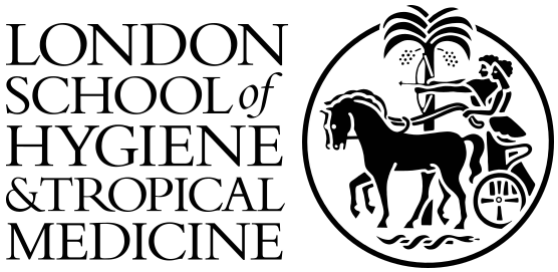

Epidemics 10 (2015) 78–82

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Seven challenges for model-driven data collection in experimental and observational studies

J. Lessler[a,*], W.J. Edmunds[b], M.E. Halloran[c,d], T.D. Hollingsworth[e,f,g], A.L. Lloyd[h,i]

[a] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21224, USA
[b] London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK
[c] Department of Biostatistics, School of Public Health, University of Washington, Seattle, WA 98195 USA
[d] Center for Statistics and Quantitative Infectious Diseases, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
[e] Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK
[f] School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK
[g] Liverpool School of Tropical Medicine, Liverpool L3 5QA, UK
[h] Department of Mathematics and Biomathematics Graduate Program, North Carolina State University, Raleigh, NC 27695, USA
[i] Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA

ARTICLE INFO

Article history:
Received 1 March 2014
Received in revised form 5 December 2014
Accepted 8 December 2014
Available online 16 December 2014

Keywords:
Modeling
Data collection
Observational studies
Experimental studies

ABSTRACT

Infectious disease models are both concise statements of hypotheses and powerful techniques for creating tools from hypotheses and theories. As such, they have tremendous potential for guiding data collection in experimental and observational studies, leading to more efficient testing of hypotheses and more robust study designs. In numerous instances, infectious disease models have played a key role in informing data collection, including the Garki project studying malaria, the response to the 2009 pandemic of H1N1 influenza in the United Kingdom and studies of T-cell immunodynamics in mammals. However, such synergies remain the exception rather than the rule; and a close marriage of dynamic modeling and empirical data collection is far from the norm in infectious disease research. Overcoming the challenges to using models to inform data collection has the potential to accelerate innovation and to improve practice in how we deal with infectious disease threats.

Introduction: What is the role of models in data collection?

When people refer to "models" of infectious disease transmission, they usually mean something far more specific than the word "model" indicates. The term is generally used to refer to a system of equations or computer program that explicitly represents the mechanisms of disease transmission and pathogenesis. This is in contrast to models purely of statistical association that are common throughout the medical literature. By setting forth a mechanistic hypothesis in a mathematically precise form, models become tools for generating (perhaps unexpected) predictions which can be used to test the underlying hypotheses through confrontation with data. Though this use of models has a long tradition throughout all branches of science (including infectious disease epidemiology and ecology), it often takes a back seat to other uses of infectious disease models. The highest profile infectious disease modeling work is often aimed at making predictions or filling in gaps in existing data. These uses of models by definition *presume that the hypothesis captured in the model is close enough to the truth to capture the dynamics of the system relevant to the task at hand*, and provide answers only conditional on the correctness of the model. While many researchers put great effort into fitting both the structure and parameters of models using existing data, data are rarely collected with the explicit purpose of testing model hypotheses, and many models go unchallenged and untested after they are first presented.

Models are powerful, in part, because they can turn a hypothesis or theory into a tool for making precise predictions. Yet even in this capacity infectious disease models are underutilized in the data collection process. For example, sample size calculations and power analyses are de rigueur for the design of observational studies and clinical trials. However, cases of transmissible infections are non-independent, limiting the utility of standard theory. Using mechanistic models that account for the transmission process can allow robust estimation in the setting of these "dependent happenings" and tell us not only how much data to collect but when to collect it (Halloran et al., 2010).

* Corresponding author. Tel.: +1 410 955 3551.
E-mail address: justin@jhu.edu (J. Lessler).

http://dx.doi.org/10.1016/j.epidem.2014.12.002

Model-driven data collection does occur, and a prime example is the Garki project (Molineaux et al., 1980). This seminal project studied the transmission of malaria under different control interventions and captured longitudinal data on both disease and vector dynamics. Thirty years after its publication it remains one of the most valuable datasets for parameterizing malaria models (e.g., Griffin et al., 2010), and has been cited over 650 times. In part because input from models was used in the design of the study, it provides an almost unique dataset on the dynamics of malaria across a number of seasons and under different interventions, providing an example for future studies which has rarely been emulated.

Through sensitivity analysis and quantification of uncertainty, models can be used to elucidate the parameters and processes that contribute most to our inability to make predictions with a high degree of confidence, and those which are less important. They can thereby indicate where experimental effort would be best directed to improve the predictive power of a model. An example of where this is being attempted is HPTN 071 (PopART), a community randomized trial of combination prevention packages including early antiretroviral therapy initiation to control HIV transmission in Africa. In the analysis of this trial, models are being used to estimate endpoints (e.g., community incidence), improve study design, and identify which processes are the main drivers of uncertainty (Cori et al., 2014).

For the modeling of infectious disease to reach its full potential, it must become more tightly integrated with the data collection process. Despite successes such as those described above, significant challenges to successful model-driven data collection remain. Some stem from technical or cultural issues that are, in principle, easily surmountable. However, others are inherent in the role that models are often called on to play in public health response and scientific research, hence may be difficult or impossible to overcome.

In this paper, we start by discussing some fundamental challenges in the relationship between models and data collection, and end with specific challenges in current practice. Throughout we highlight the potential role of "modelers" in the data collection process. Though infectious disease modeling has grown into a distinct specialty, here we refer to anyone who defines and implements a mechanistic model of disease transmission as a "modeler". In addition to our headline challenges, we have tried to identify specific research avenues under each (mentioned at the end of key paragraphs) that could result in significant progress in confronting the challenge.

1. Ensuring a strong empirical basis for models used to fill gaps in data and knowledge

One primary role of models is to use our mechanistic understanding of a system to fill gaps in available data. Gaps in knowledge may occur because data is difficult or expensive to collect, because we are trying to understand past events which can no longer be directly observed, or because we are confronting a novel disease which is not yet well understood. In each of these contexts models play a hugely important role because of the lack of information, but are at the same time hamstrung by the lack of data with which to test model assumptions or fit parameters. While data will always be scarce in these situations, each presents an opportunity to use mechanistic models to make more effective use of existing data and guide ongoing data collection.

Data on disease incidence is difficult and expensive to collect on a broad scale; and passive clinical surveillance may include only the most severe cases or be clouded by a non-specific clinical profile. However, the burden of disease is one of the most fundamental pieces of epidemiologic information used in setting the public health agenda. Hence, mechanistic models are often used to fill the gap. For instance, transmission models have been used to help estimate the global burden of measles (Simons et al., 2012), using our knowledge of how susceptibility drives epidemic dynamics to infer the true number of cases from what was observed. Likewise, models have been used to help translate observed cases of acute flaccid paralysis to polio incidence through our understanding of transmission and the symptomatic attack rate (Eichner and Dietz, 1996). There are opportunities to test and improve models that fill data gaps. Models can be used to identify efficient data collection activities that validate the model but do not require the effort and expense of collecting the data the model is meant to infer. Innovative methods and systems for updating both the model predictions and the model assumptions in real time as new surveillance data becomes available could greatly increase their value in public health practice.

Understanding disease dynamics is often dependent on measuring disease incidence years or decades in the past, making the design of suitable data collection particularly challenging. Past incidence rates may be unmeasured because a disease circulated before it was identified (e.g., HIV before the 1980s), because acute infection often occurs without identifiable symptoms (e.g., dengue, HIV), or because of poor surveillance. In the latter two cases, even recent incidence patterns may be unknown. Dynamic models can often be used to infer past incidence with current cross sectional data or historic samples. Age specific serologies can be used to estimate the past force of infection, and have been used to measure historic patterns of the force of infection for dengue and other diseases (Rodríguez-Barraquer et al., 2013). Phylogenetic models can be paired with simple epidemic models to infer past epidemic dynamics, as has been done with HIV and hepatitis C (Stadler et al., 2012). These techniques, particularly phylogenetic inference, have become quite popular, but validation has been largely limited to simulation studies (e.g., Robinson et al., 2013; Volz et al., 2012). Studies aimed at collecting prospective data specifically to evaluate serologic and phylogenetic approaches to inferring incidence would help to place these inferences on firmer footing.

When responding to emerging epidemics, the problem of missing data is particularly acute. Here we are forced to forecast the course of an epidemic with limited knowledge of the pathogen and burden of disease. Critical data must be collected to carry out this task, some of which can be measured most effectively early on in the process of disease emergence. However, this data is not routinely collected early on, whether due to the difficulty of collection, competing priorities or its value being unrecognized. In particular, the tendency is to focus almost exclusively on cases early in an epidemic, whereas those who were at risk but did not become infected may carry the most information in terms of population susceptibility and disease transmissibility. A notable exception is the 2009 H1N1 influenza pandemic in the United Kingdom, where data collection was guided by the long-term involvement of modelers in the design of control programs. Although not all the data requested was collected, these efforts enabled policy-relevant modeling during the early stages of the epidemic (Ghani et al., 2010; Baguelin et al., 2010; Eames et al., 2012). This experience illustrates how integration with the public community can pay off in better inferences to support policy. At the time of writing, modeling is playing an important role in the response to the Ebola outbreak in West Africa, making use of the detailed contact tracing data collected as part of the response (though for purposes other than modeling) (WHO Ebola Response Team, 2014). Researchers should challenge themselves to identify the most useful classes of models in an emerging epidemic and the data needed to parameterize them. They should then work with public health officials to integrate collecting this data into epidemic response plans before such a response is needed.

2. Design studies that account for the dynamic nature of disease systems

Mathematical and computational models are most useful in understanding disease systems that are dynamic, often containing structural relationships and feedback loops that lead to behaviors that are unexpected or do not follow simple functional relationships. Furthermore, these dynamics are often driven by stochastic processes which are often hard to predict. These factors present particular challenges for data collection.

While models are extremely useful in understanding disease systems, they cannot capture every factor that plays a role in driving disease incidence. Stochasticity and factors outside the modeled system may play a primary role in determining whether epidemics occur and how they develop once underway. Typical examples include disease introductions, public response to real or perceived disease threats, and pathogen evolution. It is not always desirable, or even possible, to model such exogenous factors. However, when designing studies (or even using existing data) aimed at testing models, parameterizing models, or assessing the impact of a control measure, care must be taken in separating the results of factors lying outside the system from those captured by the model.

The dynamic nature of the disease processes leads to many practical issues in study design. Accurate sample size calculations must take into account disease dynamics, and the resulting non-independence of events. Outside of adjustments for correlation within clusters (i.e., calculating design effects based on intra-class correlation coefficients (Kerry and Bland, 1998)), such calculations are rarely done and may require complex simulations. The study itself may have an effect on the process being studied, and this effect may be magnified by dynamic feedback loops in disease systems. Even "observational" studies may affect disease dynamics, as we may be ethically required to treat the cases of the disease we detect, and this treatment will affect future incidence (Valle and Clark, 2013). Finally, interpretation of our observations is contingent on the current state of the system. It is often assumed the dynamic system is in steady state, but this is rarely the case, and the phenomena we are most interested in often involve perturbations to the system that will guarantee that it is not. For instance, the implications of a particular number of individuals being infected with a disease for broader disease dynamics depend on whether data was collected when disease incidence was growing, receding or in steady state (i.e., the current dynamic regime). Since it is usually impossible to measure the full state of the system, successful model driven data collection must not only measure state variables (e.g., the number susceptible or infectious), but also attempt to determine the dynamic regime in which those variables were collected.

Simulation of trial design is a growing area of research, with numerous applications to vaccine trials (e.g., Van de Velde et al., 2007; Yang et al., 2006) and growing use in other settings (e.g., PopART Cori et al., 2014). The development of standard tools similar to those available for standard sample size calculations, or even a list of best practices, would go a long way to expanding the use of mechanistic models in study design.

3. Identifying critical data collection activities in light of high dimensional parameter spaces and substantial structural uncertainty

Infectious disease systems are often the result of a large number of interacting social and biological processes; hence they can only be fully characterized by complex models. However, model complexity carries a price, since flexibility makes models difficult to fit to data or test in a meaningful manner. Simplification may be a path to creating models that can be fit to data, but may come at the cost of failing to capture essential system dynamics. High dimensional models can fit more complex trajectories, but may have multiple parameter sets that provide almost indistinguishable fits and can be almost impossible to test. For models to drive data collection, modelers face the challenge of identifying the specific parameters or predictions of their model that are most essential in increasing our understanding of the system and testing the structure of the model itself. Fortunately, mathematical and statistical techniques exist for identifying what components have the most influence on the predictions of complex systems (e.g., Saltelli et al., 1999; Caswell, 2007). A fruitful avenue may be to expand these techniques to more directly estimate the value of critical experiments or studies in reducing our uncertainty about critical outcomes.

Even when the critical study to evaluate or improve a model has been designed, that study may be difficult to perform (e.g., outcomes may be hard to measure; sample size requirements may be unreasonably large). When specifying a model reveals these difficulties, the model is serving one of its most important roles in service of data collection: motivating precision in the statement of assumptions and hypotheses. For instance, the hypothesis that a vaccine is "85% effective" at first seems simple to test in an observational study, but once we attempt to put this into a model the subtleties of this statement become clear: since vaccination can change the course of an epidemic the comparison population must be carefully chosen depending on whether we seek to measure only direct protection or both direct and indirect effects; and our interpretation of results may differ markedly if we believe vaccination prevents infection completely (as with oral polio vaccine) or only stops the development of disease (as with injectable polio vaccine).

4. Collecting data to test and parameterize models that bridge spatial scales

Models of interventions also often rely on assumptions about how interventions work at an individual level. For example, modeling a vaccination intervention involves assumptions about whether protection is homogeneous or heterogeneous in the population, how it affects susceptibility to infection compared to the probability of developing disease given infection, and how it reduces onward transmission. Each of these differences at the individual level has important implications for how disease spreads within populations.

Recent developments in individual- or agent-based simulations have helped bridge our understanding of the dynamics of infectious disease transmission at different scales, from the individual to households, schools, towns, cities and countries (Ferguson, 2005; Eubank et al., 2004). These models have also served as powerful tools for communicating the insights from models to policy makers. The best examples in this area are based on well validated submodels fit to data on population distributions, human movement, epidemiological parameters, and within host processes. However, they are rarely tested against epidemic data at the same spatial scale (exceptions include Cauchemez et al., 2008). For these approaches to reach their full potential they require further validation, but how to do that? Collecting data at the spatial and temporal detail of the predictions created by these models is unrealistic, and the situations modeled may be so unique as to make it impossible to identify a comparison set. Evaluation of marginal results (i.e., sets of statistics summarizing more detailed model output) is the clear pathway to validating these models, but work is needed to identify which set of marginal results need to be tested for a model to be considered adequately validated for use in forecasting or inference. Innovative study designs may help to test key model predictions (e.g., the Fluscape study, Read et al., 2014), and there may be benefit in designing

models and validating data collection activities in concert to make validation feasible.

5. Collecting data over time scales commensurate with model predictions

The predictions from mathematical models often play out over many years or even decades. Such predictions as increases in herpes zoster following the introduction of varicella vaccination (Brisson et al., 2002) or reductions in cervical cancer incidence following HPV vaccination (Choi et al., 2012) take years to manifest. Likewise, proper parameterization and evaluation of dynamic models often requires decades of data, even for diseases with relatively well observed outcomes and simple dynamics (e.g., measles, Bjørnstad et al., 2002). Hence, data collection activities to test model hypotheses or improve model results may need to take place on timescales not in line with the usual process of scientific research or funding. Focusing on studies that are inexpensive but sustainable may help to identify incremental changes to normal surveillance procedures and sustainable data collection activities that will yield results over the course of decades. Starting and sustaining such long term research projects can present a substantial challenge, as they likely must be sustained outside of normal research funding mechanisms (which typically fund in increments of 5 years at best). Investigators may have to content themselves with helping future scientists, and public health and scientific priorities may have changed by the time results come in. However, long running cohort studies such as Multicenter AIDS Cohort Study (Kaslow et al., 1987), Women's Health Initiative (Rossouw et al., 2002), the Rakai Health Sciences Program (Wawer et al., 1999), and the Framingham Heart Study (Splansky et al., 2007) have shown that long term data collection activities can yield important, and often unexpected, results. If the value of these data can be demonstrated to the public health agencies, they can possibly be included in routine surveillance for many decades to come. Methods that both make incremental use of data to improve models and provide results as data is collected may help to ensure the sustained effort needed to test long term hypotheses.

6. Integrating infectious disease dynamics and modeling into the every-day scientific process

Modeling and the study of infectious disease dynamics are sometimes treated as lying outside the normal scientific process (particularly by those outside of the field), being somehow qualitatively different than laboratory science or other types of epidemiologic studies. Modelers themselves may be in part responsible for this perception due to the frequent use of a "parameters from the literature" approach to building models. However, models should be an integral part in the ongoing cycle of hypothesis generation, data collection, and hypothesis refinement (Restif et al., 2012). A framework for achieving this goal is to build collaborations that closely integrate modeling, field studies and laboratory experiments.

Experimental studies are an area where relationships between modelers and experimenters have generated questions which can be addressed directly. For instance, the interplay of mathematical models and experimental data in animals has contributed to our understanding of the generation and maintenance of CD8+ T-cell memory (Antia et al., 2005). As an example: a particular model of homeostasis suggested the total number of CD8+ T-cells remains stable regardless of exposure to additional pathogens. A study inducing new long-lived CD8+ T cells in mice was performed to test this hypothesis. Unexpectedly, the total number of memory CD8+ T-cells doubled to accommodate the new cells (Vezys et al., 2009). Thus, a new theory that the number of memory cells in mammals adapts according to immunological experience was developed by an iterative discussion between modelers and experimenters, identifying what data needed to be collected to test a clearly defined hypothesis.

While numerous instances of such collaborations exist, the perception of the scientific importance of infectious disease dynamics is often disconnected from its true importance in disease systems. Vaccines are an excellent example: licensure is based only on individual level effects (though indirect effects may be estimated); while public health decisions are heavily influenced by the indirect effects from vaccination (e.g., herd immunity) and dynamic components. Integrating cross-disciplinary experiences into the training of field, laboratory, mathematical and computational scientists could help make all more effective and ensure a role for dynamic models in the scientific process.

7. Making it common practice to identify data that would test model hypotheses

The presentation of models and model results should provide a golden opportunity for identifying those data or studies that could test the hypotheses underlying the model, reproduce results or show generalizability. However, few take this opportunity, and there is a tendency for models to be presented and then forgotten with little or no reflection on the quality of the results as more information becomes available. For models to play an important role in the progress of science, it is critical to test their qualitative and quantitative predictions and to refine the underlying models over time. Perhaps standard practice in the field should be to have a "what studies need to be done" section at the end of each paper presenting a model or modeling results. As "The Journal on Infectious Disease Dynamics", *Epidemics* is well situated to promote such a practice by encouraging or requiring such a section for research manuscripts.

Conclusion

These are broad challenges to strengthen the relationship between models and data collection across experimental, epidemiological and clinical studies of infectious diseases. The breadth of these challenges provides many opportunities to develop new techniques and standards by applying these concepts to studies in all areas of infectious disease research. The core principle behind many of the challenges discussed is that mechanistic models have a role to play early in the process of experimental and epidemiological study design, and not only be used for secondary data analysis. The challenge for the modeler is to identify priorities for what pieces of information are needed and how to collect them within the confines of an experimental or epidemiological study. This discussion will help clarify both the model and study goals, potentially leading to new "Garki projects" that will drive discovery and innovation among infectious disease researchers across disciplines.

Acknowledgments

This work was initiated during the 2013 Infectious Disease Dynamics meeting at the Isaac Newton Institute, Cambridge, UK. All authors warmly acknowledge the support of the Newton Institute. JL is a recipient of a K22 award from the National Institute of Allergy and Infectious Disease (K22-AI092150-01). MEH was supported by National Institutes of Health Grants Nos. U54-GM111274 and R37-AI032042. TDH is a member of the Center for Applied Health Research & Delivery (http://www.lstmliverpool.ac.uk/research/cross-cutting-themes/cahrd/) which is supported by a Wellcome Trust Institutional Strategic Support Award (No. 097830/Z/11/A-C)

to LSTM. ALL is supported by the Research and Policy for Infectious Disease Dynamics (RAPIDD) program of the Science and Technology Directory, Department of Homeland Security, and Fogarty International Center, National Institutes of Health, and by grants from the National Institutes of Health (P01-AI098670, R01-AI091980) and the National Science Foundation (RTG/DMS-1246991).

References

Antia, R., Ganusov, V., Ahmed, V.R., 2005. The role of models in understanding CD8+ T-cell memory. Nat. Rev. Immunol. 5, 101–111.
Baguelin, M., et al., 2010. Vaccination against pandemic influenza A/H1N1v in England: a real-time economic evaluation. Vaccine 28, 2370–2384.
Bjørnstad, O.N., Finkenstädt, B.F., Grenfell, B.T., 2002. Dynamics of measles epidemics: estimating scaling of transmission rates using a time series sir model. Ecol. Monogr. 72, 169–184.
Brisson, M., Gay, N.J., Edmunds, W.J., Andrews, N.J., 2002. Exposure to varicella boosts immunity to herpes-zoster: implications for mass vaccination against chickenpox. Vaccine 20, 2500–2507.
Caswell, H., 2007. Sensitivity analysis of transient population dynamics. Ecol. Lett. 10, 1–15.
Cauchemez, S., Valleron, A.-J., Boëlle, P.-Y., Flahault, A., Ferguson, N.M., 2008. Estimating the impact of school closure on influenza transmission from Sentinel data. Nature 452, 750–754.
Choi, Y.H., Chapman, R., Gay, N., Jit, M., 2012. Potential overestimation of HPV vaccine impact due to unmasking of non-vaccine types: quantification using a multitype mathematical model. Vaccine 30, 3383–3388.
Cori, A., et al., 2014. HPTN 071 (PopART): A cluster-randomized trial of the population impact of an HIV combination prevention intervention including universal testing and treatment: mathematical model. PLoS ONE 9, e84511.
Eames, K.T.D., Tilston, N.L., Brooks-Pollock, E., Edmunds, W.J., 2012. Measured dynamic social contact patterns explain the spread of H1N1v influenza. PLoS Comput. Biol. 8, e1002425.
Eichner, M., Dietz, K., 1996. Eradication of poliomyelitis: when can one be sure that polio virus transmission has been terminated? Am. J. Epidemiol. 143, 816–822.
Eubank, S., et al., 2004. Modelling disease outbreaks in realistic urban social networks. Nature 429, 180–184.
Ferguson, N.M., et al., 2005. Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature 437, 209–214.
Ghani, A., et al., 2010. The early transmission dynamics of H1N1pdm Influenza in the United Kingdom. PLoS Curr. 1–15, http://dx.doi.org/10.1371/currents.RRN1130.
Griffin, J.T., et al., 2010. Reducing *Plasmodium falciparum* malaria transmission in Africa: a model-based evaluation of intervention strategies. PLoS Med. 7.
Halloran, M.E., Longini Jr., I., Struchiner, C., 2010. Design and Analysis of Vaccine Studies. Springer, New York, pp. 271–312.
Kaslow, R.A., et al., 1987. The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. Am. J. Epidemiol. 126, 310–318.
Kerry, S., Bland, J., 1998. Statistics notes: the intracluster correlation coefficient in cluster randomisation. BMJ 316, 1455–1460.
Molineaux, L., Gramiccia, G., Organization, W.H., 1980. In: Molineaux, L., Gramiccia, G. (Eds.), The Garki Project: Research on the Epidemiology and Control of Malaria in the Sudan Savanna of West Africa, vol. 311. World Health Organization, Geneva.
Read, J.M., et al., 2014. Social mixing patterns in rural and urban areas of southern China. Proc. Biol. Sci. 281, 20140268.
Restif, O., et al., 2012. Model-guided fieldwork: practical guidelines for multidisciplinary research on wildlife ecological and epidemiological dynamics. Ecol. Lett. 15, 1083–1094.
Robinson, K., Fyson, N., Cohen, T., Fraser, C., Colijn, C., 2013. How the dynamics and structure of sexual contact networks shape pathogen phylogenies. PLoS Comput. Biol. 9, e1003105.
Rodríguez-Barraquer, I., et al., 2013. Revisiting Rayong: shifting seroprofiles of dengue in Thailand and their implications for transmission and control. Am. J. Epidemiol. 1–8, http://dx.doi.org/10.1093/aje/kwt256.
Rossouw, J.E., et al., 2002. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. JAMA 288, 321–333.
Saltelli, A., Tarantola, S., Chan, K.P.-S., 1999. A quantitative model-independent method for global sensitivity analysis of model output. Technometrics 41, 39–56.
Simons, E., et al., 2012. Assessment of the 2010 global measles mortality reduction goal: results from a model of surveillance data. Lancet 379, 2173–2178.
Splansky, G.L., et al., 2007. The Third Generation Cohort of the National Heart. Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. Am. J. Epidemiol. 165, 1328–1335.
Stadler, T., Kühnert, D., Bonhoeffer, S., Drummond, A.J., 2012. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proc. Natl. Acad. Sci. U. S. A. 110 (1), 228–233.
Valle, D., Clark, J., 2013. Improving the modeling of disease data from the government surveillance system: a case study on malaria in the Brazilian Amazon. PLoS Comput. Biol. 9, e1003312.
Van de Velde, N., Brisson, M., Boily, M.-C., 2007. Modeling human papillomavirus vaccine effectiveness: quantifying the impact of parameter uncertainty. Am. J. Epidemiol. 165, 762–775.
Vezys, V., et al., 2009. Memory CD8 T-cell compartment grows in size with immunological experience. Nature 457, 196–199.
Volz, E.M., Koopman, J.S., Ward, M.J., Brown, A.L., Frost, S.D.W., 2012. Simple epidemiological dynamics explain phylogenetic clustering of HIV from patients with recent infection. PLoS Comput. Biol. 8, e1002552.
Wawer, M.J., et al., 1999. Control of sexually transmitted diseases for AIDS prevention in Uganda: a randomised community trial. Rakai Project Study Group. Lancet 353, 525–535.
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., http://dx.doi.org/10.1056/NEJMoa1411100.
Yang, Y., Longini, I.M., Halloran, M.E., 2006. Design and evaluation of prophylactic interventions using infectious disease incidence data from close contact groups. J. R. Stat. Soc. C: Appl. Stat. 55, 317–330.