Epidemics 10 (2015) 72–77

Contents lists available at ScienceDirect

Epidemics

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

Six challenges in measuring contact networks for use in modelling

K. Eames [a,*], S. Bansal [b,c], S. Frost [d], S. Riley [e]

[a] Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK
[b] Department of Biology, Georgetown University, Washington, DC 20057, USA
[c] Fogarty International Center, National Institutes of Health, Bethesda, MD 20893, USA
[d] Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK
[e] MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, Norfolk Place, London W21PG, UK

ARTICLE INFO

Article history:
Received 24 February 2014
Received in revised form 21 August 2014
Accepted 22 August 2014
Available online 30 August 2014

Keywords:
Networks
Contact patterns
Disease transmission
Mathematical modelling
Network measurement
Model-driven data collection

ABSTRACT

Contact networks are playing an increasingly important role in epidemiology. A contact network represents individuals in a host population as nodes and the interactions among them that may lead to the transmission of infection as edges. New avenues for data collection in recent years have afforded us the opportunity to collect individual- and population-scale information to empirically describe the patterns of contact within host populations. Here, we present some of the current challenges in measuring empirical contact networks. We address fundamental questions such as defining contact; measurement of non-trivial contact properties; practical issues of bounding measurement of contact networks in space, time and scope; exploiting proxy information about contacts; dealing with missing data. Finally, we consider the privacy and ethical issues surrounding the collection of contact network data.

Introduction

Early mathematical models of infectious disease dynamics treated all individuals as identical, and assumed that they all interacted with each other at the same constant rate. More recent models customarily split the population into groups distinguished by characteristics such as age and location, introducing mixing rates defining interactions between groups (Mossong et al., 2008). Such models still assume that all individuals within a group are identical, and that the interaction between two given individuals is determined solely by the groups to which they belong.

In reality, social interactions are more nuanced and structured than such assumptions allow. Each person has an individual set of contacts that determine whom she may be infected by and whom she may infect. These contacts can be described by a network: a set of links between members of a population. Each link represents a (pathogen-dependent) opportunity for transmission.

There is a long history of the use of networks in epidemiology, in particular associated with contact tracing and outbreak investigations, which seek to identify risky interactions within a population (Klovdahl et al., 1994; Riley and Ferguson, 2006; Fraser et al., 2004). Likewise, there has been a great deal of recent work in developing models of transmission through networks (Keeling and Eames, 2005); challenges associated with such modelling are discussed elsewhere in this series (Pellis et al., in this issue).

There are many different types of network in epidemiology: for example, we can consider social or sexual contacts between individuals; patient movement between hospitals; airline travel between cities. In each case, the nodes represent relevant epidemiological units (individuals, hospitals, cities) and the links describe connections between nodes that could facilitate transmission. Here our focus is predominantly on the measurement of links between individual people, but many of the challenges below apply more generally; we note that recent progress has been made in using networks for understanding animal epidemiology, prompted in part by large-scale measurement of livestock movements (Brooks-Pollock et al., 2014).

In models that consider subgroups rather than individuals, rather than requiring the strength of contact between *individuals*, we need to know about contacts between *groups*. This group-level information is often collected through studies carried out at the individual level (Mossong et al., 2008; Read et al., 2012; Eames et al., 2012); thus the data-collection challenges associated with models that contain any sort of contact structure are related to those inherent in network models.

* Corresponding author. Tel.: +44 02079272469.
E-mail addresses: ken.eames@lshtm.ac.uk (K. Eames), shweta@sbansal.com (S. Bansal), sdf22@cam.ac.uk (S. Frost), s.riley@imperial.ac.uk (S. Riley).

http://dx.doi.org/10.1016/j.epidem.2014.08.006

With perfect knowledge, any outbreak of a directly transmitted infectious disease could be described by linking each infection to its infector. This *transmission tree* would show the course of infection through a population (Gardy et al., 2011). Such networks provide a natural way of visualising and conceptualising infection processes, and contain much information about those types of interactions that result in transmission (Cauchemez et al., 2011).

A transmission tree is a real and – in theory – measurable entity for pathogens that do not reproduce outside their host(s). However, the contact network – the network over which transmission *might* occur – is a more challenging theoretical concept. To make the fullest use of network methods, we require not merely the transmission tree of one outbreak of a pathogen in a population, but a network that contains all contacts relevant for transmission, whether or not they have been involved in transmission during a particular outbreak. Ideally, we would measure not only the presence/absence of a contact, but additional properties such as its strength (weight), duration, and when it occurred. Although conceptually straightforward, the challenge of obtaining information about the weights and dynamics of interactions between all possible pairs of individuals at all possible times is vast. Approximations, such as assuming that all weights are constant, are often made.

While some studies have sought to measure epidemiologically relevant networks in populations of interest (Klovdahl et al., 1994; Cauchemez et al., 2011; Conlan et al., 2011; Bearman et al., 2004; Salathé et al., 2010; Isella et al., 2011), the number of such studies is small. Below, we offer a set of challenges in collecting contact network data relevant to dynamic transmission modelling. It is not intended to be a complete list, and is biased by the preferences and interests of the authors. We hope that it will contribute to seeding conversations, research projects, and healthy disagreements.

1. Defining a contact

Our ability to define a potentially infectious contact depends on our knowledge about the dominant mode of transmission. For some infections, e.g. sexually transmitted and vector-born diseases, the relevant contacts may be difficult to measure but are clearly defined, and we can propose empirical studies to refute or confirm the existence of specific network structures (Lewis et al., 2008). However, the infection event is harder to define for respiratory pathogens, where it is not always clear precisely how infection passes from person to person. The infector and infectee must be in the same physical space within a short period of time, but it is difficult to be more precise. For example, if modelling transmission on public transport, can anyone on a bus be infected by anyone else or only those "nearby"? Decisions about what types of interaction matter are crucial when setting up network studies.

Proxy measures of contacts

We are often obliged to work with plausible proxy measures of contacts. Self-reported face-to-face conversations and skin-on-skin contact are frequently used as proxy measures of potentially infectious contacts (Mossong et al., 2008; Read et al., 2008). Age-specific mixing patterns from questionnaire studies have been highly influential in parameterising models of respiratory infection (Mossong et al., 2008), despite potential problems of inaccurate reporting and recall bias. The key challenge for the use of self-reported contact data to inform network models is to validate the relationship between reported contacts and infection. Modelling work has used different measured contact patterns to fit age-structured incidence or serology data (Goeyvaerts et al., 2010; Melegaro et al., 2011), but further work is needed to understand how to interpret the results. For example, if patterns of interactions involving physical contact provide the best fit to serological sampling, does this mean that infection actually spreads via physical contact, or just that such contacts provide a good proxy in a particular population? Extending such studies to multiple populations and multiple pathogens may shed further light on this issue.

Integrating genetic data

Genetic data potentially allow the full description of the infection tree (Gardy et al., 2011). The combination of self-reported social contacts and an accurate infection tree should permit much more accurate assessments of the relative importance of different routes of transmission, resulting in better predictive models of infection events. A necessary step for making genetic approaches useful is the collection of both genetic and detailed traditional "contact" information in the same study. A key issue is the completeness of the data collected: if only a small fraction of infections and/or a small fraction of relevant contacts are sampled, then it will be difficult to reconstruct infection trees or to draw conclusions about networks (Volz and Frost, 2013). Complete sampling is made even more difficult in cases of asymptomatic infection or when our understanding of what constitutes a relevant contact is incomplete (Resik et al., 2007).

Counterfactual contact data

Many studies do not measure contacts that have actually led to transmission; rather, they measure contacts that could potentially lead to transmission. However, there is no guarantee that individuals would behave in the same way when infectious (or when interacting with infectious individuals) (Van Kerckhove et al., 2013); such counterfactual scenarios are inherently unmeasurable. However, large-scale studies that quantify links made during infectiousness would add greatly to our ability to select the right mapping between "healthy" and "ill" contact patterns.

2. Bounding networks in space, time, and scope

An epidemiologically relevant network could, in theory, include practically everyone in the world. Although sophisticated mathematical models may include the population of the entire world (van den Broeck et al., 2011), we are unlikely to attempt to measure this network. Therefore in any study we must choose where to bound our network. The decision will depend on available time and resources, and on our understanding of what constitutes a relevant study community.

Permeable boundaries

Almost all network studies are constrained to be within a particular pre-defined study population, e.g. a school or hospital (Conlan et al., 2011; Salathé et al., 2010; Isella et al., 2011). However, it is only rarely – if ever – that there are no relevant contacts with individuals outside the study population. How much does it matter that we miss these "external" connections? In particular, how do we deal with the seeding of infection into our population without information about external contacts?

Time horizons

Studies may provide snapshots of contact networks, but the dynamic nature of interactions means that we expect networks to change over time (Bansal et al., 2010) (see Challenge 4, below). How can we best use networks collected over short time windows