

Real-time modelling of a pandemic influenza outbreak

Paul J Birrell,¹ Richard G Pebody,² André Charlett,²
Xu-Sheng Zhang² and Daniela De Angelis^{1,2*}

¹Medical Research Council Biostatistics Unit, Cambridge Institute of Public Health, University of Cambridge, Cambridge, UK

²National Infections Service, Public Health England, London, UK

*Corresponding author daniela.deangelis@mrc-bsu.cam.ac.uk

Declared competing interests of authors: none

Published October 2017

DOI: 10.3310/hta21580

Scientific summary

Modelling of a pandemic influenza outbreak

Health Technology Assessment 2017; Vol. 21: No. 58

DOI: 10.3310/hta21580

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

The UK National Risk Register lists an outbreak of pandemic influenza as the largest risk faced by the UK outside acts of terrorism. A prompt public health response to a pandemic is, therefore, vital if the effects of the pandemic are to be mitigated. A business-critical component of such a response is real-time epidemic modelling of the outbreak. As an epidemic progresses, real-time modelling should gradually refine our knowledge about the epidemic and, in particular, the burden it will place on health-care services. The model should also be able to act as a simulator, allowing for the prospective examination and evaluation of the impact of any proposed epidemic interventions.

Real-time modelling of an ongoing pandemic is, however, not a straightforward task. The model is a tool for the statistical analysis of epidemic surveillance data. Model parameters should be progressively informed as data become available over time, providing increasingly accurate assessments and predictions of the epidemic evolution. However, as experienced in responding to the 2009 'swine flu' (A/H1N1pdm virus) outbreak, real-life epidemic surveillance data are frequently far messier than anticipated, containing contamination, noise and biases that are not always straightforward to foresee. Additionally, epidemics are rarely left to play out naturally. Public health interventions designed either to contain transmission or to relieve the burden placed on health-care services have the potential to drastically alter our perception of a pandemic or, as in 2009, to interrupt the patterns observed in key data streams. These problems are addressed by using data from multiple sources, so that the different data types can complement each other, in order to eliminate noise and/or bias.

A modelling framework must, therefore, be able to accommodate this wealth of different data types (typically arriving at daily intervals) and to do so in a timely fashion. It must also be robust to interventions and be able to provide analyses stratified by age groups and geographical location as required by policy-makers. Any such model is likely to be highly complex and computationally challenging.

After 2009, a model was developed that could incorporate a number of different types of data, but that did not allow for spatial stratification. The model was also relatively costly to run in real time, as it was implemented using a technique for statistical computation called Markov chain Monte Carlo (MCMC). To this point, the model developed had only ever been used in the retrospective analysis of an entire epidemic, not for real-time purposes. Although MCMC is useful in epidemic reconstruction, it is inefficient for real-time use, as, when new data arrive, it involves the reanalysis of the entire data set. The ideal method needs, instead, to be sequential so that only the incoming data are used to update the analyses. Sequential Monte Carlo (SMC) techniques provide such an alternative approach.

Objectives

The central objective of this study is to advance the state of the art of real-time modelling of influenza epidemics and to provide a tool to monitor and predict the development of an ongoing pandemic outbreak. These advancements involve:

- Investigating spatial modelling of epidemics to understand how best to account for regionally varying epidemic activity. Two candidate approaches are proposed and examined through the analysis of pandemic surveillance data from 2009. In particular, the strengths and weaknesses of each approach are examined to prospectively assess their potential utility in a future outbreak.

- Building capacity in terms of the different types and increasing volume of data that can be used for real-time modelling.
- Improving the efficiency with which real-time statistical inference can be made.

In light of the above, a suite of software is produced to achieve the above objectives, designed to provide support to national public health bodies in the event of a pandemic, and tailored to the specific requirements of Public Health England (PHE), the responsible public health body in England.

Methods

Two candidate extensions to the existing epidemic model to accurately track a spatially diverse epidemic are proposed. The first approach, labelled the parallel-region (PR) approach, assumes that the epidemic is already established in each region by the time the modelling is initiated. From this point onwards, inter-region transmission can be considered negligible and the epidemics evolve in each region independently, although they will still share some characteristics. Spatially varying parameters, estimated separately for each region, are those dependent on population behaviour and composition: the parameters that govern health-care-seeking behaviour; the initial seeding; and the reproductive number R_0 .

The second approach, labelled the meta-region (MR) approach, uses census data on commuting behaviour to generate rates of movement between the different regions. Here, the country is thought of as a single MR, with a population stratified by both age and location. The commuting data inform the relative rates of contact between individuals of each of the strata. There are a number of competing subhypotheses within this approach; these regard the effects of density dependence on contact rates, whether or not it is appropriate to assume that individuals commute at random, and the importance of the initial seeding of infection. There will be a preliminary phase of model choice that will identify the relative performance of the model under each combination of these assumptions.

These two methods are implemented to reconstruct the 2009 pandemic. Data include a time series of general practice (GP) consultation data, a short time series of virologically confirmed cases (found through initial attempts at tracing contacts of early infections), virological swabbing data and serological sampling data. The GP consultation data, derived from syndromic surveillance, are counts of individuals reporting a collection of symptoms, known as influenza-like illness. These data, therefore, contain a significant proportion of individuals who are consulting for non-pandemic illnesses, contaminating the data. The degree of contamination is identified by the virological swabbing data, as a small subsample of individuals consulting their GP provide swab samples for testing. The tests will indicate the presence or absence of the pandemic infection. Together, the GP consultation data and the virological swabbing data give a time series that is linked to the pattern of pandemic infection. The serological data come from the testing of blood sera samples taken during the pandemic and tested for the presence of immunity-conferring antibodies. This informs the levels of cumulative incidence, giving scale to the pattern of infection estimated from the GP data. Each data component is, therefore, vital in disentangling the underlying epidemic dynamics. In future pandemics, it is anticipated that surveillance data of each of these types will be enriched. Additionally, hospitalisation data will be available from all NHS trust hospitals, recording all admissions, admissions requiring intensive care in particular, of patients with the pandemic infection.

As an epidemic reconstruction, the spatial analysis is implemented using MCMC. In a real-time monitoring context, however, it is necessary to develop and test an alternative algorithm to allow for the computationally efficient iteration of epidemic analyses. Starting with a 'basic' SMC approach taken from the literature, a number of algorithmic developments are proposed to cope with data simulated to mimic the characteristics of the 2009 pandemic. Here, it is assumed that a public health intervention drastically disrupts the temporal pattern of the observed data and tests the ability of this iterative procedure to respond and adapt appropriately, providing reliable assessment of the epidemic dynamics. The performance of the SMC algorithm is compared with the 'gold standard' MCMC in terms of both the quality of the resulting estimates and computational efficiency.

Results

Spatial modelling

Results show that both the PR and the MR approaches are able to reconstruct the epidemic dynamics well, with the PR model providing a better fit to the data than all variants of the MR model. Additionally, the PR approach offers computational benefits, as it can exploit parallel computing and can be implemented in a fraction of the time taken for the MR approach. However, the PR approach does require significant epidemic activity to have occurred in each region for reliable estimation. If epidemic prediction is required in the early stages of a pandemic, the MR approach might be more useful. Within the MR approach, strong density-dependent effects are found in contact rates between individuals, that is, the chance of two people meeting is inversely proportional to the population size of the region in which they interact. Key model parameters that influence transmission are consistent across both PR and MR approaches, with, in particular, R_0 consistently being estimated to be around 1.8. We are also able to estimate the drop in contact due to the over-summer school holidays. Contact rates among 5- to 14-year-olds fell to 0.6–0.7% of their term-time value, leading to a drop in R_0 of 43–50%.

Computational efficiency

In terms of the SMC algorithm to sequentially update estimates as new data become available, in the majority of cases, a fairly straightforward SMC implementation taken from literature would suffice. However, sequential estimation becomes more problematic when the newly observed data are highly informative, as in the situation when a 'shock' to the surveillance data is introduced to mimic the effect of a public health intervention. In this case, the SMC algorithm requires very careful construction to ensure that it retains the capacity to accurately track the epidemic. A semi-automated algorithm is pieced together to minimise the computational effort required to update the analyses. In a moderately complex example, in the immediate aftermath of the intervention, the SMC method even outperforms the MCMC, providing more reliable estimates and, with only moderate parallelisation (of the kind that most modern desktops are more than capable), SMC will prove faster to implement.

All of the above models and algorithms have been incorporated into software that is now available for use by PHE in the event of a pandemic. Key PHE personnel are receiving training in its use, which, for the foreseeable future, will continue to be offered on an ongoing basis.

Conclusions

This project divides neatly into two components: (1) developing the modelling methodology to provide information to policy-makers at the spatial resolution they require, and (2) developing the statistical computing methodology to make robust and timely inference. In terms of the spatial modelling, the PR approach is the most suited approach for real-time epidemic monitoring, even though it has little predictive power early in the epidemic. This shows that either the effects of inter-region transmission are transient, or the available commuting data do not characterise the movement of individuals between regions particularly well. As far as the computational methodology is concerned, in the real-time context, sequential methods for analysis have been shown to be equally adept at providing inference as the more established MCMC, but with a considerable computational advantage. All of the findings have been encoded into software.

The research recommendations arising out of this work are as follows.

1. To understand the impact of public health interventions. Here, interventions involving school closure and the provision of a service to relieve the burden placed on GPs have been considered, and the model has been accordingly adapted. However, investigation of the modelling adaptations required to incorporate vaccination uptake and effects of antivirals (among others) would allow the assessment of these policies, both prospectively and retrospectively. It is anticipated that adaptations to the transmission component of the model to account for such measures would be reasonably straightforward, but to accommodate the data that would inform these adaptations would be more complex.

2. To investigate the utility of alternative sources of epidemic surveillance. The statistical analysis of the epidemic is reliant on surveillance data on the uptake of health-care services (GP consultations, hospitalisations). In a widespread and/or severe pandemic, these resources could be severely stretched; hospital beds may not be available and GP appointment books may be full. At this point, the data generated from these sources may become unreliable. As mentioned in point 1, data on vaccine uptake and antiviral prescriptions administered could potentially help to fill the knowledge gap, but there are alternative influenza surveillance mechanisms that could be exploited, and these should be investigated. Could Internet searches for key influenza terms be useful? Or sales of thermometers? Simplistic modelling studies exist, but there is a real gap in the use of alternative surveillance data in real-world examples.
3. To properly account for the uncertainty in serological results. It has been shown that serological data are particularly informative to the kind of modelling effort undertaken in this work. However, the sensitivity and specificity of the testing process is rarely considered, owing to uncertainty regarding the precise level of antibodies that constitutes long-term immunity to a virus. Proper handling of the uncertainty in these data, as well as their timely provision during a pandemic, is essential.

Trial registration

This trial is registered as ISRCTN40334843.

Funding

Funding for this study was provided by the Health Technology programme of the National Institute for Health Research (NIHR). Daniela De Angelis was supported by the UK Medical Research Council (Unit Programme Number U105260566) and by PHE. She received funding under the NIHR grant for 10% of her time. The rest of her salary was provided by the MRC and PHE jointly.

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.236

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nhr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nhr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nhr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: <http://www.nets.nhr.ac.uk/programmes/hta>

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 11/46/03. The contractual start date was in October 2012. The draft report began editorial review in September 2016 and was accepted for publication in January 2017. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health.

© Queen's Printer and Controller of HMSO 2017. This work was produced by Birrell *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by the NIHR Journals Library (www.journalslibrary.nhr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Health Technology Assessment Editor-in-Chief

Professor Hywel Williams Director, HTA Programme, UK and Foundation Professor and Co-Director of the Centre of Evidence-Based Dermatology, University of Nottingham, UK

NIHR Journals Library Editor-in-Chief

Professor Tom Walley Director, NIHR Evaluation, Trials and Studies and Director of the EME Programme, UK

NIHR Journals Library Editors

Professor Ken Stein Chair of HTA and EME Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

Professor Andrée Le May Chair of NIHR Journals Library Editorial Group (HS&DR, PGfAR, PHR journals)

Dr Martin Ashton-Key Consultant in Public Health Medicine/Consultant Advisor, NETSCC, UK

Professor Matthias Beck Chair in Public Sector Management and Subject Leader (Management Group), Queen's University Management School, Queen's University Belfast, UK

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Eugenia Cronin Senior Scientific Advisor, Wessex Institute, UK

Dr Peter Davidson Director of the NIHR Dissemination Centre, University of Southampton, UK

Ms Tara Lamont Scientific Advisor, NETSCC, UK

Dr Catriona McDaid Senior Research Fellow, York Trials Unit, Department of Health Sciences, University of York, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Professor of Wellbeing Research, University of Winchester, UK

Professor John Norrie Chair in Medical Statistics, University of Edinburgh, UK

Professor John Powell Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Institute of Child Health, UK

Professor Jonathan Ross Professor of Sexual Health and HIV, University Hospital Birmingham, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Professor Martin Underwood Director, Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, UK

Please visit the website for a list of members of the NIHR Journals Library Board:
www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: journals.library@nihr.ac.uk