, laboratory results and clinical details such as symptoms, hospitalization status, treatment and outcome, and dates associated with these were reported for at least a subset of cases. The appropriate information to collect may vary depending on the disease: for Ebola, dates of isolation and funeral were relevant. Comprehensive demographic information can be used to determine risk factors for transmission or severity of infection and to project case numbers stratified by demographic characteristics. Detailed information also helps to identify and merge any duplicate entries in a line-list, which may occur when the same person visits multiple health centres over the course of illness, for instance. Finally, information on how each case was detected—for example, through hospitalization, or via contact tracing—can aid assessment of how representative the data are and allow subsequent adjustment for bias [28]; however, this was not available for Ebola.

Cases in the line-list should be classified using standardized case definitions, which is sometimes difficult for outbreaks of new pathogens, or where different case definitions are provided by different organizations (e.g. World Health Organization (WHO) and US Centers for Disease Control and Prevention (CDC)) [78]. For the Ebola response, although the WHO released official case definitions of confirmed, probable and suspected Ebola cases [79], different countries adopted different testing strategies, thereby limiting the opportunity for inter-country comparison. For example, in Guinea deceased individuals were not tested for Ebola, meaning these individuals were never classified as confirmed cases, unlike in Liberia and Sierra Leone. Encouraging use of a consistent testing protocol and case definition, and ensuring transparency in what is used where and when, would improve the validity of subsequent analyses.

Laboratory testing of clinical specimens is key for confirming cases and test results should be linked to the line-list. Understanding diagnostic test performance in field conditions is important; cross-validation of diagnostic sensitivity and specificity between laboratories is useful to assess the extent to which observed differences in case incidence may be explained by variations in laboratory conditions and practices.

In addition, recording raw test results with the case classification may help evaluation of diagnostic performance. Ebola cases were defined as confirmed cases once Ebola virus RNA was isolated from clinical specimens using reverse transcription polymerase chain reaction (RT-PCR [79]). Within each of the main affected countries, field laboratories were established [80,81] to enable prompt diagnosis of cases. Recent evidence suggests that the large fluctuations in

Downloaded from http://rstb.royalsocietypublishing.org/ on June 6, 2017

temperature and humidity to which these laboratories were subjected reduced the test performance compared with manufacturer evaluation reports [82].

During the West African Ebola epidemic, the case line-list contained a large quantity of data collected from reported cases. The information allowed estimation of the CFR, the incubation period distribution (and evaluation of differences in these by age and gender) and the reproduction number [3,4,6]. Projections of the likely scale of the outbreak were also made, either from the line-list or from aggregated case counts [3,8,83,84]. There were regular data updates [5], with a total of over 19 000 confirmed and probable cases reported, which allowed analyses to be updated as more data became available.

(b) Exposure-level data

Data on exposures allow cases to be linked to their potential sources of infection, and hence provide a better understanding of transmission characteristics.

(i) Transmission pairs

The relevant modes (e.g. airborne, foodborne) and pathways (e.g. animal–human, human–human) of transmission may be identified using information on exposure reported by cases. Cases can report contact both with sick individuals (their potential source of infection) and healthy individuals they have contacted since becoming ill (who may need to be traced and monitored as potential new cases). These data will be more informative if the majority of infections are symptomatic (and hence easily identifiable), if individuals are mostly infectious after the onset of symptoms [43] and if the time window over which exposures and contacts are monitored is as long as the upper bound of the incubation period distribution. If exposure information is collected with enough demographic information to allow record linkage, these backward and forward contacts can be identified in the case line-list, defining transmission pairs. Depending on the mutation rate of the pathogen, genetic data can also be used to identify or confirm these epidemiological links [56,85]. The increasing availability of full genome pathogen sequences offers exciting prospects in that respect. Some genetic sequence information was available during the Ebola epidemic, but most sequences could not be linked to case records, limiting the use of sequence data in this context. On the other hand, individuals who were named as potential sources of infection could often be identified in the case line-list, although this process was hindered by non-unique names and limited demographic information collected on the potential sources [10]. These exposure data were used to characterize variation in transmission over the course of infection [10], and to estimate the serial interval and the incubation period [3]. The upper bound of the incubation period distribution was estimated to be 21 days [3], which supported monitoring contacts for up to three weeks.

(ii) Transmission studies

Studies of transmission in well-defined, small settings such as households are useful to quantify asymptomatic transmission, infectivity over time and the serial interval as they capture explicitly the number and timing of secondary cases. Additionally, these studies can estimate the secondary attack rate (the proportion of contacts of a case who become infected within one incubation period), which can be used to characterize heterogeneities in transmission of different groups [59]. Estimates of the secondary attack rate have been obtained for the West African Ebola epidemic by reconstructing household data based on information reported by cases, in particular, as part of contact-tracing activities [86,87].

(c) Population-level or metadata

Although they might not immediately appear as useful as individual- or exposure-level data, metadata are crucial for answering many public health questions.

(i) Population sizes

Knowing the sizes of affected populations is important for quantifying the attack rate and informing dynamic transmission models. Census data are likely to be the most reliable source, but may be infrequently collected. Methods based on interpreting satellite imagery [88,89] can inform population size and structure, although demographic stratifications are not always available. For the West African Ebola epidemic, the most recently available age- and gender-stratified population census data were from 1996 in Guinea [90], 2004 in Sierra Leone [91] and 2008 in Liberia [92].

A particular population of interest is HCWs who, due to their contact with patients, are often at high risk of infection and may also be high-risk transmitters. Large numbers of HCWs were infected during the West African Ebola epidemic [60,61]. However, the proportion of HCWs affected at different stages of the outbreak and the relative risk of acquisition for the HCWs compared with the general population could not be estimated since the total number of HCWs was not systematically reported and changed during the course of the outbreak with the scale-up of interventions. Note that, depending on the transmission route, the definition of HCWs may need to include anyone working in a health-care setting who could be at risk (e.g. cleaners may be exposed to bodily fluids).

(ii) Mobility

Characterizing population movement is crucial to assessing the risk of exportation of the infection from one place to another. Air-travel data are the most reliable, consistently available and commonly used data source to inform models of long-distance spread [93]. Such data were widely used during the West African Ebola epidemic to quantify the risk of international spread of the disease, and to assess the potential impact of airport screening and travel restrictions on the outbreak [9,94–96].

However, air travel does not cover other population movements that may play an important role in disease spread, e.g. travel by road or on foot in Guinea, Liberia and Sierra Leone and across the porous country borders during the West African Ebola epidemic [97]. Usually, little data are available to directly characterize these typically smaller-scale movements. Gravity models, which assume that connectivity between two places depends on their population sizes and the distance between them, can be used to quantify spatial connectivity between different regions [98–101], and have proved useful to predict local epidemic spread, e.g. for Chikungunya [98]. Such models require data on population sizes and geographical distances.

Recently, mobile phone data have been explored as an alternative source of data on mobility patterns, which could be used to predict spatial epidemic spread [102–104]. However, a number of challenges (in particular related to privacy issues) meant that such data were unavailable to

Downloaded from http://rstb.royalsocietypublishing.org/ on June 6, 2017

understand the regional and local spread of the West African Ebola epidemic [105]. In addition, inter-country movement is not captured from these data. At a national level, the utility of mobile phone data may depend on whether mobile phone users are representative of the population contributing to transmission, the level of mobile phone coverage in the affected population and whether population movement is likely to remain the same during an epidemic compared to the time period of the data.

(iii) Seroprevalence

Assessing seroprevalence in a population affected by an outbreak can provide valuable information on the underlying scale of population exposure and insight into how interventions might be targeted. For instance, if there is pre-existing population immunity prior to an outbreak that varies with age (as was the case in the 2009 H1N1 influenza pandemic), vaccination could be targeted at those with lower pre-existing immunity. Dynamic transmission models incorporating such differences in susceptibility can be used to explore different targeting strategies [106].

Ideally, serological surveys would be undertaken to quantify seroprevalence [107]; however, they are expensive and not performed on a regular basis. In the absence of such data, information on historical outbreaks and vaccine use can sometimes be used to infer seroprevalence [108]. Serological studies performed before and after an epidemic can also be useful to measure the attack rate and the scale of the outbreak, and hence provide information on the level of underreporting during the outbreak [109].

It was widely assumed that the population in West Africa was entirely susceptible to Ebola at the start of this outbreak, with no known previous outbreaks in the area. However, some studies have suggested that there might have been low levels of prior immunity [110].

(iv) Recording intervention efforts across time and space

During an outbreak, multiple interventions are often implemented by different groups and organizations. Evaluating the role of interventions in interrupting transmission is important for revising and improving efforts, but it is challenging without detailed quantitative information of what has been implemented where and when [111,112]. Maintaining a systematic real-time record of the different interventions at a fine spatio-temporal scale would help, e.g. the number and location of health-care facilities and their personnel, number of beds, vaccine or treatment coverage and details of local community mobilization. Developing centralized platforms to routinely record such data once a large-scale outbreak is underway is probably unfeasible. However, developing such tools in advance of outbreaks (such as those developed for the Global Polio Eradication Initiative [113] and those recently developed to collect health-care facility data [114]) should be a priority since better information to evaluate intervention policies in real time will allow for more optimal resource allocation.

During the West African Ebola epidemic, many data on interventions were recorded at a local level by some of the numerous partners (e.g. non-governmental organizations (NGOs) and other organizations) involved in the response. For example, some data were collected on the number and capacity of hospitals over time [115] and these were used in a study modelling community transmission to assess the impact of increasing hospital bed capacity [116]. However, the decentralization of the response meant that intervention data were not systematically reported or collated and these data were not shared widely with the research community. A failure to report interventions centrally and systematically can make it difficult to disentangle a lack of intervention effect from a lack of intervention implementation. This can particularly be a problem when numerous groups coordinate their own efforts, making it impossible to draw firm conclusions about interventions. In the absence of detailed data on intervention efforts in West Africa, multiple studies have assessed the combined impact of all interventions in place, by comparing transmissibility in the early phase (with no intervention) to that in the later phases [87,117]. However, this approach provides less compelling evidence of a causal effect and does not disentangle the impact of different interventions performed at the same time, and hence is less informative for future response planning.

(v) Quantifying intervention efficacy

Vaccine or treatment trials together with case–control and cohort studies can be useful in assessing the impact of an intervention. For example, during the West African Ebola epidemic there was an urgent need to estimate the efficacy of newly developed vaccines. Trials such as the Guinea *Ebola ça suffit* vaccine trial [118] provided key data on the effectiveness of the rVSV-ZEBOV Ebola vaccine [119,120]. These trials occurred at the tail end of the epidemic and results will be useful in future outbreaks. Statistical power from trials will be maximized by implementing such studies as early as possible in future outbreaks. This will be facilitated if research on diagnostics, drugs and vaccines is promoted between, and not only during, outbreaks, e.g. through new initiatives such as the Coalition for Epidemic Preparedness Innovation [121].

(vi) Quantifying data completeness and timeliness

All of the data sources mentioned above are inevitably imperfect; what they are trying to measure is different from what they measure in practice. Quantifying the mismatch between the two is vital to appropriately account for these imperfections.

For instance, case line-lists are likely to contain information on only a proportion of all infected individuals: typically those with symptoms, or those who sought care. The level of reporting may also be influenced by the capacity of the local health systems, which can vary over time and space. During the West African Ebola epidemic, less than a third of cases were estimated to be reported [122] and severe cases were probably over-represented compared to mild cases. At the end of 2014, health-care capacity was exceeded in many parts of Guinea, Liberia and Sierra Leone [14], but new health-care facilities were subsequently built; hence the line-list of cases is likely to be more complete towards the end of the outbreak. Underreporting might also have been higher in this compared to previous Ebola outbreaks, during which the health-care systems were less overburdened. Systematic evaluation of the surveillance system [123] over different spatial units and time periods could help inform the level of underreporting. In addition, joint analysis of genetic sequence and surveillance data can provide insight into the degree of underreporting [124]. Quantifying completeness of, and potential biases in the line-list is important, e.g. to adequately quantify the CFR [28]. Although differences in the CFR were observed across

Downloaded from http://rstb.royalsocietypublishing.org/ on June 6, 2017

different health-care facilities in the West African Ebola epidemic, it was not possible to determine whether these were due to reporting differences or underlying differences between settings [125].

Similarly, exposure-level data can be incomplete, depending on the available capacity (personnel and resources) to perform contact tracing and the willingness of people to share information on their contacts. Complete data can be used to assess the route of transmission—animal to human or human to human—and the number of cases imported from other locations. If data are incomplete, these estimates may be incorrect.

Population-level data may suffer the same issues. For instance, the recording of intervention efforts (e.g. the number of personal protection equipment (PPE) kits distributed) can differ from the reality of the intervention (e.g. the number of people who used PPE in practice). Quantifying this mismatch is crucial to evaluating the impact of various interventions, and requires dedicated studies, with both qualitative and quantitative components, on the acceptability of and the adherence to given interventions. Such studies have been carried out in the past, e.g. to assess the potential impact of face masks on the risk of influenza transmission in households, or of condom use on the risk of HIV transmission [126–128].

It is also important to quantify delays encountered in the reporting of cases [129]. During the West African Ebola epidemic, there were delays in all databases and disparities between different data sources. In particular, comparison between the line-list of cases and aggregated daily case counts reported by the affected countries highlighted reporting delays in the line-list. As a result, at any point in the outbreak, naive time series of case counts derived from the line-list suggested that the epidemic was declining, due to right-censoring. Comparison between the line-list and reported aggregated case counts (which were more up-to-date) and between successive versions of the line-list allowed the reporting delays to be quantified. Analyses such as the projected incidence could then be adjusted accordingly [5].

4. Successes and challenges in data collection and analyses during the West African Ebola epidemic

In summary, the enormous quantity of detailed data collected during the West African Ebola epidemic played an invaluable role in guiding response efforts. However, several analyses could not be performed. Early on, severity, transmissibility and delay distributions were quantified and short-term projections were made, based on the case line-list and/or contact tracing data [3,5,7]. Some heterogeneity in severity and transmissibility could be identified (e.g. by age and viraemia [7,10,116,125,130–132]), but other types of heterogeneity could not be assessed. For example, it was not possible to compare the CFR between hospitalized and non-hospitalized cases, because of biases in the way cases were recorded in the line-list [28]. Long-term projections were extremely challenging due to large uncertainty in population sizes, behaviour changes and changing intervention efforts. In addition, the relative risk of Ebola acquisition for HCWs was difficult to estimate due to the absence of reliable spatio-temporal data on the number of HCWs. Finally, systematic evaluation of interventions was not feasible due to multiple control measures being carried out at the same time, with little central recording of details of each intervention.

5. Optimizing data quality

Data collection during the West African Ebola epidemic was possible, in part, due to a pre-existing case investigation form and data management system (Epi Info viral haemorrhagic fever (VHF) application [133,134]). For outbreaks of new pathogens, such systems are not usually in place. The list of data needs we have outlined above, based on our experience during the Ebola epidemic, could serve as a basis for data collection in future outbreaks. Here we outline improvements that could be made to streamline data collection, minimize delays between data collection and data dissemination, and improve data quality during future outbreaks. All of these are necessary for timely analysis to inform the response in real time. We consider this particularly for line-list and exposure data.

(a) Streamlining data collection and digitalization

Data need to be digitized before they can be analysed; streamlining this process reduces the potential for delays and errors. Using electronic data capture on tablets or phones may reduce delays and errors compared with using paper forms that then require manual digitization, with its contingent issues [135–138]. Electronic questionnaires may not be available at the start of the epidemic, but could be quickly adopted using available tools (e.g. EpiCollect [139] or EpiBasket [140]). Interpretation problems (e.g. abbreviations) and spelling errors could also be minimized at the data collection level by using multiple choice questions rather than free text and at the data entry level by using drop-down menus. For example, for spatial information the choices could be the standard administrative levels used in the country—such as district or county. As response efforts evolve over time, such as the building of new treatment centres, the lists would be updated as required. Similarly, using pop-out calendars to select dates would limit typing errors. Additionally, internal consistency checks could flag problems such as the recorded date of death being before that of symptom onset, and the system could force a manual check before the record is saved.

In the West African Ebola epidemic, data were collected using paper forms (electronic supplementary material, figures S1–S3) and manually entered into an electronic system at local operation centres. Data entry problems were particularly noticeable in clinical dates and free text variables, such as district locations and hospital names, and required considerable cleaning (e.g. to correct spelling errors) [3,5,10].

(b) Ensuring adequate logistical resources and training

With any data system there is potential for delays in data entry and dissemination due to limitations in personnel and hardware, and logistical constraints in data delivery (e.g. reliable electricity, Internet access and transport). Although the latter limitations are arguably outside the scope of outbreak control, delays from data collection to dissemination could be reduced by increasing training in data entry and providing more data entry facilities where possible. Delays in data entry and release during the Ebola epidemic meant that real-time analyses of the line-list data could not be performed on fully up-to-date data.

Downloaded from http://rstb.royalsocietypublishing.org/ on June 6, 2017

(c) Acknowledging the international nature of the threat

In the world today, an epidemic emerging anywhere is a global threat due to high population connectivity [141,142] and, as such, a response will have to operate across multiple languages. Ideally, the same data entry system would be used in every affected country to allow easy collation into a single case line-list. However, a global response requires careful translation of the questions and the system to ensure they make sense and are identical in every language. Additionally, different languages, dialects or alphabets could be challenging due to differing pronunciations, accents, alternate names or spellings, and these should be acknowledged in the form design. The language barrier was not a major problem during the Ebola response: Guinea had a form in French, while in Sierra Leone and Liberia the form was in English, though there were some minor differences in formulations between the two (for example the English version of the form asked 'In the past one month prior to symptom onset: did the patient have contact with (. . .) any sick person before becoming ill' (see §4 in the form in electronic supplementary material, figure S1), while the French form did not specify 'before becoming ill').

(d) Improving harmonization between databases

Analysis of laboratory results and sequence data can be much more powerful if they can be dated and linked to the epidemiological data recorded for each case. In the early stages of the Ebola response, there were reports that laboratory results could not always be linked to case records as labels were incorrectly written or damaged in transit [81]. Later in the epidemic, case report forms came with pre-printed unique ID barcode stickers to label all records and samples for each case (electronic supplementary material, Form S2). This would be useful if implemented early in future outbreaks, particularly if laboratory tests are not performed at the point of care. Rapid diagnostic tests were developed during the West African Ebola epidemic but not used widely [143,144]; similarly, mobile sequence tests were introduced later in the epidemic [145]: both of these would reduce delays and maximize the potential to link patient data.

6. Data availability

We have described a set of data needed to perform analyses to inform the public health response during an outbreak. We have proposed strategies to ensure fast and high-quality data collection, to enable robust and timely analyses. However, such analyses can only be performed if the data are accessible to data analysts and modellers.

(a) Data access

The West African Ebola epidemic has prompted an ongoing public debate about the ethical, practical and scientific implications of wide data access [14,146–148]. Ethical considerations require removal of data that might compromise anonymity, but such detailed information might be required to answer important public health questions. Mechanisms also need to be found to appropriately acknowledge those who collected and digitalized the data. From a scientific perspective, having several groups analysing the data is desirable, as independent analyses leading to similar results will reinforce their utility for policymaking [3,8]. Such parallelized efforts have been formalized through consortiums of highly experienced groups, e.g. for modelling of HIV [149], influenza [150,151] and malaria [152]. However, if results differ, understanding what assumptions drive these differences may confuse and delay the process of decision-making. A consequence of data sharing, therefore, needs to be an increased emphasis on evidence synthesis, such as systematic reviews of the different analyses. To enable this process, groups should explicitly state all assumptions underlying their analyses and which data they are based on. Like the original analyses, reviews need to be timely to be useful. These issues have partly been addressed through new data availability policies [153]. Furthermore, this process would ideally be performed by a group independent of those performing the primary analyses. Identifying an effective system, appropriate personnel and appropriate recognition for this important role needs to be planned in advance of future outbreaks.

(b) Data curating and data cleaning

Data from the West African Ebola epidemic required significant cleaning before it could be analysed, and this was necessary for every updated dataset, i.e. every few days during the peak. Were outbreak data to be shared more widely, collaborative or centralized cleaning would be optimal, to avoid repetition and ensure consistency across different groups. If a centralized effort was not possible, regular sharing of the cleaning code and cleaned datasets between groups would facilitate comparison of results. Even better, code could be shared on a collaborative platform such as Github (https://github.com/), leading to a common clean dataset being actively maintained by the scientific community. However, for this process to be effective, a set of best practices would need to be established in advance, such as designing a transparent workflow, establishing a fair distribution of tasks and clarifying how credit will be given for this (often lengthy) task. This is important to avoid duplication of effort and allow effective collaboration. As the process of data sharing is debated further, it is critical that the practicalities of data cleaning are discussed in parallel. Based on our experience, a centralized cleaning platform would be the most effective method.

(c) Disseminating results of data analyses

Finally, analyses will be most useful if they are shared widely across policymakers, local health teams and other research groups. The format for dissemination could be anything from a report to a scientific publication. Reports can be made available faster and regularly updated but do not undergo the peer review process of publications. Peer review can delay publication, though there are now new platforms for fast-tracking this process [154,155]. At the time of writing, as interest in Zika virus is gaining momentum, there seems to be an encouraging trend to the use of pre-print servers [156,157].

7. Discussion

By the very nature of emerging infectious diseases, we do not know which pathogen will emerge next, when or where. There have been many suggestions about how to be

Downloaded from http://rstb.royalsocietypublishing.org/ on June 6, 2017

better prepared [11–21,24,26]; here, we argue that preparedness should include development of a broad-use data collection system that can be easily and quickly adapted to any disease (in agreement with [11]) as well as the regular collection of population health data in centralized systems. Different infectious diseases may require different types of data [140]: a single approach is not applicable to all diseases. In particular, different interventions may need to be recorded for different diseases.

Data collection and management may need to evolve as the outbreak progresses and/or as more is learnt about the pathogen: the Ebola data collection forms changed late in 2014 to streamline collection and entry when the response effort was almost overwhelmed with cases (electronic supplementary material, figures S1 and S2). Similarly, both for Ebola and the recent Zika outbreak, reports of sexual transmission have led to a broadening of the contact tracing and exposure information collected.

The data collected during the Ebola epidemic allowed many analyses to be performed, which informed the response. However, as in many outbreak situations, it has not been possible to systematically quantify the relative contribution of different interventions (such as safe burials, hospitalization, contact tracing and community mobilization) in reducing transmission. This is because data on where and when these interventions took place were not centrally recorded and released in a timely fashion. Efforts have been made to collate some information from Sierra Leone, Liberia and Guinea [115]; however, this commenced late in the epidemic. As this effort relies on different organizations to contribute data, the submissions are in different formats and are unlikely to provide comprehensive information. In the midst of a global public health crisis resources are often deployed favouring implementation rather than documentation of interventions. However, we would argue that securing some resources to monitor interventions—especially during the early stages—is critical to optimally prioritizing future control efforts.

Some of the data we have suggested to be collected during outbreaks may not be obviously useful at the field or case management level, e.g. detailed demographic characteristics of cases and contacts. Collecting such data costs money and time as well as trained personnel. These three 'resources' are limited and, during an epidemic, should be prioritized where their need is greatest. However, from a population perspective, collecting these data may help quantify epidemic impact, assist in the design and evaluation of interventions, and help prevent new infections. Further studies might examine how to appropriately balance these two considerations.

We have built on the Ebola experience to draw up a list of the data needed to assist the response throughout an epidemic, which should help to collect relevant data in a standardized effort in future epidemics. To make the most of these data, epidemiologists and modellers should work now to develop tools to automatically clean, analyse and report on the data in a more timely and robust manner [158]. Based on critical review of past outbreak analyses, future studies could flag common methodological mistakes and propose good practice [159]. This includes clearly stating all assumptions underlying a model or analysis and ensuring that parametrization is either directly informed by relevant data or has appropriate sensitivity analyses, with corresponding uncertainty clearly reported [32,160].

Improving our ability to respond effectively to the next outbreak will require collaboration between all parties involved in outbreak response: those in the field, epidemiologists, modellers and policymakers as well as the populations affected. Here we have given the data analyst perspective on what data are required to answer important policy questions. It is equally important that other perspectives should be heard to be better prepared for and improve interactions during crises, thereby minimizing the impact of future outbreaks.

Authors' contributions. A.C., H.L.M. and I.M.B. conceived and designed the study, and wrote the first draft of the paper. All authors contributed to revisions and gave their final approval for publication.

Competing interests. We have no competing interests.

Funding. Authors are supported by the Medical Research Council (MRC), the Bill and Melinda Gates Foundation, the Models of Infectious Disease Agent Study of the National Institute of General Medical Sciences (National Institutes of Health), the Health Protection Research Units of the National Institute for Health Research, European Union PREDEMICS consortium (grant no. 278433), Wellcome Trust, Imperial College Junior Research Fellowship, the Forgarty International Center and the Engineering and Physical Sciences Research Council (EPSRC). The funders had no role in the study.

Acknowledgements. The authors give credit to and thank the many individuals who were involved in data collection, entry and management during the West African Ebola epidemic.

References

1. Van Kerkhove MD, Bento AI, Mills HL, Ferguson NM, Donnelly CA. 2015 A review of epidemiological parameters from Ebola outbreaks to inform early public health decision-making. *Sci. Data* **2**, 150019. (doi:10.1038/sdata.2015.19)
2. WHO Ebola Response Team. 2016 After Ebola in West Africa: unpredictable risks, preventable epidemics. *N. Engl. J. Med.* **375**, 587–596. (doi:10.1056/NEJMsr1513109)
3. WHO Ebola Response Team. 2014 Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections. *N. Engl. J. Med.* **371**, 1481–1495. (doi:10.1056/NEJMoa1411100)
4. WHO Ebola Response Team. 2016 Ebola virus disease among male and female persons in West Africa. *N. Engl. J. Med.* **374**, 96–98. (doi:10.1056/NEJMc1510305)
5. WHO Ebola Response Team *et al.* 2015 West African Ebola epidemic after one year—slowing but not yet under control. *N. Engl. J. Med.* **372**, 584–587. (doi:10.1056/NEJMc1414992)
6. WHO Ebola Response Team *et al.* 2015 Ebola virus disease among children in West Africa. *N. Engl. J. Med.* **372**, 1274–1277. (doi:10.1056/NEJMc1415318)
7. Faye O *et al.* 2015 Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: an observational study. *Lancet Infect. Dis.* **15**, 320–326. (doi:10.1016/S1473-3099(14)71075-8)
8. Meltzer MI *et al.* 2014 Estimating the future number of cases in the Ebola epidemic—Liberia and Sierra Leone, 2014-2015. *MMWR Suppl.* **63**, 1–14.
9. Read JM, Diggle PJ, Chirombo J, Solomon T, Baylis M. 2015 Effectiveness of screening for Ebola at airports. *Lancet* **385**, 23–24. (doi:10.1016/S0140-6736(14)61894-8)
10. The International Ebola Response Team. 2016 Exposure patterns driving Ebola transmission in West Africa, a retrospective observational study. *PLoS Med.* **13**, e1002170. (doi:10.1371/journal.pmed.1002170)

Downloaded from http://rstb.royalsocietypublishing.org/ on June 6, 2017

11. Burki T. 2016 Are we learning the lessons of the Ebola outbreak? *Lancet Infect. Dis.* **16**, 296–297. (doi:10.1016/S1473-3099(16)00080-3)
12. Gates B. 2015 The next epidemic—lessons from Ebola. *N. Engl. J. Med.* **372**, 1381–1384. (doi:10.1056/NEJMp1502918)
13. Heymann DL *et al.* 2015 Global health security: the wider lessons from the West African Ebola virus disease epidemic. *Lancet* **385**, 1884–1901. (doi:10.1016/S0140-6736(15)60858-3)
14. Moon S *et al.* 2015 Will Ebola change the game? Ten essential reforms before the next pandemic. The report of the Harvard-LSHTM Independent Panel on the Global Response to Ebola. *Lancet* **386**, 2204–2221. (doi:10.1016/S0140-6736(15)00946-0)
15. Beavogui AH, Delamou A, Yansane ML, Konde MK, Diallo AA, Aboulhab J, Bah-Sow OY, Keita S. 2016 Clinical research during the Ebola virus disease outbreak in Guinea: lessons learned and ways forward. *Clin. Trials* **13**, 73–78. (doi:10.1177/1740774515619877)
16. Kalra S *et al.* 2014 The emergence of Ebola as a global health security threat: from 'lessons learned' to coordinated multilateral containment efforts. *J. Glob. Infect. Dis.* **6**, 164–177. (doi:10.4103/0974-777X.145247)
17. Koch T. 2016 Ebola in West Africa: lessons we may have learned. *Int. J. Epidemiol.* **45**, 5–12. (doi:10.1093/ije/dyv324)
18. Morgan DJ *et al.* 2015 Lessons learned from hospital Ebola preparation. *Infect. Control Hosp. Epidemiol.* **36**, 627–631. (doi:10.1017/ice.2015.61)
19. Petherick A. 2015 Ebola in west Africa: learning the lessons. *Lancet* **385**, 591–592. (doi:10.1016/S0140-6736(15)60075-7)
20. Siedner MJ, Gostin LO, Cranmer HH, Kraemer JD. 2015 Strengthening the detection of and early response to public health emergencies: lessons from the West African Ebola epidemic. *PLoS Med.* **12**, e1001804. (doi:10.1371/journal.pmed.1001804)
21. Thielman NM, Cunningham CK, Woods C, Petzold E, Sprenz M, Russell J. 2016 Ebola clinical trials: five lessons learned and a way forward. *Clin. Trials* **13**, 83–86. (doi:10.1177/1740774515619897)
22. Woolhouse MEJ, Rambaut A, Kellam P. 2015 Lessons from Ebola: improving infectious disease surveillance to inform outbreak management. *Sci. Transl. Med.* **7**, p307rv305. (doi:10.1126/scitranslmed.aab0191)
23. Owada K, Eckmanns T, Kamara KB, Olu OO. 2016 Epidemiological data management during an outbreak of Ebola virus disease: key issues and observations from Sierra Leone. *Front. Public Health* **4**, 163. (doi:10.3389/fpubh.2016.00163)
24. Lipsitch M, Finelli L, Heffernan RT, Leung GM, Redd SC, H1N1 Surveillance Group. 2011 Improving the evidence base for decision making during a pandemic: the example of 2009 influenza A/H1N1. *Biosecur. Bioterror.* **9**, 89–115. (doi:10.1089/bsp.2011.0007)
25. Van Kerkhove MD, Ferguson NM. 2012 Epidemic and intervention modelling—a scientific rationale for policy decisions? Lessons from the 2009 influenza pandemic. *Bull. World Health Organ.* **90**, 306–310. (doi:10.2471/BLT.11.097949)
26. Galvani AP. 2004 Emerging infections: what have we learned from SARS? *Emerg. Infect. Dis.* **10**, 1351–1352. (doi:10.3201/eid1007.040166)
27. Garske T, Legrand J, Donnelly CA, Ward H, Cauchemez S, Fraser C, Ferguson NM, Ghani AC. 2009 Assessing the severity of the novel influenza A/H1N1 pandemic. *Br. Med. J.* **339**, b2840. (doi:10.1136/bmj.b2840)
28. Lipsitch M *et al.* 2015 Potential biases in estimating absolute and relative case-fatality risks during outbreaks. *PLoS Negl. Trop. Dis.* **9**, e0003846. (doi:10.1371/journal.pntd.0003846)
29. Pelat C, Ferguson NM, White PJ, Reed C, Finelli L, Cauchemez S, Fraser C. 2014 Optimizing the precision of case fatality ratio estimates under the surveillance pyramid approach. *Am. J. Epidemiol.* **180**, 1036–1046. (doi:10.1093/aje/kwu213)
30. Nicoll A *et al.* 2012 Influenza-related deaths–available methods for estimating numbers and detecting patterns for seasonal and pandemic influenza in Europe. *Euro Surveill.* **17**, pii. 20162.
31. Vynnycky E, White RG. 2010 *An introduction to infectious disease modelling*. Oxford, UK: Oxford Univeristy Press.
32. Metcalf CJ, Edmunds WJ, Lessler J. 2015 Six challenges in modelling for public health policy. *Epidemics* **10**, 93–96. (doi:10.1016/j.epidem.2014.08.008)
33. May RM. 2004 Uses and abuses of mathematics in biology. *Science* **303**, 790–793. (doi:10.1126/science.1094442)
34. Person B, Sy F, Holton K, Govert B, Liang A, National Center for Infectious Diseases SARS Community Outreach Team. 2004 Fear and stigma: the epidemic within the SARS outbreak. *Emerg. Infect. Dis.* **10**, 358–363. (doi:10.3201/eid1002.030750)
35. Mikler AR, Venkatachalam A, Ramisetty-Mikler A. 2007 Decisions under uncertainty: a computational framework for quantification of policies addressing infectious disease epidemics. *Stochast. Environ. Res. Risk Assess.* **21**, 533. (doi:10.1007/s00477-007-0137-y)
36. Thompson KM *et al.* 2006 Development and consideration of global policies for managing the future risks of poliovirus outbreaks: insights and lessons learned through modeling. *Risk Anal.* **26**, 1571–1580. (doi:10.1111/j.1539-6924.2006.00841.x)
37. Jha SK, Ramanathan A. 2012 Quantifying uncertainty in epidemiological models. In *Biomedical computing* (BioMedCom), 2012, pp. 80–85. ASE/IEEE International Conference.
38. Stone SP *et al.* 2007 The ORION statement: guidelines for transparent reporting of outbreak reports and intervention studies of nosocomial infection. *Lancet Infect. Dis.* **7**, 282–288. (doi:10.1016/S1473-3099(07)70082-8)
39. Ma J, Earn DJ. 2006 Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bull. Math. Biol.* **68**, 679–702. (doi:10.1007/s11538-005-9047-7)
40. Becker NG. 2015 *Modeling to inform infectious disease control*, 208. London, UK: Chapman and Hall.
41. Funk S, Bansal S, Bauch CT, Eames KT, Edmunds WJ, Galvani AP, Klepac P. 2015 Nine challenges in incorporating the dynamics of behaviour in infectious diseases models. *Epidemics* **10**, 21–25. (doi:10.1016/j.epidem.2014.09.005)
42. Lloyd-Smith JO, Funk S, McLean AR, Riley S, Wood JL. 2015 Nine challenges in modelling the emergence of novel pathogens. *Epidemics* **10**, 35–39. (doi:10.1016/j.epidem.2014.09.002)
43. Fraser C, Riley S, Anderson RM, Ferguson NM. 2004 Factors that make an infectious disease outbreak controllable. *Proc. Natl Acad. Sci. USA* **101**, 6146–6151. (doi:10.1073/pnas.0307506101)
44. Anderson RM, May RM. 1991 *Infectious diseases of humans: dynamics and control*. Oxford, UK: Oxford University Press.
45. Heesterbeek JA. 2002 A brief history of R_0 and a recipe for its calculation. *Acta Biotheor.* **50**, 189–204.
46. Cori A. 2013 EpiEstim: a package to estimate time varying reproduction numbers from epidemic curves (R package, version 1.1-2 ed).
47. Cori A, Ferguson NM, Fraser C, Cauchemez S. 2013 A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* **178**, 1505–1512. (doi:10.1093/aje/kwt133)
48. Wallinga J, Lipsitch M. 2007 How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. B* **274**, 599–604. (doi:10.1098/rspb.2006.3754)
49. Wallinga J, Teunis P. 2004 Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. Epidemiol.* **160**, 509–516. (doi:10.1093/aje/kwh255)
50. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. 2005 Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359. (doi:10.1038/nature04153)
51. Grassly NC, Fraser C. 2008 Mathematical models of infectious disease transmission. *Nat. Rev. Microbiol.* **6**, 477–487. (doi:10.1038/nrmicro1845)
52. Cauchemez S, Carrat F, Viboud C, Valleron AJ, Boelle PY. 2004 A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Stat. Med.* **23**, 3469–3487. (doi:10.1002/sim.1912)
53. Viboud C, Boelle PY, Cauchemez S, Lavenu A, Valleron AJ, Flahault A, Carrat F. 2004 Risk factors of influenza transmission in households. *Br. J. Gen. Pract.* **54**, 684–689.
54. Woolhouse MEJ *et al.* 1997 Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl Acad. Sci. USA* **94**, 338–342.
55. Didelot X, Gardy J, Colijn C. 2014 Bayesian inference of infectious disease transmission from whole-genome sequence data. *Mol. Biol. Evol.* **31**, 1869–1879. (doi:10.1093/molbev/msu121)

Downloaded from http://rstb.royalsocietypublishing.org/ on June 6, 2017

56. Jombart T, Cori A, Didelot X, Cauchemez S, Fraser C, Ferguson N. 2014 Bayesian reconstruction of disease outbreaks by combining epidemiologic and genomic data. *PLoS Comput. Biol*. **10**, e1003457. (doi:10.1371/journal.pcbi.1003457)
57. Ypma RJ, Bataille AM, Stegeman A, Koch G, Wallinga J, van Ballegooijen WM. 2012 Unravelling transmission trees of infectious diseases by combining genetic and epidemiological data. *Proc. R. Soc. B* **279**, 444–450. (doi:10.1098/rspb.2011.0913)
58. Ypma RJ, van Ballegooijen WM, Wallinga J. 2013 Relating phylogenetic trees to transmission trees of infectious disease outbreaks. *Genetics* **195**, 1055–1062. (doi:10.1534/genetics.113.154856)
59. Cauchemez S, Donnelly CA, Reed C, Ghani AC, Fraser C, Kent CK, Finelli L, Ferguson NM. 2009 Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States. *N. Engl. J. Med.* **361**, 2619–2627. (doi:10.1056/NEJMoa0905498)
60. Kilmarx PH *et al.* 2014 Ebola virus disease in health care workers–Sierra Leone, 2014. *MMWR Morb. Mortal Wkly Rep.* **63**, 1168–1171.
61. Matanock A *et al.* 2014 Ebola virus disease cases among health care workers not working in Ebola treatment units–Liberia, June–August, 2014. *MMWR Morb. Mortal Wkly Rep.* **63**, 1077–1081.
62. Carroll MW *et al.* 2015 Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. *Nature* **524**, 97–101. (doi:10.1038/nature14594)
63. Ladner JT *et al.* 2015 Evolution and spread of Ebola Virus in Liberia, 2014–2015. *Cell Host Microbe* **18**, 659–669. (doi:10.1016/j.chom.2015.11.008)
64. Balcan D, Goncalves B, Hu H, Ramasco JJ, Colizza V, Vespignani A. 2010 Modeling the spatial spread of infectious diseases: the GLobal Epidemic and Mobility computational model. *J. Comput. Sci.* **1**, 132–145. (doi:10.1016/j.jocs.2010.07.002)
65. Ferguson NM, Donnelly CA, Anderson RM. 2001 The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science* **292**, 1155–1160. (doi:10.1126/science.1061020)
66. Riley S. 2007 Large-scale spatial-transmission models of infectious disease. *Science* **316**, 1298–1301. (doi:10.1126/science.1134695)
67. Fine PE. 2003 The interval between successive cases of an infectious disease. *Am. J. Epidemiol.* **158**, 1039–1047. (doi:10.1093/aje/kwg251)
68. Flay BR *et al.* 2005 Standards of evidence: criteria for efficacy, effectiveness and dissemination. *Prevent. Sci.* **6**, 151–175. (doi:10.1007/s11121-005-5553-y)
69. Gartlehner G, Hansen RA, Nissman D, Lohr KN, Carey TS. 2006 Criteria for distinguishing effectiveness from efficacy trials in systematic reviews. Technical Review 12 (Prepared by the RTI-International-University of North Carolina Evidence-based Practice Center under Contract No. 290-02-0016.) AHRQ Publication No. 06-0046. Rockville, MD: Agency for Healthcare Research and Quality.
70. Weinberg GA, Szilagyi PG. 2010 Vaccine epidemiology: efficacy, effectiveness, and the translational research roadmap. *J. Infect. Dis.* **201**, 1607–1610. (doi:10.1086/652404)
71. Keeling MJ, Rohani P. 2008 *Modeling infectious diseases in humans and animals*. Princeton, NJ: Princeton University Press.
72. Mossong J *et al.* 2008 Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*. **5**, e74. (doi:10.1371/journal.pmed.0050074)
73. Hens N, Shkedy Z, Aerts M. 2012 *Modeling infectious disease parameters based on serological and social contact data: a modern statistical perspective* (*statistics for biology and health*), p. 316. New York, NY: Springer.
74. Cauchemez S, Ferguson NM, Wachtel C, Tegnell A, Saour G, Duncan B, Nicoll A. 2009 Closure of schools during an influenza pandemic. *Lancet Infect. Dis.* **9**, 473–481. (doi:10.1016/S1473-3099(09)70176-8)
75. Heymann DL. 2004 The international response to the outbreak of SARS in 2003. *Phil. Trans. R. Soc. Lond. B* **359**, 1127–1129. (doi:10.1098/rstb.2004.1484)
76. Eaton JW *et al.* 2014 Health benefits, costs, and cost-effectiveness of earlier eligibility for adult antiretroviral therapy and expanded treatment coverage: a combined analysis of 12 mathematical models. *Lancet Glob. Health* **2**, e23–e34. (doi:10.1016/S2214-109X(13)70172-4)
77. White MT, Conteh L, Cibulskis R, Ghani AC. 2011 Costs and cost-effectiveness of malaria control interventions—a systematic review. *Malar. J.* **10**, 337. (doi:10.1186/1475-2875-10-337)
78. Jiang L *et al.* 2015 Performance of case definitions for influenza surveillance. *Euro Surveill.* **20**, 21145. (doi:10.2807/1560-7917.ES2015.20.22.21145)
79. World Health Organization. 2014 Case definition recommendations for Ebola or Marburg virus diseases. See http://www.who.int/csr/resources/publications/ebola/ebola-case-definition-contact-en.pdf (accessed 23 May 2016).
80. Wolfel R *et al.* 2015 Mobile diagnostics in outbreak response, not only for Ebola: a blueprint for a modular and robust field laboratory. *Euro Surveill.* **20**, 30055. (doi:10.2807/1560-7917.ES.2015.20.44.30055)
81. Centers for Disease Control and Prevention. 2015 Mobilizing laboratories. See http://www.cdc.gov/about/ebola/mobilizing-laboratories.html (accessed 23 May 2016).
82. Broadhurst MJ *et al.* 2015 ReEBOV antigen rapid test kit for point-of-care and laboratory-based testing for Ebola virus disease: a field validation study. *Lancet* **386**, 867–874. (doi:10.1016/S0140-6736(15)61042-X)
83. Towers S, Patterson-Lomba O, Castillo-Chavez C. 2014 Temporal variations in the effective reproduction number of the 2014 West Africa Ebola outbreak. *PLoS Curr.* **6**. (doi:10.1371/currents.outbreaks.9e4c4294ec8ce1adad283172b16bc908)
84. Shaman J, Yang W, Kandula S. 2014 Inference and forecast of the current West African Ebola outbreak in Guinea, Sierra Leone and Liberia. *PLoS Curr.* **6**. (doi:10.1371/currents.outbreaks.3408774290b1a0f2dd7cae877c8b8ff6)
85. Scarpino SV *et al.* 2015 Epidemiological and viral genomic sequence analysis of the 2014 Ebola outbreak reveals clustered transmission. *Clin. Infect. Dis.* **60**, 1079–1082. (doi:10.1093/cid/ciu1131)
86. Dean NE, Halloran ME, Yang Y, Longini IM. 2016 Transmissibility and pathogenicity of Ebola virus: a systematic review and meta-analysis of household secondary attack rate and asymptomatic infection. *Clin. Infect. Dis.* **62**, 1277–1286. (doi:10.1093/cid/ciw114)
87. Fang LQ *et al.* 2016 Transmission dynamics of Ebola virus disease and intervention effectiveness in Sierra Leone. *Proc. Natl Acad. Sci. USA* **113**, 4488–4493. (doi:10.1073/pnas.1518587113)
88. Oak Ridge National Laboratory. 2016 Landscan™. See http://web.ornl.gov/sci/landscan/index.shtml (accessed 26 May 2016).
89. Worldpop. 2016 About worldpop. See http://www.worldpop.org.uk/about_our_work/about_worldpop/ (accessed 26 May 2016).
90. National Institute of Statistics Republic of Guinea. 2010 Population statistics of Guinea 2009. See https://knoema.com/GNPS2009/population-statistics-of-guinea-2009?tsId=1006150.
91. Statistics Sierra Leone. 2004 Population and Housing Census. 2006. See http://www.sierra-leone.org/Census/ssl_final_results.pdf.
92. Liberia Institute of Statistics and Geo-Information Services (LISGIS). 2009 Republic of Liberia, 2008 National Population and Housing Census Final Results. See http://www.emansion.gov.lr/doc/Population_by_County.pdf.
93. Colizza V, Barrat A, Barthelemy M, Vespignani A. 2006 The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl Acad. Sci. USA* **103**, 2015–2020. (doi:10.1073/pnas.0510525103)
94. Bogoch II *et al.* 2015 Assessment of the potential for international dissemination of Ebola virus via commercial air travel during the 2014 west African outbreak. *Lancet* **385**, 29–35. (doi:10.1016/S0140-6736(14)61828-6)
95. Gomes MF, Pastore YPA, Rossi L, Chao D, Longini I, Halloran ME, Vespignani A. 2014 Assessing the international spreading risk associated with the 2014 West African Ebola outbreak. *PLoS Curr.* **6**. (doi:10.1371/currents.outbreaks.cd818f63d40e24aef769dda7df9e0da5).
96. Poletto C *et al.* 2014 Assessing the impact of travel restrictions on international spread of the 2014 West African Ebola epidemic. *Euro Surveill.* **19**, 20936.
97. World Health Organization. 2014 Mali case, Ebola imported from Guinea. See http://www.who.int/mediacentre/news/ebola/10-november-2014-mali/en/ (accessed 2 November 2016).
98. Cauchemez S, Ledrans M, Poletto C, Quenel P, de Valk H, Colizza V, Boelle PY. 2014 Local and regional

Downloaded from http://rstb.royalsocietypublishing.org/ on June 6, 2017

spread of chikungunya fever in the Americas. *Euro Surveill.* **19**, 20854.

99. Grosche T, Rothlauf F, Heinzl A. 2007 Gravity models for airline passenger volume estimation. *J. Air Transport Manage.* **13**, 175–183. (doi:10.10.1016/j.jairtraman.2007.02.001)
100. Kramer AM, Pulliam JT, Alexander LW, Park AW, Rohani P, Drake JM. 2016 Spatial spread of the West Africa Ebola epidemic. *R. Soc. open sci.* **3**, 160294. (doi:10.1098/rsos.160294)
101. Riley S, Eames K, Isham V, Mollison D, Trapman P. 2015 Five challenges for spatial epidemic models. *Epidemics* **10**, 68–71. (doi:10.1016/j.epidem.2014.07.001)
102. Deville P, Linard C, Martin S, Gilbert M, Stevens FR, Gaughan AE, Blondel VD, Tatem AJ. 2014 Dynamic population mapping using mobile phone data. *Proc. Natl Acad. Sci. USA* **111**, 15 888–15 893. (doi:10.1073/pnas.1408439111)
103. Tatem AJ. 2014 Mapping population and pathogen movements. *Int. Health* **6**, 5–11. (doi:10.1093/inthealth/ihu006)
104. Wesolowski A, Qureshi T, Boni MF, Sundsøy PR, Johansson MA, Rasheed SB, Engø-Monsen K, Buckee CO. 2015 Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proc. Natl Acad. Sci. USA* **112**, 11 887–11 892.
105. Wesolowski A, Buckee CO, Bengtsson L, Wetter E, Lu X, Tatem AJ. 2014 Commentary: Containing the Ebola outbreak–the potential and challenge of mobile network data. *PLoS Curr. Outbreaks* **6**. (doi:10.1371/currents.outbreaks.0177e7fcf52217b8b634376e2f3efc5e)
106. Baguelin M, Flasche S, Camacho A, Demiris N, Miller E, Edmunds WJ. 2013 Assessing optimal target populations for influenza vaccination programmes: an evidence synthesis and modelling study. *PLoS Med.* **10**, e1001527. (doi:10.1371/journal.pmed.1001527)
107. Metcalf CJ, Farrar J, Cutts FT, Basta NE, Graham AL, Lessler J, Ferguson NM, Burke DS, Grenfell BT. 2016 Use of serological surveys to generate key insights into the changing global landscape of infectious disease. *Lancet* **388**, 728–730. (doi:10.1016/S0140-6736(16)30164-7)
108. Mangal TD, Aylward RB, Mwanza M, Gasasira A, Abanida E, Pate MA, Grassly NC. 2014 Key issues in the persistence of poliomyelitis in Nigeria: a case-control study. *Lancet Glob. Health* **2**, e90–e97. (doi:10.1016/S2214-109X(13)70168-2)
109. Dorigatti I, Cauchemez S, Ferguson NM. 2013 Increased transmissibility explains the third wave of infection by the 2009 H1N1 pandemic virus in England. *Proc. Natl Acad. Sci. USA* **110**, 13 422–13 427. (doi:10.1073/pnas.1303117110)
110. Schoepp RJ, Rossi CA, Khan SH, Goba A, Fair JN. 2014 Undiagnosed acute viral febrile illnesses, Sierra Leone. *Emerg. Infect. Dis.* **20**, 1176–1182. (doi:10.3201/eid2007.131265)
111. Woolhouse M. 2011 How to make predictions about future infectious disease risks. *Phil. Trans. R. Soc. B* **366**, 2045–2054. (doi:10.1098/rstb.2010.0387)
112. Matthews L, Woolhouse M. 2005 New approaches to quantifying the spread of infection. *Nat. Rev. Microbiol.* **3**, 529–536. (doi:10.1038/nrmicro1178)
113. Novel T. 2016 Polio Information System. See http://www.novel-t.ch/en/project/polio-information-system (accessed 26 May 2016).
114. healthsites.io. 2016 About healthsites.io. See https://healthsites.io/#about (accessed 3 October 2016).
115. The humanitarian data exchange. 2016 See https://data.humdata.org/ (accessed 26 May 2016).
116. Kucharski AJ, Camacho A, Flasche S, Glover RE, Edmunds WJ, Funk S. 2015 Measuring the impact of Ebola control measures in Sierra Leone. *Proc. Natl Acad. Sci. USA* **112**, 14 366–14 371. (doi:10.1073/pnas.1508814112)
117. Atkins KE *et al.* 2016 Retrospective analysis of the 2014–2015 Ebola epidemic in Liberia. *Am. J. Trop. Med. Hyg.* **94**, 833–839. (doi:10.4269/ajtmh.15-0328)
118. Ebola ca Suffit Ring Vaccination Trial C. 2015 The ring vaccination trial: a novel cluster randomised controlled trial design to evaluate vaccine efficacy and effectiveness during outbreaks, with special reference to Ebola. *Br. Med. J.* **351**, h3740. (doi:10.1136/bmj.h3740)
119. Henao-Restrepo AM *et al.* 2015 Efficacy and effectiveness of an rVSV-vectored vaccine expressing Ebola surface glycoprotein: interim results from the Guinea ring vaccination cluster-randomised trial. *Lancet* **386**, 857–866. (doi:10.1016/S0140-6736(15)61117-5)
120. Henao-Restrepo AM *et al*. 2016 Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola Ça Suffit!). *Lancet* **389**, 505–518. (doi:10.1016/S0140-6736(16)32621-6)
121. World Health Organization. 2016 An R&D blueprint or action to prevent epidemics plan of action. See http://www.who.int/csr/research-and-development/WHO-R_D-Final10.pdf (accessed 02 November 2016).
122. Gignoux E, Idowu R, Bawo L, Hurum L, Sprecher A, Bastard M, Porten K. 2015 Use of capture–recapture to estimate underreporting of Ebola virus disease, Montserrado County, Liberia. *Emerg. Infect. Dis.* **21**, 2265–2267. (doi:10.3201/eid2112.150756)
123. German RR, Lee LM, Horan JM, Milstein RL, Pertowski CA, Waller MN, Guidelines Working Group Centers for Disease Control & Prevention. 2001 Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. *MMWR Recomm. Rep.* **50**, 1–35; quiz CE31–37.
124. Rasmussen DA, Ratmann O, Koelle K. 2011 Inference for nonlinear epidemiological models using genealogies and time series. *PLoS Comput. Biol.* **7**, e1002136. (doi:10.1371/journal.pcbi.1002136)
125. Garske T *et al.* 2017 Heterogeneities in the case fatality ratio in the West African Ebola outbreak 2013–2016. *Phil. Trans. R. Soc. B* **372**, 20160308. (doi:10.1098/rstb.2016.0308)
126. MacIntyre CR *et al.* 2009 Face mask use and control of respiratory virus transmission in households. *Emerg. Infect. Dis.* **15**, 233. (doi:10.3201/eid1502.081166)
127. Ferng YH, Wong-McLoughlin J, Barrett A, Currie L, Larson E. 2011 Barriers to mask wearing for influenza-like illnesses among urban Hispanic households. *Public Health Nurs.* **28**, 13–23. (doi:10.1111/j.1525-1446.2010.00918.x)
128. Charania MR, Crepaz N, Guenther-Gray C, Henny K, Liau A, Willis LA, Lyles CM. 2011 Efficacy of structural-level condom distribution interventions: a meta-analysis of US and international studies, 1998–2007. *AIDS Behav.* **15**, 1283–1297. (doi:10.1007/s10461-010-9812-y)
129. Bonacic Marinovic A, Swaan C, van Steenbergen J, Kretzschmar M. 2015 Quantifying reporting timeliness to improve outbreak control. *Emerg. Infect. Dis.* **21**, 209–216. (doi:10.3201/eid2102.130504)
130. Haaskjold YL *et al.* 2016 Clinical features of and risk factors for fatal Ebola virus disease, Moyamba District, Sierra Leone, December 2014–February 2015. *Emerg. Infect. Dis.* **22**, 1537–1544. (doi:10.3201/eid2209.151621)
131. Wong JY *et al.* 2016 Assessment of the severity of Ebola virus disease in Sierra Leone in 2014–2015. *Epidemiol. Infect.* **144**, 1473–1481. (doi:10.1017/S0950268815003003)
132. Faye O *et al.* 2015 Use of viremia to evaluate the baseline case fatality ratio of Ebola virus disease and inform treatment studies: a retrospective cohort study. *PLoS Med.* **12**, e1001908. (doi:10.1371/journal.pmed.1001908)
133. The Epi Info viral hemorrhagic fever application. 2016 See https://epiinfovhf.codeplex.com/ (accessed 26 May 2016).
134. Schafer IJKnudsen EMcNamara LAAgnihotri SRollin PEIslam A 2016 The Epi Info Viral Hemorrhagic Fever (VHF) application: a resource for outbreak data management and contact tracing in the 2014–2016 West Africa Ebola epidemic. *J. Infect. Dis.* **214**(Suppl 3), S122–S136. (doi:10.1093/infdis/jiw272)
135. Blaya JA, Fraser HS, Holt B. 2010 E-health technologies show promise in developing countries. *Health Aff. (Millwood)* **29**, 244–251. (doi:10.1377/hlthaff.2009.0894)
136. Overhage JM, Grannis S, McDonald CJ. 2008 A comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions. *Am. J. Public Health* **98**, 344–350. (doi:10.2105/AJPH.2006.092700)
137. Thriemer K *et al.* 2012 Replacing paper data collection forms with electronic data entry in the field: findings from a study of community-acquired bloodstream infections in Pemba,

Downloaded from http://rstb.royalsocietypublishing.org/ on June 6, 2017

Zanzibar. *BMC Res. Notes* **5**, 113. (doi:10.1186/1756-0500-5-113)

138. Walther B, Hossin S, Townend J, Abernethy N, Parker D, Jeffries D. 2011 Comparison of electronic data capture (EDC) with the standard data capture method for clinical trial data. *PLoS ONE* **6**, e25348. (doi:10.1371/journal.pone.0025348)
139. EpiCollect. 2016 See http://www.epicollect.net/ (accessed 26 May 2016).
140. Xing W, Hejblum G, Valleron AJ. 2013 EpiBasket: how e-commerce tools can improve epidemiological preparedness. *Emerg. Health Threats J.* **6**, 19748. (doi:10.3402/ehtj.v6i0.19748)
141. World Health Organization. 2016 WHO report on global surveillance of epidemic-prone infectious diseases. 2000. See http://apps.who.int/iris/bitstream/10665/66485/1/WHO_CDS_CSR_ISR_2000.1.pdf (accessed 26 May 2016).
142. Chan M. 2016 Address to the Sixty-ninth World Health Assembly. See http://www.who.int/dg/speeches/2016/wha-69/en/ (accessed 26 May 2016).
143. Huang JY, Louis FJ, Dixon MG, Sefu M, Kightlinger L, Martel LD, Jayaraman GC, Gueye AS. 2016 Notes from the field: baseline assessment of the use of Ebola rapid diagnostic tests—Forecariah, Guinea, October–November 2015. *MMWR Morb. Mortal Wkly Rep.* **65**, 328–329. (doi:10.15585/mmwr.mm6512a4)
144. World Health Organization. 2016 Essential medicines and health products: diagnostics. See http://www.who.int/medicines/ebola-treatment/emp_ebola_diagnostics/en/ (accessed 8 June 2016).
145. Quick J *et al.* 2016 Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232. (doi:10.1038/nature16996)
146. Aarestrup FM, Koopmans MG. 2016 Sharing data for global infectious disease surveillance and outbreak detection. *Trends Microbiol.* **24**, 241–245. (doi:10.1016/j.tim.2016.01.009)
147. Yozwiak NL, Schaffner SF, Sabeti PC. 2015 Data sharing: make outbreak research open access. *Nature* **518**, 477–479. (doi:10.1038/518477a)
148. Whitty CJ, Mundel T, Farrar J, Heymann DL, Davies SC, Walport MJ. 2015 Providing incentives to share data early in health emergencies: the role of journal editors. *Lancet* **386**, 1797–1798. (idoi:10.1016/S0140-6736(15)00758-8)
149. HIV Modelling Consortium. 2016 See http://www.hivmodelling.org/ (accessed 26 May 2016).
150. Lee BY, Haidari LA, Lee MS. 2009 Mathematical modelling of the pandemic H1N1 2009. *Wkly Epidemiol. Rec.* **84**, 341–348.
151. Van Kerkhove MD *et al.* 2010 Studies needed to address public health challenges of the 2009 H1N1 influenza pandemic: insights from modeling. *PLoS Med.* **7**, e1000275. (doi:10.1371/journal.pmed.1000275)
152. Penny MA *et al.* 2016 Public health impact and cost-effectiveness of the RTS,S/AS01 malaria vaccine: a systematic comparison of predictions from four mathematical models. *Lancet* **387**, 367–375. (doi:10.1016/S0140-6736(15)00725-4)
153. Modjarrad K, Moorthy VS, Millett P, Gsell PS, Roth C, Kieny MP. 2016 Developing global norms for sharing data and results during public health emergencies. *PLoS Med.* **13**, e1001935. (doi:10.1371/journal.pmed.1001935)
154. PLOS Currents. See http://currents.plos.org/ (accessed 26 May 2016).
155. *The Lancet*. See http://www.thelancet.com/lancet/information-for-authors/fast-track (accessed 26 May 2016).
156. Cold Spring Harbor Laboratory. See http://biorxiv.org/ (accessed 26 May 2016).
157. Cornell University Library. See https://arxiv.org/ (accessed 26 May 2016).
158. Jombart T *et al.* 2014 OutbreakTools: a new platform for disease outbreak analysis using the R software. *Epidemics* **7**, 28–34. (doi:10.1016/j.epidem.2014.04.003)
159. King AA, Domenech de Celles M, Magpantay FM, Rohani P. 2015 Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola. *Proc. R. Soc. B* **282**, 20150347. (doi:10.1098/rspb.2015.0347)
160. Lessler J, Edmunds WJ, Halloran ME, Hollingsworth TD, Lloyd AL. 2015 Seven challenges for model-driven data collection in experimental and observational studies. *Epidemics* **10**, 78–82. (doi:10.1016/j.epidem.2014.12.002)