COMMUNICABLE DISEASES AND DATA ANALYSIS

M. Elizabeth Halloran
Department of Biostatistics, Rollins School of Public Health, Emory University, USA

Keywords: Transmission probability, basic reproductive number, vaccines, infectious, secondary attack rate.

Contents

Summary

The special aspect of communicable diseases is the presence of the infectious agent in the host population. The key distinguishing process is transmission of the infectious agent from one host to another. The key relation in infectious diseases is the dependence of infection events among individuals in a population, called dependent happenings by Sir Ronald Ross in 1916. Two key measures in communicable diseases are the transmission probability and the basic reproductive number. Once a host is infected, the natural history of the infectious agent within the host can be described with reference to either infectiousness or disease. The binomial model of transmission is one of the most commonly used approaches for estimating the transmission probability. The secondary attack rate (SAR) is a special case of the transmission probability. Secondary attack rate studies are often conducted by identifying infectious individuals, then recruiting their transmission units or contacts into the study to establish what proportion become infected. It is generally recommended to use the simplest transmission model that captures the essence of the scientific question at hand. Evaluation of indirect effects of an intervention in a population involves comparison of populations or communities, not just individuals.

1. Introduction

The special aspect of communicable diseases is the presence of the infectious agent in the host population. The host population can be humans or other animals. The life cycle of the infectious agent may also include a vector, such as a mosquito or a snail. The infectious agent has a life and agenda of its own. Its presence creates a population-level dynamic that is absent from noncommunicable diseases. Due to the active infectious agent, there are measures of natural history and disease frequency that distinguish the epidemiology and analytic problems in infectious diseases from those of other fields. The key distinguishing process is transmission of the infectious agent from one host to another. The key relation in infectious diseases is the dependence of infection events among individuals in a population, called dependent happenings by Sir Ronald Ross in 1916. That is, how many people become infected at any time depends on how many people are already infected and infectious. Thus, infection events are not independent. One consequence of the dependence of events is that an intervention in one individual can have consequences for the outcomes in other individuals. The indirect effects of interventions and their evaluation are of increasing interest in infectious disease epidemiology.

Two key measures in communicable diseases are the transmission probability and the basic reproductive number. The transmission probability is the probability that, given a contact between an infective source and a susceptible host, successful transfer of the infectious agent will occur so that the susceptible host becomes infected. The basic reproductive number R_0 is the expected number of new infectious hosts that one infectious host will produce during his or her infectious period in a large population that is completely susceptible. Both the transmission probability and the basic reproductive number are measures of the success of an infectious agent in a population. The basic reproductive number is a function of the transmission probability, the duration of infectiousness, and the contact process. The dependent happening relation defines the incidence rate of infection as a function of the prevalence of infectious persons as well as the transmission probability, the contact process, and the contact patterns. Thus, both the incidence rate of infection and the basic reproductive number are functions of the transmission probability and the contact process. However, the incidence rate of infection is from the point of view of the susceptible, and the basic reproductive number is from the point of view of the infectious person.

Once a host is infected, the natural history of the infectious agent within the host can be described with reference to either infectiousness or disease. While the disease process and its associated time line are important to the infected person and to a physician, the dynamics of infectiousness are important for propagation of the infectious agent and for public health. The relation of the two time lines to one another is specific to each infectious agent and can have important implications for study design and public health.

The natural history of infectiousness includes the latent period, the time interval from infection to becoming infectious, and the infectious period, during which time the host could infect another host or vector. Eventually the host becomes noninfectious, either by clearing the infection, possibly developing immunity, or by death. The host can also become noninfectious while still harboring the infectious agent. The host may also

become an infectious carrier if he recovers from disease and becomes asymptomatic, but remains infectious.

In contrast, the natural history of disease in the infected host includes the incubation period, the time from infection to symptomatic disease, and the symptomatic period. The probability of developing symptomatic disease after becoming infected is the pathogenicity of the interaction of the infectious agent with the host. Eventually the host leaves the symptomatic state, either by recovering from the symptoms or by death. An inapparent case or silent infection is a successful infection that does not produce symptoms in the host. Inapparent cases can be infectious.

Historically, early 20th century epidemiology and biostatistics were dominated by applications in infectious diseases. Sir Ronald Ross wrote his paper on dependent happenings and developed the early mathematical models of malaria and the effect of intervention measures. Greenwood and Yule wrote a seminal paper on evaluating vaccination in general, but with the particular case of typhoid and cholera inoculation during the Balkan Wars. Frost developed the concept of the cohort effect in elucidating age and period related changes in tuberculosis disease. Kendrick and Eldering wrote a lengthy paper on the conduct and analysis of pertussis vaccine trials in 1939. In the 1950's attention was focused on the design and implementation of the community-based polio vaccine trials.

However, in the mid-20th century the focus of epidemiology and biostatistics turned towards cancer and heart disease. With the advent of the antibiotic era and the success of numerous vaccines, scientists really believed that infectious diseases were licked. It was just a matter of mopping up the little that was left. Besides which, if there was a magic bullet like penicillin around, fancy statistics were not required to tease out any subtle results. Infectious diseases fell out of favor with biostatisticians, as well as many other people.

By the mid-1980's, the trend reversed. Many infectious agents including those responsible for malaria, tuberculosis, and gonorrhea, were developing drug resistance. The sudden appearance of HIV/AIDS drew many highly qualified statisticians and epidemiologists out of the cancer arena to tackle the analytic problems posed by that infection. Also, the development of new vaccines against important diseases, such as HIV, human papilloma virus, influenza and pertussis began to pose more difficult analytic problems. Advances in molecular epidemiology and the use of microarray chips to study immune responses have opened up whole new scientific horizons.

As a consequence, methodological developments in the study of infectious diseases have received increasing attention. Of particular interest is in understanding the indirect and population-level effects of interventions, not just the direct protective effects. One question is how a particular vaccine might reduce infectiousness and therefore secondary transmission to others. It is possible that a vaccine might not confer good direct protection, but reduce transmission and therefore be an important public health intervention tool. Another question is how immunization of one segment of a population might affect incidence in another segment. For instance, there is considerable interest in what the effect of vaccinating children against influenza would have on the incidence in

adults. These questions, results of the dependence of events in infectious diseases, raise challenging issues of study design and analysis. Estimating the effects of interventions on secondary transmission generally requires information on contacts between infectives and susceptibles. Estimating indirect effects of interventions requires comparison across whole communities. These problems are the current focus of much active methodological research.

2. Transmission probability

One measure of the success of an infectious agent is how effectively it is transmitted. The transmission probability p is the probability that, given a contact between an infective source and a susceptible host, successful transfer of the infectious agent will occur so that the susceptible host becomes infected. The transmission probability depends on characteristics of the infective source, the infectious agent, the susceptible host, and the type of contact. Estimation of the transmission probability and its distribution is important for planning public health interventions, for evaluating interventions, and for understanding the population biology of the infectious agent.

2.1 The binomial model of transmission

The binomial model of transmission is one of the most commonly used approaches for estimating the transmission probability. The basic idea of the binomial model is that exposure to infection occurs in discrete contacts and that each contact is independent of another. If p is the transmission probability during a contact between a susceptible person and an infectious source, then the probability that the susceptible person will not be infected during the contact is $q = 1 - p$. The quantity q is called the escape probability. If a susceptible person makes n contacts with infectious people, then, assuming all contacts are equally infectious, the probability of escaping infection from all of the n contacts is $q^n = (1-p)^n$. The probability of being infected after n contacts with infectives is $1 - q^n = 1 - (1-p)^n$.

To estimate the transmission probability, generally information is needed on contacts between susceptibles and infectives and which of the contacts result in successful infection. Many different approaches can be used to estimate the transmission probability from the binomial model when the appropriate data are available. Most commonly used are maximum likelihood and, more recently, Bayesian Markov chain Monte Carlo methods. The challenges of estimating the transmission probabilities of HIV have pushed the field forward in the past decade. Issues such as interval censoring of infection data, measurement error in exposure to infection data, and heterogeneities have been taken into account. Covariate or intervention effects can be easily included in the model as multiplicative effects on the transmission probability.

The concept of a contact is very broad and must be defined in each particular study. The infectious agent's transmission mode determines what types of contact are potentially infectious. Contacts can be defined between two individuals, or an individual and a vector. More generally, contacts can also be defined within small transmission units, such as households, child care centers, school classes, or retirement homes. Within

small transmission units, mixing is often assumed to be random. A small transmission unit can also be defined as two individuals, such as a steady sexual partnership or a household with just two people. The definition of a contact within a study can depend on the definition of the transmission units. The small transmission unit can also be thought of as a minicohort.

2.2 Contacts with persons of unknown infection status

Ascertaining and counting contacts can be difficult. Even more difficult can be to know the infection status of all of the ascertained contacts. Thus, the infection status of the contacts is often unknown. Under these circumstances, it is not possible to estimate the transmission probability directly. However, a model-based approach can be used to incorporate data on the prevalence of infection in the population along with assumptions about the mixing within the population to model the probability of infection per contact with a person of unknown infection status.

Under the assumption of random mixing, that is randomly making contact with other people in the population, the probability that a person with whom a susceptible makes contact is infectious equals the prevalence of infectious people in the population of contacts, denoted by P Then the probability of being infected from a contact of unknown infection status is $\rho = pP$. The quantity ρ is not a transmission probability in the strict sense, but an infection probability. The probability of escaping infection from contact with someone of unknown infection status is $1-\rho = 1-pP$. Under the binomial model, the probability of becoming infected after n such contacts is $1-(1-pP)^n = 1-(1-\rho)^n$. With external estimates of the prevalence in addition to data on number of contacts and the occurrence of the infection, estimation of the transmission probability can proceed exactly as described above. The estimates of the transmission probability are sensitive to the accuracy of the prevalence estimates. Use of prevalence estimates to aid in estimating transmission probabilities is an area of current research.

-
-
-

Bibliography

N.G. Becker. *Analysis of Infectious Disease Data*. Chapman and Hall, London, 1989. [This book covers elements and definitions of infectious disease data analysis, with a focus on stochastic models and the Reed-Frost and Greenwod models]

K. Dietz. Transmission and control of arbovirus diseases. In D. Ludwig and K.L. Cook, editors, Proceedings of the SIMS Conference on Epidemiology Held at Alta, Utah, 8-12 July 1974. SIAM–SIMS

Conference series, no. 21, pages 104–121, Philadelphia, 1974. Society for Industrial and Applied Mathematics. [This paper covers several of the aspects of the basic reproductive number.]

C.P. Farrington, M.N. Kanaan, and N.J. Gay, Estimation of the basic reproductive number for infectious diseases from age-stratified serological survey data. Appl. Statist. 2001, 50:251-292 (with discussion). [This paper discusses issues related to estimating the basic reproductive number from serologic survey data, particularly the problem of identifiability.]

H. Grosskurth, F. Mosha, J.Todd, et al. Impact of improved treatment of sexually transmitted disease on HIV infection in rural Tanzania: randomised controlled trial. Lancet, 346:530-536, 1995. [This paper is a good example of a group randomized trial to estimate overall effects of an infectious disease intervention.]

M.E. Halloran and I.M. Longini. Using Validation Sets for Outcomes and Exposure to Infection in Vaccine Field Studies. Am J Epidemiol, 2001, 154:391-398. [This paper proposes the use of validation sets to improve the accuracy of vaccine efficacy and effectiveness estimates when only a non-specific outcome or exposure to infection measure is available for most people in the study.]

M.E. Halloran and C.J. Struchiner. Study designs for dependent happenings. Epidemiology, 2:331-338, 1991. [This paper defines the different types of study designs for evaluating direct, indirect, total, and overall effects of infectious disease interventions.]

M.E. Halloran, C.J. Struchiner, and I.M.Longini. Study designs for different efficacy and effectiveness aspects of vaccination. American Journal of Epidemiology, 146:789-803, 1997. [This paper provides a systematic framework of the relation of different efficacy and effectiveness measures and the study design types required to estimate them.]

I.M. Longini and J.S. Koopman. Household and community transmission parameters from final distributions of infections in households. Biometrics, 38(1):115-126, 1982. [This paper is the first in a series that demonstrates the joint estimation of the transmission probability within a household or other small mixing structure and the probability of being infected within the community.]

P.H. Rhodes, M.E. Halloran, and I.M. Longini. Counting process models for differentiating exposure to infection and susceptibility. J R Statist Soc B, 58:751-762, 1996. [This paper gives the technical framework for understanding that the efficacy parameters for direct protection form a hierarchy depending on how much information about the transmission process is used.]

R. Ross. An application of the theory of probabilities to the study of a priori pathometry, Part 1. Proc R Soc Series A, 92:204-230, 1916. [This seminal paper lays down the theory of happenings, dividing them into dependent happenings and independent happenings. Infectious diseases fall into the latter category. In dependent happenings, the number of new events depends on the number of individuals already affected in the population.]

Biographical Sketch

M. Elizabeth Halloran was born in Wooster, Ohio, USA. She obtained her B.S. in General Science from the University of Oregon, Eugene, Oregon, USA, in 1972. She obtained her medical degree from the Freie Universitaet, West Berlin, Germany in 1983. She obtained her M.P.H. in Tropical Public Health in 1985 and her D.Sc. in Population Sciences in 1989 from the Harvard School of Public Health, Boston, Massachusetts, U.S.A. She did a brief post-doctoral fellowship with Lord Robert May in Britain in the summer and fall of 1989, based at Imperial College, London. In December, 1989, Dr. Halloran joined the faculty in Biostatistics at Emory University in Atlanta, Georgia, USA, where she is Professor of Biostatistics. Her interests focus on methodological problems, novel study designs, and new methods of analysis for infectious disease interventions, in particular, vaccine field studies. She is interested in Causal Inference and Bayesian methods.