Available online at www.sciencedirect.com
ScienceDirect
Social Networks 30 (2008) 127–135
SOCIAL NETWORKS
www.elsevier.com/locate/socnet

Collecting social contact data in the context of disease transmission: Prospective and retrospective study designs

Rafael T. Mikolajczyk, Mirjam Kretzschmar *

Department of Public Health Medicine, School of Public Health, University of Bielefeld, PO Box 100131, D-33501 Bielefeld, Germany

Abstract

Patterns of direct contacts between people ultimately determine the transmission rate for airborne pathogens like influenza; therefore, data regarding contact behaviour are essential for designing infectious disease control. In order to collect reliable data about social contacts, decisions about the survey methodology have to be made. A series of relevant surveys was conducted in 2003–2006 among university students in Bielefeld, Germany. Each survey focussed on specific methodological questions related to the number of contacts encountered during 1 day. In this study, we report on the analysis of different survey designs.

Keywords: Social contacts; Airborne infectious diseases; Survey methodology

1. Introduction

Many infectious diseases are transmitted via the air during contact between a contagious and a susceptible person. Transmission can occur through either the small infectious droplets released when coughing or speaking, or through an aerosol of infectious material. In the following analysis, we will refer to both modes as "airborne". Examples of diseases with an airborne transmission route include influenza, measles, smallpox, pertussis and many others. The probability of transmission and the resulting spread of these infectious diseases through a population depend on the number of contacts between individuals and the patterns of these contacts. By contact patterns, we refer to a description of who has contact with whom; paramount to this, we refer to the age-dependent probability that individuals contact each other. Other attributes of individuals can also influence their propensity for coming into contact including: socio-economic status, geographic location and the social network that they are part of. These contact patterns, together with a disease specific transmission probability upon contact determine the way that an infectious disease spreads through a population leading to the epidemiological outcomes that are measured on population level (Wallinga et al., 1999).

Collecting information about contact networks has a long tradition among social scientists. Fu (2007) provided a recent review of methodologies used in this field. Specifically, he examined observational studies, interviews and the diary approach. He identified the work by Gurevich (1961) as the first attempt to use the diary approach. This first study included 18 participants; the same approach was used later in a study with 27 (de Sola Pool and Kochen, 1978–1979) and subsequently with 78 participants (Lonkila, 1999). Several further studies followed, many of them assessing the accuracy of reporting social contacts (Higgins et al., 1985; Brewer, 2000; Brewer and Webster, 1999; Killworth and Bernard, 1979–1980) or other kinds of interactions like sexual contacts or drug use (Brewer and Garrett, 2001; Brewer et al., 1999, 2002). In these studies, full network information was collected; here however, we report on studies based on egocentric data. Recently, studies were performed in the area of infectious disease epidemiology with the aim of elucidating the role of social contacts in disease transmission. Until now this work has remained largely disconnected from the social networks literature, instead focussing on providing quantitative input for mathematical modelling of infectious dis-

* Corresponding author. Tel.: +49 521 1063879; fax: +49 521 1062968.
E-mail address: mirjam.kretzschmar@uni-bielefeld.de (M. Kretzschmar).

0378-8733/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.socnet.2007.09.002

ease transmission. In this paper we refer to several concepts developed in social network theory to analyze survey data about social contacts. We are interested in comparing a diary approach which includes individual records for each contact to questionnaires where only aggregate numbers of contacts are reported. After comparing retrospective and prospective study designs, we discuss the problems of recall accuracy and digit preference.

Although conceptually it is clear that direct contacts between individuals play a pivotal role in disease transmission, collecting empirical data which characterizes these contacts in a more quantitative way has been notoriously difficult to do. Edmunds et al. (1997) conducted a study where respondents in a university setting were asked to record the persons they talked with during 1 day. Contact data was obtained using a diary approach during one randomly assigned day, where respondents were asked to keep records of persons with whom they had a two-way conversation. Data was collected from 65 individuals between the ages of 22 and 66. This study demonstrated the feasibility of obtaining data about contact behaviour in this way, even if questions surrounding this methodology remain to be studied (Edmunds et al., 2006). Beutels et al. (2006) investigated whether contact behaviour could be assessed via a web-based interface in a sample of 73 university students. Mikolajczyk et al. (2007) reported on a contact study among 235 school children in a primary school in Germany. Recently, Wallinga et al. (2006) analyzed data from a much larger study performed in a general population in Netherlands and estimated age-dependent contact rates from those data. They also demonstrated that a transmission model for mumps provided a better fit for the seroprevalence data collected in the same study when using the estimated contact rates as compared to contact matrices chosen for mathematical convenience such as proportionate mixing matrices. This shows that collection of contact data and its statistical analysis is a prerequisite for good modelling studies in infectious disease epidemiology.

Contact frequencies and patterns can be studied in various ways. First a distinction has to be made between prospective and retrospective surveys, depending on whether respondents know in advance that they should report on their contacts or whether they are asked to recall the contacts they had during a specified time period. For prospective studies, one can distinguish between observational studies by an observer, e.g. by personal observation, video-taping or electronic sensing (Pentland, 2007) and self-reported studies. In a diary-based approach respondents are asked to fill in a contact diary during the day or at the end of the day. When the diary is filled in during the course of the day, forgetting some contacts is less likely, but clearly some people might not be able to record each contact immediately and thus the advantage of this approach might be lost. In a retrospective study, respondents are asked to fill in a diary for a past time period without being informed in advance. They can be asked to give information in a diary-based approach, where they have to give information on every individual contact, or they can be asked to give aggregated information, for example about the number of contacts they had within specific age groups or in specific situations (Fu, 2005). Prospective studies and questionnaires with more detailed information can be expected to have a higher validity; however, at the same time they are more demanding in terms of necessary resources. While for epidemiological studies, prospective study designs are generally perceived as being superior to retrospective designs, no sufficient information exists for comparing the advantages and disadvantages of those different approaches with respect to the assessment of contact patterns. Additional questions regarding contact patterns relate to the issues of variability in the contact numbers within individuals and whether contacts are repeated contacts with the same person within certain time periods.

We conducted a series of pilot studies for collecting data about contact patterns. The general purpose of this project was to assess the suitability of different approaches for obtaining relevant contact patterns data, with a special focus on comparing prospective and retrospectives study designs. The specific questions we wanted to answer were: (1) How does the number of reported contacts vary across different types of studies? (2) How does the number of reported contacts vary between measurements on different days within the same individuals? (3) What is the perceived ability to remember contacts? (4) What is the validity of reporting contacts in a retrospective study as compared to a prospective design? Here we report on the results of this series of studies and discuss their similarities and differences. Knowledge about the advantages and disadvantages of different survey designs will guide future larger surveys for collecting data to inform mathematical modelling.

2. Methods

The contact surveys were conducted with university students at the University of Bielefeld, Germany, from 2003 to 2006. Each survey was conducted with a specific methodological question in mind. Although the sample sizes are quite small and the population is not representative of a larger general population, the comparison of the different surveys and their methodological differences gives valuable information about how to conduct contact surveys.

2.1. Definitions of contact and questionnaires used in the studies

Two ways to assess contacts were used in these studies: One question asked about aggregate numbers of contacts on six levels of proximity: intimate contacts, close contacts (same household), direct conversation (>2 min duration, max. 2 m distance), small group (with conversations, but less intensive than in direct conversations), larger group (seminary or lecture room) and occasional contacts (in the range of 2 m in local transportation, cinema, etc.). We refer to the data collected in this way as "aggregate contacts". A second question asked for more detailed information about contacts with other persons involving a direct conversation: gender, age or approximate age group of the person in contact, duration of the time in contact and whether contact with this person occurs frequently or not. The latter was cate-