Epidemiological Data and Model Requirements to Support Policy

Marc Baguelin
Public Health England
London, UK
marc.baguelin@phe.gov.uk

Elizabeth Buckingham-Jeffery
School of Mathematics
Manchester, UK
e.buckingham-jeffery@manchester.ac.uk

Ian Hall
School of Mathematics
Manchester, UK
ian.hall@manchester.ac.uk

Thomas House
School of Mathematics
Manchester, UK
thomas.house@manchester.ac.uk

Timothy Kinyanjui
School of Mathematics
Manchester, UK
timothymuiruri.kinyanjui@manchester.ac.uk

Lorenzo Pellis
School of Mathematics
Manchester, UK
lorenzo.pellis@manchester.ac.uk

ABSTRACT

Often, the task of epidemiological modelling is seen as one of improving biological and social realism, with increasing data availability enabling increased realism. Here, we consider ways in which models designed to support policy have different data and algorithmic requirements from those aiming at realism or insight, via a series of case studies. In particular, calculation and communication of uncertainty is often more important than refinement of model structure without confirmation of validity.

CCS CONCEPTS

• **Applied computing** → **Life and medical sciences**; • **Computing methodologies** → *Artificial intelligence*; *Modeling and simulation*;

KEYWORDS

Epidemic; Disease Dynamics; Uncertainty Quantification; Inference

ACM Reference Format:
Marc Baguelin, Elizabeth Buckingham-Jeffery, Ian Hall, Thomas House, Timothy Kinyanjui, and Lorenzo Pellis. 2018. Epidemiological Data and Model Requirements to Support Policy. In *Proceedings of epiDAMIK, ACM SIGKDD (epiDAMIK)*. ACM, New York, NY, USA, Article tbc, 5 pages. https://doi.org/tbc

1 OVERVIEW

Epidemiology is a data-driven science; however, the sources of data involved are typically observational, since controlled experiments are seldom practical. This means mathematical modelling has a key role to play in the field, both in terms of inference and of prediction [17].

Inference is challenging because data is usually scarse and only indirectly informing our knowledge, as most events involved in the transmission process (e.g. infection, or beginning/end of viral shedding, which does not necessarily correspond to onset/end of symptoms) are rarely observed. Models offer an opportunity to codify our biological understanding (or belief) about the infection mechanism, but their integration with data requires sophisticated statistical techniques, particularly those that successfully cope with missing data.

Models are also key for predictions, as they allow for the representation of phenomena such as potential interventions that are not directly observable. Unfortunately, validation of model predictions is problematic, as no epidemic is ever identical to any other and testing alternative control policies is anyway constrained by ethical or political considerations.

When modelling to inform policy, particularly when a decision must be made under time pressure, there are particular practical and theoretical challenges. These are often markedly different from those that arise in curiosity-driven science. Figure 1 shows an overall picture of modelling for policy support. A successful policy-driven model will integrate available data, and propagate forward the most significant uncertainties to allow interventions to be optimised, all the while improving methodology. The rest of this position paper is structured as a set of case studies that illustrate this point.

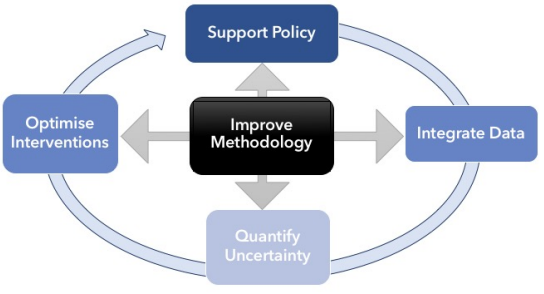

Figure 1: Representation of an idealised modelling support cycle for policy.

epiDAMIK, August 2018, London, UK

ACM ISBN tbc.
https://doi.org/tbc

2 TYPES OF MODELS USED

Broadly speaking, there are two classes of models (although these can be related to each other and embedded in more general frameworks) that we now present simple examples of for clarity.

The first of these is the Poisson process for a non-communicable disease. Here we imagine that we observe a number of cases of disease y in a population of size N over a time period of length t. In the simplest model, we assume that $N \gg y$ and that cases arise independently at a rate λ, leading to a Poisson likelihood of

$$\Pr(y) = \frac{(\lambda t)^y \mathrm{e}^{-\lambda t}}{y!} \,. \tag{1}$$

This Poisson model, if correct, would allow us to estimate λ from data (for example if 365 cases are observed in a year then we would estimate the rate to be $\hat{\lambda} = 1\ \mathrm{days}^{-1}$) and on that basis could use (1) to predict the probability that we observe significantly more cases than expected, leading to a lack of healthcare capacity, on a given week.

The second is the SIR model, in which individuals are split into compartments according to whether they are Susceptible to the disease, Infectious, or Removed. This model is often represented through the non-linear system of differential equations:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\frac{\beta}{N} SI \,; \qquad \frac{\mathrm{d}I}{\mathrm{d}t} = \frac{\beta}{N} SI - \gamma I \,. \tag{2}$$

If we know the values $S(t = 0)$ and $I(t = 0)$, as well as the mean duration of infectiousness $1/\gamma$ and the basic reproductive ratio $R_0 = \beta/\gamma$ then this model (and related approaches [23]) can be used to calculate quantities such as the value of t at which $I(t)$ is maximised – i.e. the peak time and height – as well as the total number of individuals infected during the epidemic, $N - S(\infty)$.

Elaborating on the back-bone of the SIR model (2), a range of more sophisticated model structures has been proposed to relax unrealistic assumptions and capture various aspects of human social patterns, including: multitype models that distinguish between classes of individuals (e.g. age or risk behaviour); metapopulation models (e.g. cities connected by flights); households models, capturing social grouping imposed by household, school and workplace structures; network models; spatial models, e.g. incorporating transmission reduction with distance or movements dependent on population density; and complex individual-based stochastic simulations, which are flexible but involve many parameters. For reviews, see Rock et al. [37] and Keeling and Rohani [23].

3 REAL-TIME DECISION MAKING DURING PANDEMICS

Infectious disease modelling is increasingly used to support decision making in real-time in order to choose best control strategies. In 2001, for example, modelling was used to estimate the impact of potential control strategies during the outbreak of foot-and-mouth disease in the UK [14, 24]. During the last influenza pandemic in 2009, a cost-effectiveness analysis involving a transmission model fitted to incoming data informed the UK government on the likely impact of alternative vaccination strategies [2]. More recently, during the Ebola outbreak in West Africa, models were used to forecast the likely course of the outbreak [41], evaluate the benefits and risks of introducing Ebola community centres [26], estimate the impact of new beds [6] and evaluate clinical trials for experimental treatments [11].

In these situations, trying to develop increasingly more complex models in order to integrate additional parameters relevant to decision makers is often not possible using traditional methods. The reason is that traditional methods of evidence synthesis are based on computationally intensive algorithms poorly adapted to quick responsive real-time inference. The quality of real-time data involved integrating additional modelling layers to reflect potential censoring and sources of uncertainties. Finally, these models need to incorporate simple summary outputs which can be handled during decision making.

A real-time modelling toolbox needs to be developed to provide a set of modular methodologies which can be picked up to build a model flexible enough to integrate relevant complexities while still being fitted in a short amount of time. Such tools could involve heavily parallelised methods that exploit multi-core computer architecture such as particle filters, variational Bayesian approaches, or approximation [5]. It is, however, possible that any computational gain could be offset by approximation error. It would thus also be necessary to develop, prior to these crises, studies quantifying the level of bias introduced by using 'fast' methods.

4 PREDICTING THE PEAK DEMAND FOR HEALTHCARE

Prediction of height and timing of peak incidence is crucial to estimate the stress on the health care system, and hence to inform decision-makers on how to allocate resources to manage the outbreak most cost-effectively. Hospitals running out of beds during more severe influenza seasons are not uncommon [7, 8], with patient care severely delayed and other hospital services postponed to after the winter crisis. Conversely, enough resources in terms of bed capacity in treatment centres has been shown to have contributed to controlling the 2014 Ebola outbreak [27].

Given the abundance of studies discussing, comparing and testing methods for prediction of influenza epidemics [32, 34, 40, 44], we focus here on the case of influenza, although comments naturally extend to other infections. Among the many epidemiologically relevant epidemic characteristics [40], peak timing and height are among the most widely considered [32, 33, 38].

Stochastic simulations of simple epidemic models based on the 'mass-action' mixing assumption highlight how peak epidemic timing can vary widely as a consequence of the random delays in the early epidemic phase. However, once the epidemic takes off and the number of cases becomes large enough to motivate a deterministic approximation, the explosive nature of exponential growth is such that uncertainties regarding the size of the population under consideration or the fraction effectively susceptible to infection has only marginal impact on how quickly the peak is reached. Mathematically speaking, the stochastic counterpart of model (2), if not for the initial and final epidemic phases dominated by random events, consists of a 'deterministic' central phase the duration of which is $O(1)$ (i.e. independent of the population size N) [1]. It is unsurprising, therefore, that numerous retrospective studies of influenza concluded peak timing prediction can be already accurate weeks in advance (cited numbers generally range form 4 to 7 [32, 33, 38]), at

least in those years characterised by a single epidemic wave and provided enough data is promptly available. Prospective, real-time forecasts [39, 42] also proved reasonably accurate even up to 9 weeks in advance.

The delays between infections and reliable data on confirmed cases becoming available (1-2 weeks [34] or even longer [9]) may threaten the usefulness of fast prediction methods. Alternatives using more promptly available data, such as influenza-like illness, acute respiratory infection data or absenteeism, or data surrogates such as Google 'Flu Trends and search engine queries, unavoidably inherit the inaccuracies of such data sources. For example, the wide media coverage during the 2012-2013 epidemic in the USA has been claimed as a potential cause for the time discrepancy between Google search activity and the peak in real infections [39]. However, forecast methods based on combinations of social (e.g. Internet searches) and physical indicators (e.g. absolute humidity) appear to outperform those based on single indicators alone, and can be further strengthened by accounting for uncertainty due to offical estimates undergoing revision after publication [9].

If peak timing predictions appear broadly successful, the prediction of peak height is much more challenging [32, 34]. A potential explanation is that, even when mass-action mixing is a reasonable assumption, the height of the peak is highly dependent on the exact speed of epidemc growth [32] and other factors that are hard to estimate, such as: the distribution of (often partial) susceptibility in the population, the rate of under-reporting, and the effect of control policies (e.g. vaccination) and behavioural change (e.g. self-quarantining, reduced mixing, but also "flu parties" [29]).

When the mass-action assumption is not justifiable, models with a more complex structure might need to be employed, complicating the matter further. In a metapopulation framework, the height and timing of peak incidence in the full population results from the superposition of the subpopulation dynamics. As such, they depend on subpopulation sizes, but also, crucially, on the times at which the epidemic jumps between subpopulations, which are often 'rare' events that are hard to predict accurately. However, because as discussed above peak timing can be predicted in each subpopulation provided local data is promptly available, the work of Shaman et al. [39], which ignores a metapopulation structure and treats different cities as independent of each other, could still forecast peak timing for many of the 108 cities considered.

Forecasting methods typically involve complex simulation models and relatively simple statistical fitting procedures [33] or relatively simple models whose potential misspecification is compensated for by sophisticated statistical approaches, such as particle filters [32] and data assimilation ensemble approaches originally adopted in numerical weather prediction [38, 39]. No single method appears uniformly better than others and even after the peak has passed there are instances when peak timing forecasts are inaccurate [44]. In particular, all the methodologies tested in [44] struggled in performing prediction for those years characterised by two (or even three) separate peaks, as they were the result of separate outbreaks of different strains of influenza that the models were not designed to capture. Integrating the sophisticated statistical techniques with more realistic models of seasonal influenza is a challenge for the future, but the lack of accurate and promptly available data remains one of the main limitations in epidemiology, especially compared to weather and climate predictions [28].

5 SYNDROMIC SURVEILLANCE FOR SITUATIONAL AWARENESS

Syndromic surveillance is the monitoring of the number of cases of illness in a population with a specified syndrome [36]. Priorities for the algorithms used to monitor real-time syndromic signals are robustness, speed, and the ability to model signals at different scales (some signals have many days of zero cases and some have many cases every day).

Methods used in practice to monitor signals such as vomiting or influenza-like illness at different points in the healthcare system include the 'early aberration reporting system' developed by the CDC [21], the 'moving epidemic method' [43], and the 'rising activity, multi-level mixed effects, indicator emphasis' (RAMMIE) method developed by Public Health England [31]. These are statistical methods not based on dynamical systems of transmissionthat are not transmission dynamic. RAMMIE, for example, is based on a Poisson model (1) with an additional multi-level structure to exploit signals from hierachical geographies (national, regional, and local levels) [31].

Such approaches, applied to daily data, can improve situational awareness during mass gatherings [18], provide support during environmental problems [13], and detect and follow trends in larger seasonal influenza outbreaks [20]. However, smaller outbreaks, for example small gastrointestinal outbreaks [10] or anthrax releases [30], are unlikely to be detected; the possibility of greater sensitivity therefore remains open.

6 RESPONDING TO DISEASE OUTBREAKS WITH ENVIRONMENTAL SOURCES

Diseases which arise from environmental sources rather than by person to person spread – for example anthrax and Legionnaires disease as opposed to measles or influenza – do not have the nonlinearity inherent in SIR-type models. This makes them mathematically simpler but still have their own modelling challenges [12]. These outbreaks will involve fewer people than expected from a pandemic influenza wave but may be deliberate or arise during high profile events such as the Olympic Games meaning the timescales for decisions are shorter and uncertainty in data greater.

As was shown in [30] traditional syndromic surveillance schemes are unlikely to detect the atypical emerging disease outbreaks. Instead such infections are likely to be detected within hospital settings. This means the key data required to run models must be collected on the fly from cases. Given the speed of data collection there is likely to be uncertainty in quality of data. Efficient data transfer is essential (using electronic data capture software and transfer formats [15]) and wider adoption of such technology by responders is key, where practical.

As shown in [16], methods may be adapted on the fly to support outbreaks. Work between outbreaks can extend methods more robustly (by allowing for additional delays or multiple sources) but every outbreak is unique and so the perfect model design and data demands are difficult to write down *a priori*.

A key uncertainty is the location of people at the time of infection and the number of people present that did not get infected. Traditional epidemiological methods of interviewing observed cases mean that one can reasonably expect home and work location of cases. This may be matched to home and work locations of the general population from surveys or censuses but if infection arises during travel or leisure activity the infection location may be missed. Even with more detailed travel history from cases this denominator population is hard to define. In [19] the authors have looked at transient movements using novel data set including detailed travel histories for defining denominators but such commercial datasets are expensive to maintain and often appear as 'black boxes' to eventual model users. To be viable for translation to public health organisations the tools must be affordable and transparent. Emerging data sources such as mobile phone location data may be a benefit to public health authorities, to define the at risk population, but access to such data is hard to ensure (particularly ahead of time so methods can be developed) and accuracy is hard to quantify.

Whilst inferential methods for reverse epidemiology (finding exposure locations given cases) exist [12], further methodological development is required, for example fusing intervention models [35] with inference tools to ensure realistic modelling of interventions. Futhermore, given that evidence for human dose response arises from animal studies and for relatively high doses, the infection of humans receiving low doses of biological agent is uncertain.

Given uncertainty in the models, and their justification, another challenge is around model selection (in absence of clear nesting of models) and presentation of key assumptions in an intuitive way to a lay audience. Models may have challenges in explaining concepts to such audiences but are critical in such situations to support outbreak control team work.

7 OUTBREAK CONTROL IN CARE HOMES

Care homes are an integral setting for disease transmission and control, especially within the context of an ageing population. A recent modelling study of scabies in care homes [25] suggested that early detection of an infectious index case is critical in establishing who and when to treat. Uncertainty in determining the index case leads to possibly a greater number of residents being infected with the potential of infecting staff who maintain frequent links between the care home and the general population.

Traditionally, the study of transmission of an infectious agent within such a setting has been studied using either an agent-based simulation model or a mean-field model [23]. For the sake of argument and without loss of generality, a care home can be regarded as a large household. These models capture the complete range of stochastic behaviours using a large set of ordinary differential equations which can be expressed succinctly as

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \boldsymbol{Q}\mathbf{p}\,, \qquad \mathbf{p}(0) = \mathbf{p}_0\,, \tag{3}$$

where $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ is the household transition matrix and $\mathbf{p}$ is the probability that a household is in a certain infection configuration. For example, if we consider the SIR model as in (2) then we would have $\mathbf{p}_{s,i,r}(t)$ being the probability that a household has s susceptibles, i infectious and r removed. This is made possible because we count the number of events of each type that can occur rather than keep track of the population numbers. For example, in the SIR model only two events can occur, namely infection $(S, I, R) \to (S-1, I+1, R)$ and recovery $(S, I, R) \to (S, I-1, R+1)$. For a detailed overview of this type of models, see [3, 4]. The solution for (3) involves the exponential of a matrix:

$$\mathbf{p}(t) = \exp(t\boldsymbol{Q})\mathbf{p}_0\,. \tag{4}$$

To fully achieve the modelling ideal of sufficiently accounting for uncertainty, in both structure and parameter values, and being able to propagate it within a modelling framework, computationally efficient algorithms for solving the master equation are needed. Jenkinson and Goutsias [22] have considered an implicit Euler implementation to approximate the solution of (4). More recently, Kinyanjui et al. [25] have demonstrated that there exist computational advantages in solving the matrix exponential using expansion-based methods for the time-homogeneous case, i.e. when $\boldsymbol{Q}$ is time-invariant. The computational advantages gained allowed for a full Bayesian quantification of scabies transmission and control within households fully accounting for uncertainty [25]. However, there is still an open research question as to what happens when we have interactions between households making $\boldsymbol{Q}$ temporally heterogeneous.

8 CONCLUSIONS

In summary, we have outlined a series of policy-driven modelling contexts where data, time pressures and uncertainty limit the detail that can be incorporated into models, highlighting where further – particularly methodological – work is needed.

REFERENCES

[1] Håkan Andersson and Tom Britton. 2000. *Stochastic Epidemic Models and Their Statistical Analysis.* Springer Lectures Notes in Statistics, Vol. 151. Springer, Berlin.

[2] Marc Baguelin, Albert Jan Van Hoek, Mark Jit, Stefan Flasche, Peter J. White, and W. John Edmunds. 2010. Vaccination against pandemic influenza A/H1N1v in England: A real-time economic evaluation. *Vaccine* 28, 12 (mar 2010), 2370–2384. https://doi.org/10.1016/j.vaccine.2010.01.002

[3] Frank G Ball and Owen D. Lyne. 2001. Stochastic Multitype SIR Epidemics among a Population Partitioned into Households. *Advances in Applied Probability* 33, 1 (2001), 99–123. http://projecteuclid.org/euclid.aap/999187899

[4] Andrew J. Black and Joshua V. Ross. 2015. Computation of epidemic final size distributions. *Journal of Theoretical Biology* 367 (2015), 159–165. https://doi.org/10.1016/j.jtbi.2014.11.029 arXiv:arXiv:1407.3887v2

[5] Elizabeth Buckingham-Jeffery, Valerie Isham, and Thomas House. 2018. Gaussian process approximations for fast inference from infectious disease data. *Mathematical Biosciences* 301 (2018), 111–120.

[6] Anton Camacho, Adam Kucharski, Yvonne Aki-Sawyerr, Mark A White, Stefan Flasche, Marc Baguelin, Timothy Pollington, Julia R Carney, Rebecca Glover, Elizabeth Smout, et al. 2015. Temporal changes in Ebola transmission in Sierra Leone and implications for control requirements: a real-time modelling study. *PLoS currents* 7 (2015).

[7] Denis Campbell, Pamela Duncan, and Sarah Marsh. 2018. NHS patients dying in hospital corridors, A&E doctors tell Theresa May. https://www.theguardian.com/society/2018/jan/11/nhs-patients-dying-in-hospital-corridors-doctors-tell-theresa-may. First published: 2018-01-11.

[8] Denis Campbell and Sarah Marsh. 2017. NHS crisis: 20 hospitals declare black alert as patient safety no longer assured. https://www.theguardian.com/society/2017/jan/11/nhs-crisis-20-hospitals-declare-black-alert-as-patient-safety-no-longer-assured. First published: 2017-01-11.

[9] Prithwish Chakraborty, Pejman Khadivi, Bryan Lewis, Aravindan Mahendiran, Jiangzhuo Chen, Patrick Butler, Elaine O Nsoesie, Sumiko R Mekaru, John S

Brownstein, Madhav V Marathe, et al. 2014. Forecasting a moving target: Ensemble models for ILI case count predictions. In *Proceedings of the 2014 SIAM international conference on data mining*. SIAM, 262–270.

[10] Felipe J. Colón-González, Iain R. Lake, Roger A. Morbey, Alex J. Elliot, Richard Pebody, and Gillian E. Smith. 2018. A methodological framework for the evaluation of syndromic surveillance systems: a case study of England. *BMC Public Health* 18, 1 (2018), 544. https://doi.org/10.1186/s12889-018-5422-9

[11] Ben S. Cooper, Maciej F. Boni, Wirichada Pan-ngum, Nicholas P. J. Day, Peter W. Horby, Piero Olliaro, Trudie Lang, Nicholas J. White, Lisa J. White, and John Whitehead. 2015. Evaluating Clinical Trial Designs for Investigational Treatments of Ebola Virus Disease. *PLOS Medicine* 12, 4 (04 2015), 1–14. https://doi.org/10.1371/journal.pmed.1001815

[12] Joseph R. Egan and Ian M. Hall. 2015. A review of back-calculation techniques and their potential to inform mitigation strategies with application to non-transmissible acute infectious diseases. *Journal of The Royal Society Interface* 12, 106 (2015).

[13] Gillian E.Smith, Zharain Bawa, Yolande Macklin, Roger Morbey, Alec Dobney, Sotiris Vardoulakis, and Alex J.Elliot. 2015. Using real-time syndromic surveillance systems to help explore the acute impact of the air pollution incident of March/April 2014 in England. *Environmental Research* 136 (2015), 500–504.

[14] Neil M. Ferguson, Christl A. Donnelly, and Roy M. Anderson. 2001. Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature* 413, 6855 (oct 2001), 542–548. https://doi.org/10.1038/35097116

[15] Thomas J.R. Finnie, Andy South, Ana Bento, Ellie Sherrard-Smith, and Thibaut Jombart. 2016. EpiJSON: A unified data-format for epidemiology. *Epidemics* 15 (2016), 20–26.

[16] Maya Gobin, Jeremy Hawker, Paul Cleary, Thomas Inns, Daniel Gardiner, Amy Mikhail, Jacquelyn McCormick, Richard Elson, Derren Ready, Tim Dallman, Iain Roddick, Ian Hall, Caroline Willis, Paul Crook, Gauri Godbole, Drazenka Tubin-Delic, and Isabel Oliver. 2018. National outbreak of Shiga toxin-producing Escherichia coli O157:H7 linked to mixed salad leaves, United Kingdom, 2016. *Eurosurveillance* 23, 18 (2018).

[17] Nicholas C Grassly and Christophe Fraser. 2008. Mathematical models of infectious disease transmission. *Nature Reviews Microbiology* 6, 6 (2008), 477.

[18] Adi V. Gundlapalli, Jonathan Olson, Sean P. Smith, Michael Baza, Robert R. Hausam, Louise J. Eutropius, Stanley L. Pestotnik, Karen Duncan, Nancy Staggers, Pierre Pincetl, and Matthew H. Samore. 2007. Hospital electronic medical record-based public health surveillance system deployed during the 2002 Winter Olympic Games. *American Journal of Infection Control* 35 (2007), 163–171. Issue 3.

[19] Penelope A. Hancock, Yasmin Rehman, Ian M. Hall, Obaghe Edeghere, Leon Danon, Thomas A. House, and Matthew J. Keeling. 2014. Strategies for Controlling Non-Transmissible Infection Outbreaks Using a Large Human Movement Data Set. *PLOS Computational Biology* 10, 9 (2014), e1003809. https://doi.org/10.1371/journal.pcbi.1003809

[20] S. E. Harcourt, G. E. Smith, A. J. Elliot, and R. Pebody. 2012. Use of a large general practice syndromic surveillance system to monitor the progress of the influenza A(H1N1) pandemic 2009 in the UK. *Epidemiology and Infection* 140 (2012), 100–105.

[21] Lori Hutwagner, William Thompson, G Matthew Seeman, and Tracee Treadwell. 2003. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). *Journal of urban health : bulletin of the New York Academy of Medicine* 80, 2 Suppl 1 (2003), i89–i96. https://doi.org/10.1007/PL00022319

[22] Garrett Jenkinson and John Goutsias. 2012. Numerical integration of the master equation in some models of stochastic epidemiology. *PloS one* 7, 5 (jan 2012), e36160. https://doi.org/10.1371/journal.pone.0036160

[23] Matt J Keeling and Pejman Rohani. 2007. *Modeling infectious diseases in humans and animals*. Princeton University Press.

[24] Matt J Keeling, Mark EJ Woolhouse, Darren J Shaw, Louise Matthews, Margo Chase-Topping, Dan T Haydon, Stephen J Cornell, Jens Kappey, John Wilesmith, and Bryan T Grenfell. 2001. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* 294, 5543 (2001), 813–817.

[25] Timothy Kinyanjui, Jo Middleton, Stefan Güttel, Jackie Cassell, Joshua Ross, and Thomas House. 2018. Scabies in residential care homes: Modelling, inference and interventions for well-connected population sub-units. *PLOS Computational Biology* 14, 3 (2018), e1006046.

[26] Adam A.J. Kucharski, Anton Camacho, Francesco Checchi, Ron Waldman, Rebecca R.F. Grais, Jean-Clement J.-C. Cabrol, Sylvie Briand, Marc Baguelin, Stefan Flasche, Sebastian Funk, William John Edmunds, and W. John Edmunds. 2015. Evaluation of the Benefits and Risks of Introducing Ebola Community Care Centers, Sierra Leone. *Emerging infectious diseases* 21, 3 (2015), 393–399. https://doi.org/10.3201/eid2103.141892

[27] Adam J Kucharski, Anton Camacho, Stefan Flasche, Rebecca E Glover, W John Edmunds, and Sebastian Funk. 2015. Measuring the impact of Ebola control measures in Sierra Leone. *Proceedings of the National Academy of Sciences* 112, 46 (2015), 14366–14371.

[28] Tom Lindström, Michael Tildesley, and Colleen Webb. 2015. A bayesian ensemble approach for epidemiological projections. *PLoS computational biology* 11, 4 (2015), e1004187.

[29] Donald G. McNeil Jr. 2009. Debating the Wisdom of 'Swine Flu Parties'. https://www.nytimes.com/2009/05/07/world/americas/07party.html. First published: 2009-05-06.

[30] RA Morbey, AJ Elliot, A Charlett, S Ibbotson, NQ Verlander, S Leach, I Hall, I Barrass, M Catchpole, B McCloskey, et al. 2014. Using public health scenarios to predict the utility of a national syndromic surveillance programme during the 2012 London Olympic and Paralympic Games. *Epidemiology & Infection* 142, 5 (2014), 984–993.

[31] Roger A. Morbey, Alex J. Elliot, Andre Charlett, Neville Q. Verlander, Nick Andrews, and Gillian E. Smith. 2015. The application of a novel 'rising activity, multi-level mixed effects, indicator emphasis' (RAMMIE) method for syndromic surveillance in England. *Bioinformatics* 31, 22 (2015), 3660–3665. https://doi.org/10.1093/bioinformatics/btv418

[32] Robert Moss, Alexander Zarebski, Peter Dawson, and James M McCaw. 2016. Forecasting influenza outbreak dynamics in Melbourne from Internet search query surveillance data. *Influenza and other respiratory viruses* 10, 4 (2016), 314–323.

[33] Elaine Nsoesie, Madhav Mararthe, and John Brownstein. 2013. Forecasting peaks of seasonal influenza epidemics. *PLoS currents* 5 (2013).

[34] Elaine O Nsoesie, John S Brownstein, Naren Ramakrishnan, and Madhav V Marathe. 2014. A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza and other respiratory viruses* 8, 3 (2014), 309–316.

[35] Gabriel Rainisch, Martin I Meltzer, Sean Shadomy, William A Bower, and Nathaniel Hupert. 2017. Modeling Tool for Decision Support during Early Days of an Anthrax Event. *Emerging Infectious Diseases* 23, 1 (2017), 46–55.

[36] Arthur Reingold. 2013. If syndromic surveillance is the answer, what is the question? *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science* 1, 2 (2013).

[37] Kat Rock, Sam Brand, Jo Moir, and Matt J Keeling. 2014. Dynamics of infectious diseases. *Reports on Progress in Physics* 77, 2 (2014), 026602.

[38] Jeffrey Shaman and Alicia Karspeck. 2012. Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences* 109, 50 (2012), 20425–20430.

[39] Jeffrey Shaman, Alicia Karspeck, Wan Yang, James Tamerius, and Marc Lipsitch. 2013. Real-time influenza forecasts during the 2012–2013 season. *Nature communications* 4 (2013), 2837.

[40] Farzaneh Sadat Tabataba, Prithwish Chakraborty, Naren Ramakrishnan, Srinivasan Venkatramanan, Jiangzhuo Chen, Bryan Lewis, and Madhav Marathe. 2017. A framework for evaluating epidemic forecasts. *BMC infectious diseases* 17, 1 (2017), 345.

[41] WHO Ebola Response Team. 2014. Ebola Virus Disease in West Africa – The First 9 Months of the Epidemic and Forward Projections. *New England Journal of Medicine* 371, 16 (oct 2014), 1481–1495. https://doi.org/10.1056/NEJMoa1411100

[42] Sherry Towers and Zhilan Feng. 2009. Pandemic H1N1 influenza: predicting the course of a pandemic and assessing the efficacy of the planned vaccination programme in the United States. *Eurosurveillance* 14, 41 (2009), 19358.

[43] Tomás Vega, Jose Eugenio Lozano, Tamara Meerhoff, René Snacken, Joshua Mott, Raul Ortiz de Lejarazu, and Baltazar Nunes. 2013. Influenza surveillance in Europe: Establishing epidemic thresholds by the Moving Epidemic Method. *Influenza and other Respiratory Viruses* 7, 4 (2013), 546–558. https://doi.org/10.1111/j.1750-2659.2012.00422.x

[44] Wan Yang, Alicia Karspeck, and Jeffrey Shaman. 2014. Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics. *PLoS computational biology* 10, 4 (2014), e1003583.