Epidemiologic Methods for the Study of Infectious Diseases

Edited by

James C. Thomas, M.P.H., Ph.D.
Department of Epidemiology
School of Public Health
University of North Carolina
Chapel Hill, North Carolina

David J. Weber, M.D., M.P.H., M.H.A.
Department of Medicine
School of Medicine
and
Department of Epidemiology
School of Public Health
University of North Carolina
Chapel Hill, North Carolina

OXFORD
UNIVERSITY PRESS
2001

Contents

Part I Foundations

Part II Data Sources and Measurement

5

Overview of Study Design

M. ELIZABETH HALLORAN

Concepts of study design in infectious disease epidemiology have much in common with those in noninfectious disease epidemiology. However, the presence of the infectious agent, separate from but interacting with the human host population, introduces further complexities. Whether a person becomes infected depends on who else in the population is already infected and infectious. Alternatively it may depend on environmental sources of infection. Sir Ronald Ross (Ross 1916) coined the term "dependent happenings" to describe the characteristic of contagious diseases that the number of people becoming newly infected depends on how many are already infected. The transmission of the infectious agent and dependent happenings produce the special aspects of study designs in infectious diseases.

Infectious disease epidemiology encompasses the study of diverse scientific questions:

- Is a disease *communicable*? What infectious organism is causing a disease? What is the mode of transmission? How effectively is the infectious agent transmitted? What are the contact patterns and patterns of spread within the host population? What is the source or reservoir of a point source epidemic?
- What is the *natural history* of infection in individuals? What is the latent period, and the duration and degree of infectiousness? What is the probability of becoming symptomatic? What is the incubation period from acquisition of infection to symptoms? What is the duration of symptoms? What is the probability of dying?
- What are the *population biology*, epidemiology, and dynamics of the infectious agent and any vector? Is the microbe endemic or epidemic? Is an epidemic occurring? Is a disease reemerging? What is the age distribution of infection and disease in the host population? Are there important temporal and spatial aspects of the agent and any vectors? How diverse are genetic variants of the microbe? Is the microbe developing drug resistance or evolving owing to some other pressure? Does transmission intensity influence microbial diversity?

- What are the effects of *covariates or interventions* on infection, disease, and infectiousness? What facilitates infection (risk factors for exposure and susceptibility)? What facilitates disease progression (risk factors)? How can infection and disease be prevented? What is the effect of intervention at the individual level and at the population or community level?

The choice of study design needs to be tailored to the question being asked. Several of the questions listed above are concerned with the etiology and natural history of the infectious agent. The Henle-Koch postulates for evidence that an organism causes a disease are useful (Evans 1976). New techniques in molecular epidemiology permit more accurate tracking of transmission (Glynn et al. 1999, Small et al. 1994), studies of the natural history of the disease, and study of the evolution of microbes (Lipsitch 1997) than before. Several of the questions concern the study of the dynamics and interaction of the host population with the infectious agent population (see Chapter 4).

In this chapter, we focus on the dependent happening relation and its consequences for design and interpretation of studies in infectious disease epidemiology. For coherence of presentation, we also present some definitions and concepts from general epidemiology. In the next section, we present measures of disease frequency, with a focus on the transmission probability. We give a formal expression for the dependent happening expression and show its usefulness. In the third section, we present measures of causal effects and association, with a focus on the transmission probability ratio, and the indirect, total, and overall effects of interventions in populations. We demonstrate the difference between causal effects and association using a formal model for causal inference. We also show how the formal dependent happening relation can aid in interpreting risk ratios. In the fourth section, we present cohort study designs, focusing on those appropriate for estimating the transmission probability and the secondary attack rate. The context is developed as an expanded approach to cohort studies. We also discuss case-control studies. In the fifth section, we consider cross-sectional and community-level studies. In the last section, we touch on aspects of estimation and inference.

MEASURES OF DISEASE FREQUENCY

In this section we define several common measures of disease frequency. Specific to infectious disease epidemiology are the transmission probability, the secondary attack rate, and the basic reproductive number. Common to all fields of epidemiology are the incidence rate, hazard rate, incidence proportion, and prevalence. Because of the phenomenon of dependent happenings in infectious diseases, the common measures of disease frequency have additional intrinsic relations to one another through the underlying transmission process. We give a formal definition of the dependent happening relation as a function of the measures of disease frequency. We discuss how this relation contributes to the design and interpretation of infectious disease studies.

Transmission Probability and Secondary Attack Rate

In Chapter 4, we defined the *transmission probability, p,* as the probability that, given a contact between an infective source and a susceptible host, successful transfer of the microbe will occur so that the susceptible host becomes infected. The transmission probability in a population is

$$p = \frac{A}{n},$$

where n is the total number of contacts made between susceptibles and infectives in a population, and A is the number of infections that occur during those contacts. The transmission probability depends on characteristics of the infective source, the microbe, the susceptible host, and the type of contact.

The *secondary attack rate* (*SAR*) is a special case of the transmission probability. The secondary attack rate is the expected proportion of susceptibles who become infected when exposed to an infectious person. In the secondary attack rate, the contact between the infectious susceptible persons may be defined as occurring over some time period, such as the duration of infectiousness or over the period of the study. For example, the *household SAR* is the probability that a susceptible individual living in the same household with an infectious person during his or her period of infectiousness will become infected (Fine et al. 1988, Orenstein et al. 1988). The secondary attack rate is defined

$$SAR = \frac{A}{M},$$

where M is the total number of susceptible exposed persons and A is the number of persons exposed who develop disease. The SAR is a proportion, not a true rate.

Both the transmission probability and secondary attack rate are defined conditionally on the susceptibles being exposed to infection. Being conditional on exposure to infection distinguishes the transmission probability and secondary attack rate from the general measures of disease frequency such as incidence rate, hazard rate, and incidence proportion presented below.

The probability, ρ, of becoming infected given a contact with a source of unknown infection status is related to the transmission probability p, but it is not strictly a transmission probability. It is an infection probability. Under random mixing, the probability of becoming infected from a contact with a source of unknown infection status is $\rho = pP$, where P is the prevalence of infectious people in the population of contacts.

Incidence rate and hazard rate

The incidence rate, I, of an event in a population is the rate at which the event occurs per unit of person-time at risk. The incidence rate is

$$I = \frac{A}{T},$$

where A is the number of cases observed during a total of T units of person-time at risk. Incidence that varies over time we denote at time t by $I(t)$. If the incidence rate changes in a time interval but is estimated as an average over that interval, the estimate will not reflect the fluctuations that occur within the interval. The hazard rate is the instantaneous probability of an event occurring in a small interval of time. The hazard rate at time t is denoted by $\lambda(t)$. The hazard rate and incidence rate are defined somewhat differently, but both are measures of the probability of an event in an individual in a small unit of time at risk. The term *force of infection* is used in infectious disease epidemiology to denote either the hazard rate or the incidence rate of infection. The incidence rate in infectious diseases can vary rapidly, such as during an epidemic or due to seasonality of the vector population. The rapid changes in incidence rates are a source of some of the challenges of infectious disease epidemiology.

Incidence proportion

The *incidence proportion, R,* is the number of people who experience an event in a closed group of susceptible people over the course of study. The incidence proportion is expressed

$$R = \frac{A}{N},$$

where N is the number of people in the population and A is the number of people who experience the event. We can be explicit that the incidence proportion is measured over a time interval $(0,T)$ by writing $R(T)$. In infectious disease epidemiology, the incidence proportion is often called the attack rate (AR). The *infection* attack rate or incidence proportion is the proportion of the population who become infected. The *disease* attack rate or incidence proportion is the proportion who develop disease.

The mirror image of the incidence proportion is the survival probability, the probability of not experiencing an event in a time interval $(0, T)$. Use of the incidence proportion in the form given here requires that the population be a closed group from the beginning to the end of the study. That is, no one can leave the population. However, analytic methods in survival analysis allow estimation of the probability of experiencing an event in the time interval $(0, T)$ even when some people leave the study.

Prevalence

The *prevalence*, P, is the proportion of a population that has the disease or outcome of interest at a given time. *Seroprevalence*, also denoted P, is the proportion of a population that has a serological marker at a given time. An example is the seroprevalence of immunoglobulin G (IgG) against a specific microbe. Current seroprevalence can reflect either past or current infection, depending on which immune markers are measured. We denote prevalence at time t by $P(t)$. Prevalence of a disease or of infectious people can change rapidly with time, especially during epidemics or due to seasonality of the microbe.

Basic Reproductive Number, R_0

The *basic reproductive number*, R_0, of a microbe in a population is the expected number of new infectives produced by one infective in a large, completely susceptible population during his or her period of infectiousness. For microparasitic diseases, the basic reproductive number is expressed as

$$R_0 = cpd,$$

where c is the number of contacts per unit time, p is the transmission probability, and d is the duration of infectiousness. The basic reproductive number is a measure of the reproductive capacity of a microbe in a particular host population. It is discussed in more detail in Chapter 4.

The Dependent Happening Relation

The key relation in infectious diseases is the dependence of infection events among individuals in a population, called dependent happenings. Under random mixing, the dependent happening relation can be expressed as

$$I(t) = cpP(t), \qquad (1)$$

where $I(t)$ is the incidence rate, c is the constant contact rate, p is the transmission probability, and $P(t)$ is the prevalence of infectious persons at time t. The dependent happening relation means that the incidence rate of infection is dependent on the prevalence of infectious persons. The incidence rate also depends on the contact process and contact patterns, as well as the transmission probability.

The formal expression (1) of the dependent happening relation helps clarify our thinking about several issues. First, in designing and interpreting studies in infectious diseases, it is crucial to distinguish risk factors or interventions related to exposure to infection from those related to susceptibility. The dependent happening expression (1) makes explicit the different components related to the risk of becoming infected. All three factors on the right in relation (1) contribute to exposure to infection. If individuals increase their rate of contact, c, it could increase their exposure to infection. The transmission probability, p, depends on the degree of infectiousness of the contact as well as the type of contact, and so plays a role in determining the level of exposure to infection. The prevalence of infectious people in the population $P(t)$ also helps determine the level of exposure to infection. Behavioral changes aimed at lowering exposure to infection could be aimed at reducing the contact rate, c, altering the transmission probability, p, or reducing the probability that a person makes contact with someone who is infectious. The latter would mean being more selective about with whom one makes contact, with the effect of reducing the prevalence of infectives $P(t)$ in one's contact groups.

Susceptibility of the person at risk to become infected enters into the dependent happening relation primarily through the transmission probability, p. That is, conditional on actually being exposed to a certain level of infection, the susceptibility of the exposed person determines whether the person becomes infected. Although in any given study, the separate components of the dependent happening relation (1) may not be measured, assumptions about the relation of the incidence rate to the contact process, transmission probability, and prevalence are fundamental in designing and interpreting studies.

Second, the dependent happening relation (1) applies in epidemic and rapidly changing situations as well as in stationary situations. It does not rely on the assumption of equilibrium to be valid. Contrast this with another well-known relation from epidemiology that does rely on the assumption of equilibrium incidence rate and prevalence. If D is the average duration of disease, and I is the equilibrium incidence rate of disease, then prevalence approximately equals the product of the incidence rate and average duration (Freeman 1980):

$$P \doteq ID \qquad (2)$$

Relation (2) holds approximately for prevalence less than 0.10. At higher prevalences, the left side would be better represented by the prevalence odds. In Chapter 4 on transmission dynamics, we presented a hypothetical example of gonorrhea in men and women. The dependent happening relation (1) in that example is that incidence of infection in each gender depends on the prevalence of infectious people in the other gender. This does not require that gonorrhea is at equilibrium in the population. In contrast, at equilibrium, expression (2) says that prevalence of infection in each gender depends on the incidence and duration in the same gender. The capital D for duration of disease distinguishes it from the lower case d for duration of infectiousness. The word disease emphasizes that relation (2) is more closely related to the natural history of disease, whereas the dependent happening relation (1) is more closely related to the course of infectiousness, as discussed in Chapter 4.

Third, expression (1) not only demonstrates the relation between the transmission probability and incidence rate as measures of disease frequency but also clarifies their difference. While the transmission probability is defined conditional on exposure to infection, the incidence rate is defined as events per person time. The incidence rate as well as the incidence proportion rely on the notion that the people being studied are potentially exposed to infection, but do not require that any particular individual is actually exposed. Halloran and Struchiner (1995) call the transmission probability a *conditional* measure of disease frequency, while the incidence proportion, incidence rate, and hazard rate are *unconditional* measures. The parameters transmission probability, incidence rate, and incidence proportion form a hierarchy requiring decreasing amounts of information about the transmission and contact processes (Rhodes et al. 1996).

Fourth, the dependent happening relation for the incidence rate $I(t) = cpP(t)$ can be contrasted with the expression for the basic reproductive number, $R_0 = cpd$. Both contain the product of the contact rate and the transmission probability, cp, a fundamental expression of the transmission process. However, the incidence rate reflects the point of view of the susceptible and the probability of becoming infected per time unit. The basic reproductive number, R_0, reflects the point of view of the infectious host as the number of people he or she will infect.

Finally, the dependent happening relation (1) can be used to estimate different quantities, depending on which components have been measured. The product of the contact rate and the transmission probability equals the more easily estimable ratio of the incidence rate to the prevalence of infectives, $cp = I(t)/P(t)$. Thus we do not need to observe the underlying contact process and transmission probabilities to obtain some information about their product cp. The transmission probability can be estimated if the other three components are measured, $p = I(t)/cP(t)$. To estimate c separately from p, however, generally information is needed about the contact process. That is, c and p are not separately identifiable from cp without information on contacts.

MEASURES OF EFFECT AND ASSOCIATION

Evaluating interventions and determining risk factors for infection and disease are important goals of infectious disease epidemiology. Risk differences and risk ratios are formed from the measures of disease frequency discussed earlier. Measures of effect and association based on the transmission probability are specific to infectious diseases epidemiology, while those based on the incidence rates or incidence proportion are common to all fields. Because of the dependent happening relation, interventions in infectious diseases can also have important indirect effects on individuals not receiving the intervention directly. In this section, we focus on the measures of effect and association that are particularly important for infectious disease epidemiology. We begin with a discussion of the difference between causal effects and association.

Causal Effects Versus Association

Suppose we do a study of condom use and its relation to the risk of sexually transmitted infection. We observe that the difference between the proportion of people who contract a sexually transmitted disease in the group using condoms and the group not using condoms is 0.4. We can definitely say we have observed an association between condom use and risk of infection, and quantify the association using the observed risk difference. Can we claim, however, that condom use has a causal effect on reducing risk of infection compared to no condom use? No, we cannot say that condom use is the cause of the reduction in risk without further restrictions. To clarify the difference between association and causal effects, we turn to a formal structure for defining effects of causes.

The approach for defining the effects of causes requires that the effect of a cause be defined relative to another cause. The causes could be different treatments, preventive interventions, or risk factors. In our example, the comparison is between condom use and no condom use. The causal effect of condom use compared to no condom use by an individual is defined as the difference between what the infection outcome would be if the person used condoms and what it would be if the person did not use condoms. This approach to defining causal effects assumes that an individual has some potential outcome for each of the various interventions or treatments under study. The causal effect in an individual is the difference between his or her potential outcomes under the two treatments (Rubin 1978, Holland 1986).

Consider four individuals who are at risk for a sexually transmitted disease. The potential outcomes of the four individuals are listed in Table 5–1. For individual i, let Y_{i0} and Y_{i1} denote the potential infection outcomes under no condom use and condom use, respectively. Then, for any individual i, the individual causal effect of condom use versus no condom use is $Y_{i1} - Y_{i0}$. For subject one in Table 5–1, it is $0 - 1 = -1$, that is, condom use prevents infection in subject 1. For subject 2, the difference in the potential outcomes is $1 - 1 = 0$. That is, there is no causal effect of condom use in subject 2. The person becomes infected in either case.

The fundamental problem of causal inference is that only one of these potential outcomes is observable in any individual, since we can observe the individual only either using condoms or not using condoms. A statistical approach to solving the fundamental problem of causal inference is to define the average causal effect in a population. The average causal effect, C, in the population is the average of the individual causal effects. This, in turn, equals the difference between

Table 5–1 Potential Outcomes under Condom Use or No Condom Use *

	Potential Outcome*	
Subject	Condom Use ($X=1$) (Y_{i1})	No Condom Use ($X=0$) (Y_{i0})
1	0	1
2	1	1
3	0	0
4	0	1

*0,1 denote uninfected and infected, respectively.

the average value of the potential outcomes if everyone received one intervention and the average if everyone received the other intervention. Thus

$$C = E[Y_1 - Y_0] = E[Y_1] - E[Y_0],$$

where E means the average or expected value. In Table 5–1, the average causal effect of condom use compared to no condom use as measured by sexually transmitted infection is $(1 - 3)/4 = -0.50$.

Of course we still cannot observe the potential outcomes of each individual under each intervention. What we can observe is each person's potential outcome under the intervention that he or she actually used. The potential outcomes that we do not observe are called *counterfactual*. We can observe the difference in the average potential outcomes in the people who actually used a particular intervention ($X = 1$) and the average of the potential outcomes in people who did not use the intervention ($X = 0$). We denote this actual observable difference as A, and write

$$A = E[Y_1|X = 1] - E[Y_0|X = 0],$$

where $E[Y_1|X = 1]$ is the average of the potential outcomes in people who actually received $X = 1$, and similarly for $E[Y_0|X = 0]$. In the example above, we observed a difference of 0.4 between risk of infection in the two groups, so $A = 0.4$ in the example. The value of A expresses an association between the intervention and the outcome, but does this association equal the average causal effect in the population C? The answer is, not in general. That is, in general

$$C = E[Y_1] - E[Y_0] \neq E[Y_1|X = 1] - E[Y_0|X = 0] = A,$$

except under certain conditions.

Under two important assumptions, the observable association, A, will equal the average causal effect in the population C. The first assumption is that the potential outcomes in one person are independent of the treatment assignments in the other people. This allows Table 5–1 to have only two columns of potential outcomes for each person, one for each treatment. For instance, the assumption is that subject 1's condom use does not affect the potential outcomes of subject 2. This is sometimes called the *noninterference of units* assumption (Cox 1958). The assumption is obviously violated in many studies in infectious diseases.

The second assumption is that the intervention assignment for each individual is independent of his or her potential outcomes. An example of an independent assignment mechanism is randomization. That is, under randomization, we do not assign people whose potential outcome is infection to use condoms, while assigning people whose potential outcome is no infection not to use condoms. This would obviously bias our measure of effect. Formally, under randomization and the noninterference of units, the causal risk difference equals the observed risk difference,

$$C = E[Y_1] - E[Y_0] = E[Y_1|X = 1] - E[Y_0|X = 0] = A.$$

This statement can be interpreted that in a large population, if half of the people are randomly assigned to each of the two treatments, the difference in the observed average outcome of the two groups would be the same as if it had been possible to observe the entire population under each of the treatments. Assignment mechanisms are usually not random. That is, people decide for their own reasons whether they want to use condoms, and it could very well be associated with their probability of becoming infected. This is a simple formal argument for using randomization to estimate causal effects. It also clarifies the difference between association and causal effects.

Most studies are not randomized. Many studies are *observational*, with the investigator having little influence over the events under study. In many *experimental* or intervention studies, the investigator may have some control over allocation or the interventions or covariates of interest, but does not randomize them. The goal of observational and nonrandomized studies may be to elucidate causal effects, but since they are not randomized, the objective is difficult to achieve. Under these circumstances, the difference in the average outcomes of two groups could be due to something other than the measured risk factor or intervention. An estimated association between the outcomes of interest and the intervention or risk factors of interest could be due to unmeasured confounders.

For instance, people who use condoms may also be very careful about whom they choose for sex partners. Thus people who use condoms may also have a lower exposure to infection. Although we may observe a fourfold decrease in sexually transmitted diseases among people who use condoms compared to those who do not, the reduction may have nothing to do with condom use itself. Thus it would be incorrect to conclude that use of condoms has a causal effect on reducing sexually transmitted diseases. It would be correct to say that there is an observed association. Causal inference will generally rely on some untestable assumptions.

There are many sources of bias in observational studies. Ascertainment or selection biases result in the actual study population not being representative of the population that was targeted to be studied. Ascertainment bias can be important in infectious disease studies that ascertain transmission units through an infectious person, such as an index case. The types of infected persons so ascertained might not represent the infected population. Also, larger transmission units in a population would tend to be ascertained more often than smaller transmission units because they have more people in them. Other sources of biases and potential confounders are covered in detail in Rothman and Greenland (1998) as well as many other texts.

It is important to measure possible confounders in nonrandomized studies to take them into account in the analysis. However, it is difficult to measure all confounders. *Sensitivity analyses* can be used to quantify potential *hidden biases* due to unobserved covariates (Rosenbaum 1995, Robins et al. 1999). The point of departure for sensitivity analyses is quite often the paradigm of the randomized study and causal inference described previously. When the goal is to make causal statements based on observational studies, sensitivity analyses can provide some measure of uncertainty about the bias and how large the bias would need to be to swamp out the observed association. Although randomization helps in interpreting study results as causal effects, it does not control for all confounding of estimates. The interested reader can find more on this topic in Greenland and Robins (1986), Greenland (1987), Greenland et al. (1999), Gail (1986, 1988), and Gail et al. (1984, 1988).

Despite its general usefulness, the potential outcome approach to causal effects encounters difficulty when applied to dependent happenings, such as in infectious diseases. The assumption commonly made when using the potential outcome paradigm is that the potential outcomes in any individual are independent of the treatment assignments in other individuals. This is not true for many of the situations in infectious diseases.

Suppose a person is vaccinated and does not become infected, but if he had not been vaccinated, he would have become infected and infected another person. This second person's infection outcome is dependent on the intervention assignment of the first person. Although the assumption that the potential outcome in one person does not depend on the treatment assignment in another person is not conceptually necessary for this approach to causal inference (Rubin 1978, Rubin 1990), the problems arising when the assumption is violated have not been solved.

For example, in Table 5–1, related to potential outcomes for two treatments, the two columns need to be expanded for each individual to include all the treatment and outcome possibilities of people with whom he or she may make potentially infectious contact. However, this is not generally feasible. Quite simply put, because of the indirect effects in infectious diseases, the population causal rate ratio of receiving an intervention compared to not receiving the intervention does not necessarily equal the observed rate ratio. Another option is to condition on exposure to infection, as in the transmission probability. This solution runs into other problems. Studies that challenge humans with inoculation by the microbe are unethical if they pose more than a minimal risk, so exposure to infection can, in general, not be randomized. Halloran and Struchiner (1995) discuss in detail the problem of using the potential outcome approach to causal inference in infectious diseases. Although solutions are still being sought for applying the approach to dependent happenings, the paradigm is increasingly being used to study causal effects, association, nonadherence, and confounding. Infectious disease epidemiologists need to be familiar with its strengths and its shortcomings.

Measures of Effect and Association

Commonly, the same ratio and difference measures are used for estimating both causal effects and associations. Their interpretation is simply different. For simplicity of presentation, we generally use the term *effect* in the following discussion. In this section we present an overview of many of the commonly used risk ratios. Table 5–2 contains a summary of some important relative risk measures in infectious disease epidemiology by choice of comparison group and level of information required. In the top row are relative risk measures based on the transmission probability. These measures are specific to infectious disease epidemiology. They estimate the relative susceptibility and infectiousness associated with risk factors or covariates conditional on exposure to infection. In the bottom part of the first column are the unconditional relative risk measures based on the incidence rate, hazard rate, and incidence proportion. These relative risk measures are not specific to infectious disease epidemiology. The unconditional relative risk measures estimate either relative exposure to infection or susceptibility depending on the design of the study and assumptions regarding exposure to infection. In the bottom right portion of the table are measures of community level relative risk in which the comparison groups are transmission dynamically separate populations. They include the indirect, total, and overall effects of intervention. The indirect effects of intervention are important in the dependent happening situation.

Transmission Probability Ratio

The transmission probability ratio, *TPR*, is a measure of the relative risk of transmission to susceptibles between different pairs of risk factors in infectives during a contact. For any given type of contact and infectious agent, we can estimate the effect of a covariate on susceptibility, infectiousness, or their combination by our choice of comparison pairs in the transmission probability ratio. We may want to compare the male-to-male, male-to-female, female-to-male, and female-to-female transmission probabilities of gonorrhea. We may want to know how transmission of influenza between children compares to that between adults, or also between children and adults. We may want to compare the ability of two types of mosquitoes to transmit malaria to humans. The goal of a study might be to estimate the effect of vaccination on reducing susceptibility and infectiousness as measured by the secondary attack rate. We can also estimate the transmission probability of differing types of contacts, infectious agents, routes of infection, or strains of an infectious agent. For instance, one clade (i.e., strain) of HIV may be more transmissible than another.

Suppose that there are two types of infectives and susceptibles making a specified type of contact for a given type or strain of microbe. We denote the two risk levels as 0 and 1. The risk factors might be vaccinated and unvaccinated, for example. Then there are four different possible combinations of the risk factors in the transmission probability. If the first subscript denotes the infectious person and the second denotes the susceptible in the contact, then the four transmission probabilities are p_{00}, p_{01}, p_{10}, p_{11}. For instance, p_{10} denotes the transmission probability of an infective with risk factor level 1 to a susceptible with risk factor level 0. The relative susceptibility as measured by the transmission probability ratio, TPR_S, is

Table 5–2 Various Measures of Relative Risk.

		Comparison Groups and Effect			
Level	Parameter Choice	Susceptibility	Infectiousness	Combined Change in Susceptibility and Infectiousness	
A. Parameter conditional on exposure to infection					
I	Transmission probability, p Secondary attack rate (SAR)	$\text{TPR}_S = \frac{p_{01}}{p_{00}}$	$\text{TPR}_I = \frac{p_{10}}{p_{00}}$	$\text{TPR}_T = \frac{p_{11}}{p_{00}}$	
B. Parameter not conditional on exposure to infection					
		Study Design			
		I Direct	IIA Indirect	IIB Total	III Overall
II	Incidence rate (I)	$\text{IR}_\text{I} = \frac{I_{A1}}{I_{A0}}$	$\text{IR}_\text{IIA} = \frac{I_{A0}}{I_{B0}}$	$\text{IR}_\text{IIB} = \frac{I_{A1}}{I_{B0}}$	$\text{IR}_\text{III} = \frac{I_A}{I_B}$
	Hazard (λ)	$\text{HR}_\text{I} = \frac{\lambda_{A1}}{\lambda_{A0}}$	$\text{HR}_\text{IIA} = \frac{\lambda_{A0}}{\lambda_{B0}}$	$\text{HR}_\text{IIB} = \frac{\lambda_{A1}}{\lambda_{B0}}$	$\text{HR}_\text{III} = \frac{\lambda_A}{\lambda_B}$
III	Proportional hazards (PH)	$\text{HR}_\text{PH} = e^{\beta_1}$	NA	NA	NA
IV	Incidence proportion (R) Attack rates (AR)	$\text{RR}_\text{I} = \frac{R_{A1}}{R_{A0}}$	$\text{RR}_\text{IIA} = \frac{R_{A0}}{R_{B0}}$	$\text{RR}_\text{IIB} = \frac{R_{A1}}{R_{B0}}$	$\text{RR}_\text{III} = \frac{R_A}{R_B}$

Adapted from Halloran et al., Am J Epidemiol 146:789–803, 1997.

The subscripts 0 and 1 describe two levels of risk. The subscripts *S*, *I* and *T* denote susceptibility, infectiousness, and combined effects, respectively. The Cox proportional hazards estimator is denoted by e^{β_1}. Time has been omitted from the table for notational clarity.

measured by comparing the transmission probabilities to susceptibles with different covariates from infectives. The relative infectiousness, TPR_I, of infected people with the two covariate levels is measured by comparing the transmission probabilities from infectives with different covariate levels to susceptibles. To measure the combined effect of the covariates, TPR_T, the transmission probability between people who are both covariate level 1 is compared to that between people in which both are covariate level 0. The transmission probability ratios are:

$$\text{relative susceptibility: } TPR_S = \frac{p_{i1}}{p_{i0}};$$

$$\text{relative infectiousness: } TPR_I = \frac{p_{1j}}{p_{0j}}; \qquad (3)$$

$$\text{and combined effect: } \text{TPR}_\text{T} = \frac{p_{11}}{p_{00}}.$$

The transmission probability ratios are in the top row of Table 5–2. In general, there could be several levels of covariates, with p_{ij} denoting the transmission probability from an infective with covariate status *i* to a susceptible with covariate status *j*.

One of the covariates might be considered a control or baseline value. Table 5–3 presents an example from a measles vaccine study of the secondary attack rates from vaccinated or unvaccinated index cases to vaccinated or unvaccinated susceptibles. Recall that the secondary attack rate is a special case of a transmission probability, so we can use the SAR_{ij} in place of the p_{ij} in the expressions for the transmission probability ratios. We use the subscripts 0 and 1 to denote unvaccinated and vaccinated, respectively. As an example, consider the data in Table 5–3. The secondary attack rate if both the index case and the exposed children are unvaccinated is $SAR_{00} = 0.38$. If both are vaccinated,

Table 5–3 Secondary Attack Rates by Vaccination Status of the Index Child and the Vaccination Status of the Exposed Children in a Measles Epidemic in Senegal, 1994–1995

	Secondary Attack Rate		
Index Case	Vaccinated, Exposed Children	Unvaccinated, Exposed Children	All Children
Vaccinated	6/83 (0.07)	3/17 (0.18)	9/100 (0.09)
Unvaccinated	41/374 (0.11)	47/124 (0.38)	88/498 (0.18)
Total	47/457 (0.10)	50/141 (0.35)	97/598 (0.16)

From Cisse et al. 1999.

$SAR_{11} = 0.07$. If we calculate TPR_S separately for children exposed to unvaccinated or vaccinated index cases, the estimates are $TPR_S = SAR_{01}/SAR_{00} = 0.11/0.38 = 0.29$ and $TPR_S = SAR_{11}/SAR_{10} = 0.07/0.18 = 0.39$, respectively. Without stratifying on infective vaccination status, the effect of vaccination on susceptibility is estimated as $TPR_S = SAR_{.1}/SAR_{.0} = 0.10/0.35 = 0.29$. The dot in the subscript indicates summation over both the 0 and the 1 strata. The interpretation is that the average transmission probability to vaccinated children is 0.29 that of the transmission probability to unvaccinated children. The analogous calculations for the effect of vaccination on infectiousness are $TPR_I = SAR_{10} / SAR_{00} = 0.07/0.11 = 0.64$, $TPR_I = SAR_{11} \neq SAR_{01} = 0.18/0.38 = 0.47$, and $TPR_I = SAR_{1.} \neq SAR_{0.} = 0.09/0.18 = 0.50$, respectively. The vaccine seems to have a stronger effect on susceptibility than on infectiousness. The ratio of the transmission probability if both the index case and the exposed children are vaccinated compared to if both are unvaccinated is $TPR_T = SAR_{11} / SAR_{00} = 0.07/0.38 = 0.18$.

The corresponding vaccine efficacies based on the transmission probability ratios can be calculated from $VE_S = 1 - TPR_S$, $VE_I = 1 - TPR_I$, and $VE_T = 1 - TPR_T$ (Halloran et al. 1997). In this case, the average vaccine efficacy for susceptibility is $VE_S = 1 - 0.29 = 0.71$, the average vaccine efficacy for infectiousness is $VE_I = 1 - 0.50 = 0.50$, and the efficacy if both are vaccinated compared to if neither are vaccinated is $VE_T = 1 - 0.18 = 0.82$.

A slightly different approach to the TPR_S can be used in the binomial models described in Chapter 4. Assume that the effect of the covariates on infectiousness and susceptibility are multiplicative on the transmission probability, and that the two effects are independent. Denote the relative susceptibility of risk factor level 1 to 0 by θ, so that $TPR_S = \theta$, and the relative infectiousness of level 1 to 0 by ϕ, so that $TPR_I = \phi$. By the assumption that the two effects are independent, then $TPR_T = \theta\phi$. Assume that p_{00} is the baseline transmission probability, denoted simply by p. The transmission probability between an infective of covariate status μ and a susceptible of covariate status ν is written $\theta^{\nu} \phi^{\mu} p$. For example, if both people in the contact are of covariate status 0, this reduces simply to $p = p_{00}$. If the infectious person has covariate status $\mu = 1$ and the susceptible person has covariate status $\nu = 0$, then the expression reduces to $\phi p = p_{10}$. A simple extension of the binomial model to include covariates is to insert the appropriate expression for the transmission probability for each contact observed. The expression can be solved using numerical methods for the estimates of p, θ, and ϕ to obtain the desired TPR_S. Other more complex models for estimating transmission probability ratios are mentioned in the Study Designs section.

Incidence and Hazard Rate Ratios

Consider the situation that there is just one covariate with two levels, denoted by 0 and 1. The incidence rate ratio at time t is

$$IR(t) = \frac{I_1(t)}{I_0(t)},$$

where $I_0(t)$ and $I_1(t)$ are the incidence rates in the two covariate groups. The hazard rate ratio is $\lambda_1(t)/\lambda_0(t)$. If the hazard rate ratio in the two groups is constant over time, the proportional hazards model is said to hold (Cox 1972). The proportional hazard ratio is often denoted e^{β}, where β is the estimated parameter. In the proportional hazards model, the baseline hazard rate in the two groups cancels out and does not need to be estimated. The incidence rate ratio and hazard rate ratio do not condition on exposure to infection. They are not specific to infectious diseases. In Table 5–2 the incidence rate ratio and hazard rate ratio are the second and third row of parameters. The fourth row contains the proportional hazard parameter, but only under the column for direct effects. Vaccine efficacy estimated by the incidence rate ratio is $VE_{IR}(t) = 1 - IR(t)$. Vaccine efficacy can also be estimated from the hazard rate ratio.

Relative Incidence Proportion

Assume again that there is just one covariate with two levels, denoted by 0 and 1. The incidence proportion ratio at time T in a study that goes from time $(0, T)$ is

$$RR(T) = \frac{R_1(T)}{R_0(T)},$$

where $R_0(T)$ and $R_1(T)$ are the incidence proportions up to time T in the two covariate groups. The bottom row in Table 5–2 contains the incidence proportion ratio. The incidence proportion ratio does not require information on exposure to infection and is not specific to infectious disease epidemiology. It is sometimes called the attack rate ratio. Vaccine efficacy can be estimated from $VE_{RR}(T) = 1 - RR(T)$.

Conditional Versus Unconditional Relative Risks Measures

The relative risk measures require differing levels of information for their estimation. The greatest difference is between the conditional parameters, such as the transmission probability ratio, and the unconditional parameters, such as the incidence rate ratio and the incidence proportion ratio. To estimate the *TPR*, information on contacts between susceptibles and infectives and knowledge of infection events is generally required. The transmission probability ratios are specific to infectious diseases. For estimation of the incidence rate ratio, the time at which each event occurs and the time at potential risk are required. Similar time-to-event data are needed to estimate the relative hazard rates. For the incidence proportion ratio, only information on whether an event occurs by the end of the study is required. Thus the ordering of the rows in Table 5–2 corresponds to a hierarchy of information needed for estimating the relative risks (Rhodes et al. 1996).

In designing a study, a choice needs to be made about which relative risk measure will be used in the analysis to help determine what data to collect. The primary choice is between using the transmission probability ratio or one of the unconditional measures, such as the incidence rate ratio. We can use the dependent happening relation (1) to clarify some of the implications of using the incidence rate ratio. Analogous arguments would apply to the hazard rate ratio and the incidence proportion ratio. We expand the dependent happening relation (1) to include two covariate groups. We let $I_1(t)$ in covariate group 1 be the product of the contact rate c_1, the transmission probability from an average infectious person with whom they make contact, $p_{.1}$, and the prevalence of infection in those people with whom they make contact, $P^1(t)$. The index is in the superscript to indicate it is the prevalence in those people with whom people in covariate group 1 make contact. This might not be covariate group 1 itself. Similarly, the incidence rate $I_0(t)$ in covariate group 0 is the product of the contact rate c_0, the transmission probability $p_{.0}$, and the prevalence in their contacts, $P^0(t)$. The incidence rate ratio can then be expressed

$$IR(t) = \frac{I_1(t)}{I_0(t)} = \frac{c_1\, p_{\cdot 1}\, P^1(t)}{c_0\, p_{\cdot 0}\, P^0(t)}. \qquad (4)$$

Although the incidence rate ratio compares the incidence rates in two groups of susceptibles, its interpretation is not limited to being a measure of the relative susceptibility of the two groups. In expression (4), the incidence rate ratio could differ from 1 for a variety of reasons. The contact rates, c_0 and c_1, of the comparison groups could differ. The transmission probabilities, $p_{\cdot 0}$ and $p_{\cdot 1}$, could differ either because the susceptibility of the comparison groups differs, the groups make different types of contacts, or they make contacts with infective people of differing infectiousness. The proportion of contacts the groups make with infective people $P_0(t)$ and $P_1(t)$ could differ because they circulate in differing subpopulations. For example, we might observe that the incidence rate of yellow fever is three times higher in men than in women. The higher incidence rate in men could result from: a higher contact rate with the mosquito vector for yellow fever; men may be more susceptible to developing yellow fever when exposed; or men may spend time in areas where a higher proportion of the yellow fever vector mosquitoes are infected.

Consider again a study of the effect of condoms on sexually transmitted infection. Assume that condom use reduces infectiousness, so that the transmission probability to people whose partners use condoms is $p_1 = 0.25p_{0\cdot}$. Assume we conduct a study in which we do measure contacts of the study subjects with infectives. Assume that there are 100 contacts between infectives in each group. In the group using condoms, four people become infected, while in the other group, 16 people become infected. Then we estimate that $p_{1\cdot} = 0.04$, $p_{0\cdot} = 0.16$, and that $TPR_I = 0.25$. Suppose that instead of collecting information on contacts with infectives, we collect only time-of-event data and person-time-at-risk data, and use the incidence rate ratio. If the study is randomized and people do not change their behavior after randomization except to use condoms, the contact rates, prevalence of infection, and infectiousness in the sexual partners of the two groups might be equal. Then

$$IR(t) = \frac{I_1(t)}{I_0(t)} = \frac{c_1\, p_{1\cdot}\, P^1(t)}{c_0\, p_{0\cdot}\, P^0(t)} = \frac{p_{1\cdot}}{p_{0\cdot}} = 0.25.$$

In this simple case, we would get a similar estimate from both the TPR_I and the IR.

Suppose that the study is observational and that people using condoms have a three times higher contact rate than people not using condoms, $c_1 = 3c_0$. However, we do not collect information on the relative contact rates in the groups. Then the expected estimate of $IR(t)$ would be

$$IR(t) = \frac{I_1(t)}{I_0(t)} = \frac{c_1\, p_{1\cdot} P^1(t)}{c_0\, p_{0\cdot} P^0(t)} = \frac{3c_0(0.25p_0)}{c_0\, p_{0\cdot}} = 0.75.$$

Interpretation of the estimate 0.75 would be difficult without further information. If we falsely assumed that the behavior of the two groups was similar, then the results suggest that condom use reduces transmission by less than a factor of two, rather than by a factor of four.

If the contact rate were the same in the two groups, but people who asked their partners to use condoms chose their sexual partners from a partner pool in which prevalence was five times higher, then $P^1(t) = 5P^0(t)$. The expected estimate would be

$$IR(t) = \frac{I_1(t)}{I_j(t)} = \frac{c_1\, p_{1\cdot} P^1(t)}{c_0\, p_{0\cdot} P^0(t)} = \frac{(0.25p_0)5p^0(t)}{p_{0\cdot} P^0(t)} = 1.25.$$

It would appear that condom use actually increases incidence. Thus there could be a difference in the incidence rates of two groups for a variety of reasons.

Under what circumstances could we interpret a difference in the incidence rates in two groups as due to a difference in susceptibility? First, the risk factor (e.g., age) or intervention (e.g., vaccination) in question would have to be associated with the study individual's susceptibility, not exposure to infection. Second, the exposure to infection in the comparison groups would have to be the same. Under randomization, and assuming nothing changed postrandomization, we would expect exposure to infection to be equal in the two groups. Since several factors contribute to exposure to infection, it may be that not each of the factors is the same in each group, but the overall exposure to infection is the same. If exposure to infection in the groups is not the same, however, just as with any confounder, stratifying on a surrogate measure for exposure to infection can improve the estimates of the effects on susceptibility. To stratify by surrogates or risk factors for exposure to infection is not the same as conditioning on actual contacts with infectives.

In designing and interpreting studies, it is important to distinguish risk factors for exposure to infection from risk factors for susceptibility. For any particular risk factor or intervention, one must give thought to the component of the dependent happening expression to which it corresponds. It should then be clear whether the risk factor corresponds to exposure to infection or to susceptibility. Behavioral interventions could affect the contact rate, the transmission probability, or the probability that a given contact is infectious (Halloran et al. 1994). Randomization can help interpretation of the results. With randomization and masking, on the average, the comparison groups should be similar in the absence of intervention. Although estimating the transmission probability ratio requires more information than the unconditional measures, it has clear advantages. Estimates of the transmission probability ratio can be more directly interpreted as evaluating relative infectiousness and susceptibility (Koopman et al. 1991). In fact, estimation of the relative infectiousness is generally not possible except by using the transmission probability ratio. Also, by controlling for contacts between susceptibles and infectives and exposure to infection, the transmission probability ratio is more robust than the unconditional measures to deviations from randomization.

Population and Community Level Relative Risks

Because of the dependent happenings in infectious diseases, interventions often have effects not only on the people receiving the intervention but also on people not receiving the intervention. The indirect effects are defined not only with regard to a kind of intervention, such as vaccination, but for the allocation of the intervention in the entire population. Although outcomes will still be measured on individuals, evaluation of indirect effects of an intervention in a population involves comparison of populations or communities, not just individuals. The primary unit of analysis and inference is the population.

We define three different types of effects at the population level (Fig. 5–1). Indirect effects are benefits, or detriments, from an intervention program in a population to individuals not directly receiving the intervention, compared to their hypothetical experience if their population had not had the intervention program. Total effects are the combined direct effect in individuals actually receiving the intervention and the benefits due to the indirect effects of the intervention program as a whole. The overall effect of an intervention program is the effect on the population as a whole, including both those receiving and those not receiving the intervention.

Vaccination programs are a common example in which individuals receive the intervention but its widespread application can have indirect effects on those who were not vaccinated. The indirect effects in the unvaccinated people may be different from those in the vaccinated people, which is why we define both indirect and total effects. For example, the average age of first infection may be shifted in both the vaccinated and the unvaccinated people. However, it may be shifted even more in the vaccinated people because of the protection directly conferred

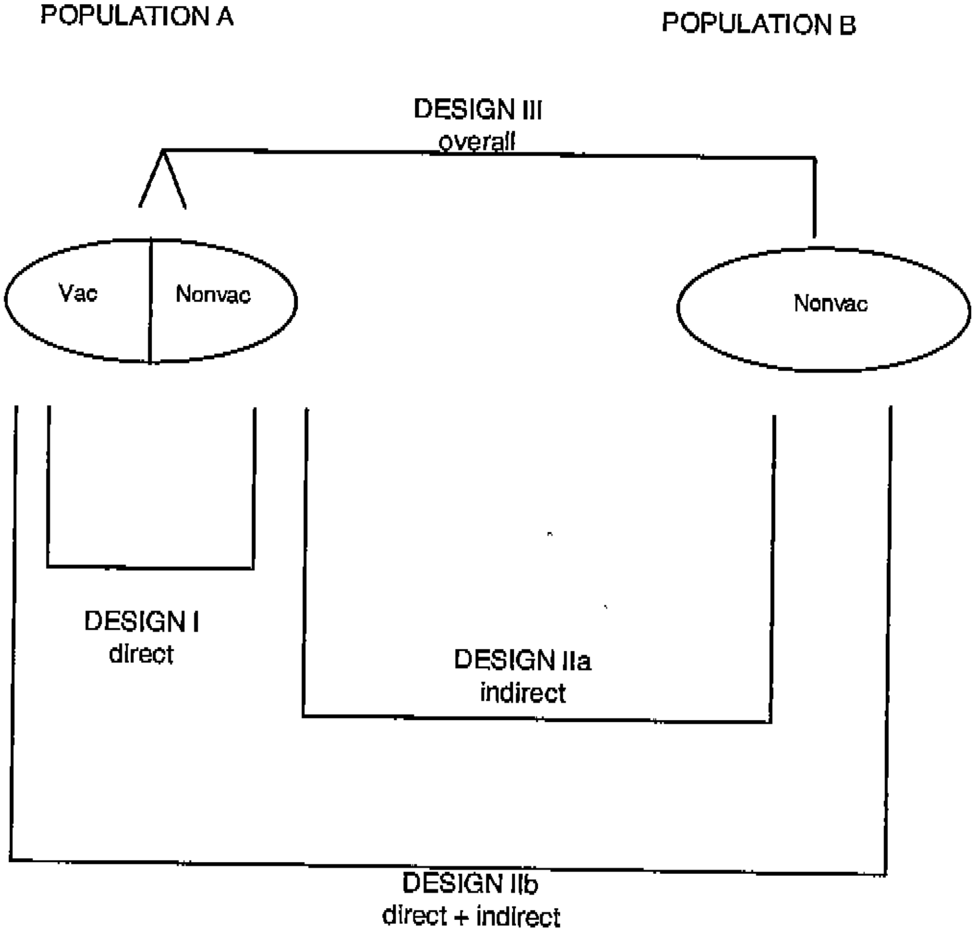

Figure 5–1. Types of effects of interventions against infectious disease and different study designs based on comparison populations for their evaluation.
Source: Halloran and Struchiner 1991.

by the vaccine. As an example of estimating overall effects, Hayes and colleagues (1995) studied the effect of improved treatment of sexually transmitted disease on HIV infection in rural Tanzania with a community randomized controlled trial. Bed net studies for protecting against malaria infection can also be evaluated for their effect on population level incidence. Some interventions are not applied at the level of the individual and have only overall effects. For example, draining mosquito breeding sites is intended to reduce transmission of malaria by reducing the abundance of mosquitoes. Introducing wells for obtaining water is supposed to reduce Guinea worm infection (dracunculiasis).

In Figure 5–1, the assumption is that some individuals in population A received the intervention program and population B did not receive the intervention program. The different kinds of effects are measured by comparing different subpopulations in population A to population B. For indirect effects, individuals in population A not receiving the intervention are compared to individuals in population B, all of whom did not receive the intervention. To measure total effects, the subpopulation in A composed of individuals receiving the intervention is compared to population B. For the overall public health benefits, the average outcome in the entire population A is compared to that in B. The comparison of different subgroups from population A to population B are designated study designs IIA, IIB, and III, respectively They are called the study designs for dependent happenings (Struchiner et al. 1990, Halloran and Struchiner 1991,1995). Different strata within the subpopulations such as age groups or gender can also be compared.

The bottom right portion of Table 5–2 contains examples of possible comparisons using the unconditional estimators to estimate the indirect, total, and overall effects. For example, if the incidence rate ratio is used to measure the indirect effects, then $IR_{IIA}(t) = I_{A0}(t)/I_{B0}(t)$, where $A0$ and $B0$ denote those not receiving the intervention in A and B, respectively. The total effects are estimated from $IR_{IIB}(t) = I_{A1}(t)/I_{B0}(t)$. The proportional hazards parameter is not included in that portion of the table, because the assumption that the baseline hazard rate in the comparison groups is equal would presumably be violated. A change in the basic reproductive number R_0 or shift in age distribution could also be used for comparison of the overall or indirect effects.

Study designs of type I measure those direct effects discussed previously with unconditional parameters and compare people receiving the intervention with people not receiving the intervention within the same population. One important difference to noninfectious disease epidemiology can be made explicit here. In noninfectious diseases, the prevented fraction in a covariate group is generally measured by comparing the incidence proportion in a group with a particular covariate value with that in a group with another covariate value in the same population. The actual number of prevented cases can be calculated by knowing what fraction of the population has the covariate of interest and the relative difference in the two incidence proportions. However, in infectious disease, because of the indirect effects of intervention, this would not be the appropriate comparison. For example, if many people were vaccinated, then the incidence proportion would be lower in both the vaccinated and the unvaccinated groups than it would have been without vaccination. The number of prevented cases is larger than would be estimated by the comparison used in noninfectious diseases. The appropriate comparison would use a study design IIB. However, generally, use of study design IIB is not possible to estimate the prevented number of cases. If study design type I is used, then the estimated prevented fraction will generally underestimate the total number of cases prevented by the covariate or intervention. This should be kept in mind when interpreting the results.

STUDY DESIGNS

Many study designs are the same for infectious diseases as for noninfectious diseases. Here we focus on concepts relating to these study designs that are specific to infectious diseases, such as the interaction of the study cohorts with the population at large, designs that allow estimation of the transmission probability ratio, and the role of assumptions about population mixing structures in the use and interpretation of data. Other books such as Rothman and Greenland (1998) describe principles of study designs to estimate the unconditional relative risks.

Cohort Studies

In cohort studies, usually the investigator identifies a group of disease-free people and follows them over time to see how their disease outcome depends on varying levels of risk factors or interventions. The question of interest is how the different levels of risk factor affect the time to onset of infection or disease or whether they are infected or not by the end of the study. Cohort studies are generally used to estimate the unconditional relative risks such as incidence rate ratio, hazard rate ratio, or incidence proportion ratio. With information on contacts between infectives and susceptibles, cohort studies can also be used to estimate the transmission probability ratio. The unit composed of the susceptible, the infective, and the contact between them is the irreducible element in the study of transmission. In the infectious disease setting, the cohort may be composed of small transmission units, such as partnerships, households, or schools. The small transmission units can be considered independent of one another and analyzed as minicohorts. Alternatively, the small transmission units may be thought of as embedded within a larger community in which members of the small units mix with one another. The contact patterns of the cohort members either with one another or with the population at large can influence the transmission dynamics and the interpretation of the study results. In this section, we present some of these different study designs.

A *fixed* cohort is assembled at one time and followed, with no new additions to the cohort. If a fixed cohort does not lose people during follow up, it is a *closed* cohort. A closed population in an epidemic is a closed cohort. If people leave a cohort before experiencing an event or they do not experience an event before the end of a study, their event times are not observed. Their event times are said to be *right censored. Dynamic* cohorts can have people entering and leaving the risk set during the period of observation. Natural history studies describe the course of infection, disease, mortality, and infectiousness within the human host. *Longitudinal* studies in which several readings are obtained on the same individual are quite common. Since the observations within an individual are correlated, longitudinal studies usually require special methods of analysis (Diggle et al. 1994). *Baseline studies* observe cohorts before an intervention to learn about the feasibility of doing an intervention and to estimate preintervention incidence of infection. For example, a baseline study will yield information on retention rates and how large the study population needs to be.

The cohort may be composed of people who are infected, with the goal being to estimate disease progression or infectiousness based on parasite shedding. Ideally, people will enter the study cohort at the time they are infected. However, sometimes study cohorts are recruited among people who were infected before they entered the study. A cohort of people recruited after they were infected is called a *prevalent* cohort. At the beginning of the HIV epidemic, of necessity, cohorts of already infected people were assembled. Brookmeyer and Gail (1987) discuss biases in risk ratios estimated in prevalent cohorts.

Cohort within a population

We can assemble either a fixed or a dynamic cohort of susceptibles within a larger dynamic population and follow them as in a usual cohort study. The transmission process in the population at large outside the cohort under study can affect incidence within the cohort. If we enroll a cohort of susceptibles and follow them as they become infected, we will observe an epidemic within the cohort. If the contacts made by the cohort members are made at random predominantly with people outside the cohort, the prevalence of infection in the contacts will be similar to the prevalence in the population at large. Prevalence in the population at large may be changing rapidly over time or it may be fairly constant.

Suppose data are collected on the infection status of the cohort members and the number of contacts made by them. Assume also that an estimate of the prevalence, P, of infection in the pool of potential contacts is available, either from the study itself or from other sources. The expression for the infection probability from a contact with unknown infection status can be used to estimate the transmission probability. The probability of infection after n total contacts is $\rho_n = 1 - (1 - Pp)^n$. In this case, the actual infection status of the contacts does not need to be known to estimate the per contact infection probability with someone of unknown infection status. If some estimate of prevalence P is available, then the expression can be used to estimate the transmission probability p. In one study, Hooper and colleagues (1978) used the number of contacts made by men, the number of men who became infected, and an estimate of gonorrhea prevalence in sex workers to estimate the transmission probability of gonorrhea per sex act from women to men. Hudgens and associates (2001) used an estimate of

HIV prevalence in injection drug users, the number of needle sharing acts and the number of injections, and the number of infections to estimate the transmission probability of HIV per needle sharing. If the contacts are predominantly with other members of the initially susceptible study cohort, then in the early phase of the study there will be few infectious contacts. However, the number of infectious contacts will increase as the epidemic within the study cohort spreads as in an epidemic process. If contacts are made with other members of the cohort, then it is likely the infection status will be known.

Transmission probability and contact studies

To estimate the transmission probability or the transmission probability ratio we generally need information on contacts between susceptibles and infectives. The concept of a *contact* is very broad and must be defined in each particular study. The microbe'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 susceptible 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.

Different definitions of a potentially infective contact and transmission unit are possible for the same microbe, and even within the same study. In a study of chickenpox transmission, a potentially infective contact could be defined as being in the same school on one day with someone with chickenpox. Alternatively, it could be defined as living in the same house during the presumed infectious period of the person with chickenpox. In the first case, the transmission unit is the school, and in the latter, it is the household. In the first case, the contact is defined over one day, and in the latter, it is defined over the entire infectious period. In tuberculosis, a contact could be defined as riding on the same bus with someone with open tuberculosis, or as being in the same prison with someone with tuberculosis. In the former case, the transmission unit is the bus, and in the latter, it is the prison.

There could be different definitions of a contact for one definition of transmission unit. In an HIV study, a potentially infective contact could be defined as each sex act between two sexual partners in a steady relationship, one of whom is infected with HIV. Alternatively, the partnership over its entire duration or over the duration of the study could be defined as one potentially infective contact.

In Chapter 4, we discussed transmission units in the contexts of chain binomial models and of nonrandom mixing. Here we further discuss the implications of thinking about transmission units and contacts within populations for study design and analysis.

Identification of infectives

In one approach to ascertaining transmission units or contacts, infectious individuals are identified, then their transmission units or contacts are identified. The initially identified infectious person in each unit is called the *primary* or *index case*. The transmission probability or secondary attack rate is estimated by observing the proportion of the people in the transmission unit who become infected. Alternatively, a cohort of susceptible individuals could be recruited and followed over time. As individuals become infected the transmission units or contacts might be ascertained. As mentioned in the discussion on observational studies, ascertainment bias can be substantial when ascertaining transmission units by infectives. The latter method of ascertainment of transmission units would be less prone to ascertainment bias.

To estimate the conventional household SAR, data on the time of onset of disease for each case in the household as well as knowledge of who is susceptible are required. Also needed are estimates or assumptions about the minimum and maximum incubation periods, the latent period, and the maximum time that a person remains infectious. Using this information, one then needs to define the time interval after the occurrence of the index case that would include the secondary cases. Based on the time of onset data within each household or transmission unit, each case is defined as being either a secondary case or not. In the measles study in Senegal presented earlier, assumptions about all of these factors were made to define the secondary cases presented in Table 5–3. The estimated household secondary attack rate is the total number of secondary cases in all households divided by the total number of at-risk susceptibles in all households. A generalized estimating equation (GEE) (Liang and Zeger 1986) approach can be used to deal with clustering effects that may occur. In some cases, tertiary or higher generation cases may be included in the analysis by calling the secondary cases the index case. Note that the index cases are not included in the analysis, nor are any coprimaries. *Coprimaries* are people who developed disease too soon after the index case to have been infected by the index case. The assumption is that the households or other small transmission units are independent of one another.

Similar to the household secondary attack rate is the *case-contact* approach. In the case-contact approach, an index case is identified, then the people who have made contact with the index case are identified. For example, in tuberculosis or HIV, through contact tracing the people who have made contact with the infective person might be identified and their infection status ascertained. One difficulty in estimating the transmission probability from such a study is in determining the temporal order of infection in the contacts. Difficulties in estimating the conventional SAR and case-contact rates include determination of the latent and incubation periods, ascertainment of onset times of cases, and determining when an exposure to infection has taken place. The actual value of the estimated SAR can depend on the choice of the transmission unit and the definition of contact. For example, in a study of measles transmission in Senegal, the SARs estimated in schools, at homes, and in huts differed (Cisse et al. 1999). Kemper (1980) discusses biases in conventional SAR estimation.

Susceptibles exposed to infective contacts
Another study design approach to estimate the transmission probability is to assemble a cohort of susceptibles. The study then follows the susceptibles and collects information on their contacts with infectives or potential infectives. Study subjects might give information on the average number of contacts rather than the exact number of contacts they each make per unit time. From this, the expected number of contacts during the study period can be estimated. The binomial model is probably the most commonly used model for estimating the transmission probability when susceptibles make more than one potentially infectious contact. It can take on very complicated forms, depending on assumptions about variability in the transmission probability, time-varying covariates, and the amount and quality of data available. The model can be embedded in complex Markov or survival models. The principles of the binomial model were discussed in Chapter 4. One approach to including covariate effects as multiplicative factors on the transmission probability was discussed previously in the section on the transmission probability ratio. The data required are infection outcome, number of potentially infective contacts, and covariate status for each person in the study. Parameter estimates are obtained using numerical methods. Unfortunately, only limited software is available for estimating transmission probabilities using the binomial model. Software is usually written for particular situations.

The secondary attack rate can also be defined based on the binomial model for several contacts. For example, let m be the number of sexual contacts between two partners over the course of a study, where one of the partners is infected. Then the probability of being infected after m contacts is $1 - (1 - p)^m$, the per partnership SAR over the study interval.

Studies in a community of transmission units

In the study designs to estimate the transmission probability ratios described before, the transmission units such as houses or partnerships were assumed to be independent of each other. The susceptibles in the transmission unit were assumed to be exposed to infection only by the index case, who had somehow become infected. In Chapter 4, we discuss having small transmission units within a larger community. If the small transmission units (e.g., households or partnerships) are part of a community in which individuals from different transmission units interact, then the individuals can become infected either within the transmission unit or in the community at large.

The model developed by Longini and Koopman (1982) for transmission in a community of households takes into account both sources of infection. Two parameters are estimated. One is the SAR, the probability of being infected within the transmission unit from one infective. The other is the community probability of infection (CPI), the probability of being infected in the community at large over the course of the study or during an epidemic. Thus the model allows estimation of parameters from two different levels. The SAR is a conditional parameter from level I of Table 5–2. The CPI is an unconditional parameter from level IV of Table 5–2, and is closely related to the incidence proportion. The simplest version of the model assumes that mixing is random within the small transmission units and random within the community outside of the transmission units.

In the simplest study design the data requirements to fit the model are to know for each individual his transmission unit and his infection status at the beginning and the end of the study or epidemic. That is, simple final value data and the distribution of susceptibles and infected people within the transmission units are sufficient to estimate the SAR and the CPI. For any particular person, there would be uncertainty about the source of infection. The community could be a town, and the transmission units be households, schools, or other units. This method can be used to study transmission of diseases such as measles, influenza (Longini and Koopman, 1982), or dengue (Dantes et al. 1988). Alternatively, the community could be composed of sexually active people, with nonmonogamous partnerships forming the transmission units.

Table 5–4 presents data from an Asian influenza epidemic from households with three initially susceptible people in them, which we assume to be the whole community. The data are the number of households that had either 0, 1, 2, or all 3 people infected by the end of the epidemic. Using the model developed by Longini and Koopman, the estimated SAR is 0.166, and the estimated CPI is 0.114. The interpretation is that the probability of being infected from one infective in a household is 0.166, while the probability of being infected in the community at large, allowing for transmission within households, is 0.114. We emphasize that the two estimates have very different meanings. The SAR is conditional on being exposed to infection, while the CPI is an unconditional measure related to the incidence proportion.

In fact, we can estimate the usual incidence proportion from these data by simply ignoring the household structure. That is,

Table 5–4 Observed and Expected Distributions of Asian Influenza Data (Sugiyama 1960) in Households of Size Three as Analyzed by Longini and Koopman (1982)

Number of Cases	Observed of Number Households	Expected Number of Households
0	29	29.17
1	9	7.87
2	2	3.62
3	2	1.34
Total	42	42.00

suppose we do not have information on households. There are 42 households with three people each, so the total population is 126 people. From Table 5–4, we can calculate that 19 people became infected. The incidence proportion is $R = 19/126 = 0.151$. The incidence proportion is interpreted as the probability of becoming infected within the population without any further assumptions about the dynamics of interaction. Note that the incidence proportion, R, is higher than the estimate of community probability of infection, CPI. The simple incidence proportion is higher than the community probability of infection, because the incidence proportion includes the portion of the infected individuals who, under the model that included the SAR, were estimated to have been infected within households. This simple example illustrates the importance of considering the mixing assumptions within a population when developing models for estimating meaningful population parameters in infectious disease epidemiology.

At the other extreme, we could fit the Reed-Frost model, presented in Chapter 4, to these data. That model assumes that households are independent of one another. The probability of becoming infected from the community would be 0 and the estimated transmission probability within the household would be higher than that estimated with the model of Longini and Koopman. The Reed-Frost analysis would also not include the 29 households in Table 5–4 in which no one was infected. The Reed-Frost model, similar to the conventional SAR approach, assumes that there is at least one index case in each transmission unit included in the analysis. Note that in this influenza example, we do not have the information required to estimate the conventional SAR, because we have no data on the time of onset of infection, and we have made no assumptions about the latent, incubation, or infectious periods. Also, we have not made assumptions about who became infected from inside the household or outside in the community. We have partially replaced our data requirements with model assumptions.

Covariates are easily incorporated into the model to estimate the effect of risk factors on both the SAR and the CPI (Longini et al. 1988, Haber et al. 1988, Magder and Brookmeyer 1993, O'Neill et al. 2000). In a study of dengue transmission, Dantes and associates (1988) used the model to estimate the relative risk of transmission at both the individual and the household level.

The general principle of modeling transmission in small units embedded within a community can be extended in many different ways. Time can be incorporated into the model and time-to-event data used (Addy et al. 1991, Rampey et al. 1992). We can collect information from study participants on their number of sexual contacts both within their partnerships and with other people in the community (Longini et al. 1999). The parameters of the model to be estimated are then the transmission probability per sexual contact within the partnership and the probability of infection with a person of unknown infection status in the community at large. If an estimate of prevalence of infection in the population is available, then the transmission probability in the community at large can also be estimated (Hudgens et al. 2001).

The *augmented* study design is another extension of the idea of small transmission units within a community (Longini et al. 1996, Datta et al. 1998). In the augmented study design, individuals are recruited and possibly randomized to intervention. The individual recruitment and randomization is similar to standard randomized studies that aim to estimate relative risks based on one of the unconditional measures, such as incidence rate. However, then individuals with whom the primary study participants make contact, such as in a household or partnership, are also recruited. That is, the transmission unit of the participant is recruited into the study and augments the original primary study. The augmented participants may or may not be also randomized to intervention. In this way, the design is similar to that discussed under the conventional SAR studies. The advantage of the augmented design over conventional individual recruitment with randomization is that it permits estimation of the transmission probability ratios and, in particular, the effect of risk factors or interventions on infectiousness.

Comparison of Assumptions and Data Structures

There are more variants of study designs that incorporate information and assumptions about contact structures and transmission units than those presented here, but they will follow the same principles. To estimate the transmission probability and effects of risk factors on susceptibility and infectiousness, generally some information about contacts between susceptibles and infectives is required. Assumptions about how a population mixes in small transmission units and how the transmission units interact influence the transmission model that is developed. This in turn determines how data are analyzed, and ultimately what the parameter estimates are and how we interpret them.

In conventional SAR studies, the assumption is that the households or transmission units are independent, while in the model of households within communities, infection can take place both within and outside the small transmission unit. If the transmission probability or SAR is estimated without taking into account the opportunity to become infected outside of the transmission unit, it will overestimate the actual probability of becoming infected per contact. In general, ratio measures are less biased by this problem. The drawback in using a model such as that developed by Longini and Koopman is that it contains strong modeling assumptions about the mixing in the community. It also requires that the transmission units in fact are part of a community. An advantage of the conventional SAR studies or case-contact study designs is that the minicohorts or transmission units do not need to be within a single community. The minicohorts are assumed to be independent of one another.

The data requirements and use of the data are different in the different approaches. While in the conventional SAR studies, the index cases are excluded from the analysis, in the approach assuming transmission units within a community, all cases are included in the analysis. We leave it as an exercise for the reader to create a hypothetical community composed of small transmission units. Assign to each individual a covariate status (0,1) and also an infection time and infection status at the end of an epidemic. Consider the various approaches for estimating the effect measures, such as the conventional SAR, the SAR and the CPI simultaneously, and the simple incidence proportion. How do the data being used for each approach differ? What parameters can be estimated? What is the interpretation of the measures under each approach?

Case-Control Studies

Case-control studies can produce good estimates of either the incidence rate ratio or the incidence proportion ratio (Greenland and Thomas 1982). In case-control studies, cases are ascertained from the population of interest, or source population. Rather than following an entire cohort or gathering information on the entire source population, however, controls are sampled from the source population to estimate the relative person-time at risk or the relative proportions of the source population in the different treatment or covariate groups. In infectious disease epidemiology, case-control studies can also be used to estimate the transmission probability ratio and for preliminary etiologic studies in outbreak investigations. If properly conducted, case control studies are important, efficient alternatives to cohort studies as well as randomized trials (Smith 1982, 1987, Smith et al. 1984, Rodrigues and Smith 1999).

A case-control study might be conducted within the cohort that is under study (i.e., a nested case-control study), or at least within a well-defined population. Thinking of the case-control study as being nested within a cohort or a well-defined source population enables clearer formulation of assumptions about the underlying dynamics and covariate distributions. This, in turns, aids in choosing the appropriate sampling method and method of analysis for estimating the incidence rate ratio, the incidence proportion ratio, or the transmission probability ratio. We consider first estimating the incidence rate ratio of two covariate groups with a case-control study. The *odds ratio*, *OR*, is

$$OR = \frac{(A_1 / A_0)}{(B_1 / B_0)} = \frac{A_1 B_0}{A_0 B_1},$$

where A_0 and A_1 are the number of cases in the two different covariate groups, and B_0 and B_1 are the number of controls selected in the two groups. To estimate the incidence rate ratio using the odds ratio, the goal in sampling the controls is to estimate the relative person-time at risk in the two groups.

There are two main ways to sample the controls that give a consistent and unbiased estimate of the incidence rate ratio under certain conditions if the incidence rate ratio is constant in time (Greenland and Thomas 1982). One approach is density sampling, also called risk set sampling. In this approach, controls are selected from the population at risk at the time of onset of each case. By selecting the controls matched on time with the cases, density sampling samples the relative distribution of person-time in the two covariate groups. Another approach to sample controls does not match the sampling on time with the cases. In time-unmatched sampling, controls are selected so that the expected ratio of the number of controls in one covariate group to the number in the other covariate group equals the expected ratio of the total person-time at risk in one covariate group to the person-time at risk in the other covariate group over the entire case ascertainment period. Thus the probability that any control is selected should be proportional to the amount of time that he or she is at risk in the study to become a case.

The underlying cohort can be a dynamic cohort as long as the assumptions are satisfied. If people enter and leave the group at risk so that individuals have different person-time at risk, then the probability of being sampled should be proportional to the person-time at risk. This will occur as a consequence of time-matched sampling, but would need to be computed with time-unmatched sampling. In both the time-matched and the time-unmatched sampling schemes, controls should be sampled independently of the covariates of interest.

If controls are sampled matched on time with cases using the density sampling, then the odds ratio can be computed using either a time-unmatched or a time-matched analysis. If the odds ratio is computed using an unmatched analysis of the time-matched cases and controls, then it is a consistent estimator of the constant incidence rate ratio if the proportion of the population at risk that has a particular covariate value is constant. This assumption would be violated, for instance, if a vaccination program were beginning so that the proportion of people who were vaccinated increased over the course of the study. If the odds ratio is calculated using a matched pair or discordant pair analysis that is matched on time, then it is a consistent estimator of the constant incidence rate ratio with no further assumptions. That is, if vaccine coverage were increasing, density sampling with a time-matched analysis could still be used to estimate the incidence rate ratio, and thus, the vaccine efficacy. In both of these situations, as long as the incidence rate ratio is constant, the baseline incidence rate may vary. For instance, there could be seasonal variation over the course of the study, such as in malaria, or there could be an epidemic, as with influenza.

If controls are sampled without matching on time, then the analysis cannot be matched on time. The odds ratio computed from the unmatched sampling scheme is a consistent estimator of a constant incidence rate ratio if either (1) the baseline incidence rate or (2) the proportion of those at risk who are in each of the covariate groups is constant. For example, if people were all vaccinated before the influenza season, then the time-unmatched approach could be used to estimate the incidence rate ratio.

If the incidence ratio is not constant, then there is no unique effect to estimate with the odds ratio. A useful illustration of these principles is found in Struchiner and colleagues

(1990). Using the example of malaria vaccination and seasonal transmission of malaria, they compare the three different odds ratio estimators of the incidence rate ratio.

To estimate the relative incidence proportion ratio using a case-cohort study, the controls are used to estimate the relative proportions of the population in each covariate group. That is, the goal is to use the ratio B_1/B_0 to estimate the distribution of the covariate among the cohort members rather than among the person-time at risk. The controls in the cohort are sampled regardless of their person-time at risk (Wacholder 1991, Rothman and Greenland 1998). Individuals who become cases may also be sampled as controls. Again, the controls should be selected independently of their covariate groups.

Case-control studies in infectious diseases need to satisfy the same assumptions as case-control studies in noninfectious diseases. The assumptions underlying many types of case-control studies may, however, be dramatically violated in studies of infectious diseases. Stationarity (i.e., dynamic equilibria of the human and parasite populations) assumptions commonly do not apply, the incidence rate ratio may change with time if the effect of an intervention wanes, and the proportion of the population with a particular covariate value can change quickly (Struchiner et al. 1990). Thus the underlying assumptions should be examined closely for their applicability.

To estimate the transmission probability ratio for susceptibility, cases are those people in the population for whom information on exposure to infection is available. Controls are selected conditional on being exposed to infection, possibly matched on a similar level of exposure, to estimate the odds of having a particular covariate status. The use of case-control studies to estimate the transmission probability ratio needs more formal research.

The preceding sampling designs do not rely on the rare disease assumption for the odds ratio to be a consistent estimator of the effect measure of interest (Rothman and Greenland, 1998). However, a study design frequently used in infectious diseases does rely on the rare disease assumption for the odds ratio estimator to be a good approximation to the incidence rate ratio or the incidence proportion ratio. In outbreak investigations where a point source epidemic is suspected, the potential controls are usually considered to be those people who did not get the illness. Sampling of controls generally takes place after the outbreak has occurred, so it is not matched on time. In this situation, if a large portion of the population became ill, the odds ratio could differ substantially from the population parameter of interest. However, in such studies, the main interest may be in simply identifying that people in one covariate group have a higher risk of being ill than those in the other covariate group. For example, it may be of interest to determine that people who ate potato salad had a higher risk of being ill than those who did not. An unbiased estimate of the underlying relative incidence rate or relative incidence proportion is probably not important.

Two-stage case control studies and studies with validation sets

Exposure to infection is often difficult to measure accurately. Also, definitive diagnosis of a case of a particular infectious disease can be expensive or difficult. For example, in influenza studies the case definition in a study might be a set of symptoms such as coughing, fever, aches, or sore throat, but not include culture-positive confirmation. In either case, with poorly measured exposure to infection or a nonspecific case definition, estimates of effects could be very biased and, in particular, attenuated. Study designs have been developed in nutritional and cancer epidemiology that have potential use in infectious disease epidemiology. The general idea is to measure an inexpensive or easily available covariate or outcome measure on everyone in the study. In a smaller subsample of the study, called a validation set, the more accurate exposure or outcome measure that is somehow correlated with the poorer value is measured. Statistical methods have been developed to combine the two levels of information (Pepe and Fleming 1991, Carroll and Wand 1991, Reilly and Pepe 1995, Robins et al. 1994, 1995). The small group with the good measurement helps to get more accurate effect estimates, while the larger study helps to have smaller variance in the estimate. Case-control studies can be done as two-stage studies (Cain and Breslow 1988, Breslow and Cain 1988, Flanders and Greenland 1991, Zhao and Lipshitz 1992, Breslow and Holubkov 1997), where the more accurate measures or additional covariates are collected on a sample of the cases and controls. Golm and associates (1998, 1999) showed the potential for using two levels of exposure to infection information for good estimates of vaccine efficacy for susceptibility and infectiousness in HIV vaccine trials. Increased use of validation sets and two-stage case-control methods could greatly improve the design of efficient studies in infectious disease epidemiology (Halloran and Longini 2000).

OTHER TYPES OF STUDIES

Cross-Sectional Studies

A cross-sectional study takes place within a short time window and includes all people or a sample of the people in the population at that time. *Prevalence* studies using a cross-sectional study design are used to estimate the current status of infection in a population. Similarly, *seroprevalence* studies measure the prevalence of immune response to an infectious agent and give information on the history of infection in a population. Estimating incidence rates, also age-specific incidence rates, from prevalence data is possible, assuming that the conditions of disease transmission have remained fairly stable and that immunity does not wane (i.e., once infected, the serologic test remains positive) (Grenfell and Anderson 1985, Keiding 1991).

As shown in Chapter 4, seroprevalence can be used as a measure of herd immunity. Seroprevalence can also be used for a simple method to estimate the basic reproductive number, R_0, if the transmission system is assumed to be in dynamic equilibrium, that is, not changing a lot over time. The underlying idea is that when the average incidence rate and prevalence of disease are not changing, an infectious case produces on average one other infectious case, so the reproductive number $R = 1$. From the relation $R = R_0x = 1$, the proportion susceptible at equilibrium would be $x = 1/R_0$. Assuming random mixing, then R_0 is roughly estimated by the reciprocal of the proportion susceptible. In the study of hepatitis A and E in Vietnam (Hau et al. 1999), seroprevalence of anti-HAV IgG was 0.97 and of anti-HEV IgG was 0.09. The proportion susceptible to each is then 0.03 and 0.91, respectively. The estimate of R_0 for hepatitis A in this population is $R_0 = 1/0.03 = 33$ and for hepatitis E is $R_0 = 1/0.91 = 1.1$. Hau and colleagues (1999) express concern that conditions such as flooding or poor hygiene could favor the epidemic spread of hepatitis E. Essentially, a worsening of conditions would increase the R_0 of hepatitis E.

Spatial mapping and GIS systems

Spatial studies of infectious diseases, including vector-borne diseases, are becoming more common. These studies often include the use of geographical information systems (GIS). For instance, they may be used to map the mosquito breeding grounds in relation to houses.

Community level studies

As mentioned previously, estimation of the indirect, total, and overall effects of interventions using the study designs for dependent happenings requires comparison of populations, not just individuals. Such community trials fall into the category of cluster or group randomized trials where whole social units, rather than independent individuals are randomly assigned to treatment groups (Hayes et al. 2000, Koepsell et al. 1992, Donner 1998, Prentice and Sheppard 1995, Klar et al. 1995, Murray 1998). Observational studies in which the community is the level of observation are called ecologic studies. In choosing the communities or populations to include in a study, it is important to assure that they are separated as much as possible in every way that is relevant for transmission. If the populations are not transmission dynamically separated, then the intervention in one population will affect transmission in the other population. The indirect effects might be similar in the two populations. A study that compares nonseparate populations will yield an attenuated estimate of the potential indirect effects of intervention. Transmission patterns that differ greatly among communities can also mask the indirect effects of intervention. Matching by transmission characteristics is an option to consider (Hayes et al. 1995). In selecting communities, some thought is required about the transmission patterns and sources of exposure to infection in a population. These transmission patterns will greatly influence the magnitude of the indirect effects.

Analysis needs to be by unit of observation. For population-level studies, the unit of observation and level of analysis is the population, not the individual, so sample size calculations must be done accordingly. That is, if a study takes place in two populations each with 10,000 people and the comparison is how population A compares to population B, then the sample size is two, not 20,000.

ESTIMATION AND INFERENCE

For estimation of incidence rate ratios or hazard rate ratios, Poisson regression or stratified survival analysis is used most often. Special to infectious disease epidemiology is the possibility of using the dependent happening relation (1) to incorporate information on people who are infected in any given time interval to model the shape of the baseline hazard (Longini and Halloran 1996). The proportional hazards model (Cox 1972) is often used to estimate the regression parameters when time-to-event data are available. In the proportional hazards model, the baseline hazard rate need not be estimated, but just the ratio of the two hazard rates. For example, in malaria, with the high variability of mosquito densities as the seasons change, it is possible to estimate the hazard rate ratio of two covariate groups without having to estimate the actual seasonal variation in transmission. However, if using the proportional hazards model, it is important to check whether the assumption of proportionality holds. These methods of analysis are discussed in detail elsewhere (Cox and Oakes 1984, Andersen et al. 1993).

In the discussion of the incidence rate ratio as the ratio of two dependent happening expressions, we made some strong assumptions without making them explicit. By writing the expressions as we did, there is an implicit assumption that within each of the covariate groups, everyone is the same. That is, each covariate group is assumed homogeneous with respect to the contact rate, transmission probability, and the prevalence of infection in their contacts. However, it is likely there will be unmeasured heterogeneities within study groups. Then, even if the effect of the risk factor in question does not change over time, the effect may appear to change. Some people may be exposed to infection more than others. Some may be more susceptible to infection than others. Those people with the higher susceptibility or higher exposure tend to develop the disease first. The estimated relative risk will change with time. If the estimated relative risk changes with time, the question is whether it is a true time-varying effect or an artifact of the unmeasured heterogeneities. If the effect is truly changing over time, then models for time-varying effects should be used (Schoenfeld 1982, Durham et al. 1998). If it is possible to measure the heterogeneities, then the analysis can be stratified accordingly. It is generally not possible to measure all heterogeneities, however. If the effect seems to vary because of unmeasured heterogeneities, frailty models can be tried. These are random effects models for time-to-event data (Vaupel et al. 1979, Longini and Halloran 1996).

Logistic regression is often used to analyze data obtained on whether an event occurs because the outcome data are binary, not time dependent. The model allows incorporation of covariates. In a cohort study, the estimates of the logistic regression parameters can be transformed to obtain an estimate of the incidence proportion ratio. Logistic regression can also be used to analyze some case-control data.

Little standard software exists for estimating transmission probabilities or the models that are variants of transmission units within communities. Conventional secondary attack rate ratios can be estimated using logistic regression or generalized estimating equations (Liang and Zeger 1986), for which software is available. Generalized estimating equations take into account clustering within households in the variance estimates.

In this chapter we have primarily discussed questions of estimation. Inference as a general topic goes beyond the scope of this chapter. The epidemiologist is well advised to include a biostatistician in the study team early in the design stage. Statistical inference has to do with predicting what might be expected of further observations or further studies, and quantifying degrees of certainty or uncertainty about the results we have obtained. The design of epidemiologic studies needs to include clear statements about the degree of certainty desired in the results. These have important consequences for sample size and power required in the studies.

There are different approaches to statistical inference, including the frequentist, likelihood, and Bayesian approaches. They differ in their emphasis on use of prior information, whether testing or estimation is more important, whether decision or inference is central, and in their sensitivity to the sampling procedure (Oakes 1990). Because of their emphasis on estimation and inference, rather than on testing and decision making, likelihood and Bayesian approaches to inference are more natural than frequentist approaches for epidemiologic studies. Bayesian approaches are being used increasingly as the complex computational methods they require become more feasible. The Bayesian approach allows integration of information from different sources in a natural way, and thus is particularly useful for observational studies. In epidemiology, inference using confidence intervals is preferred over using p-values. The usual confidence intervals depend on a normal approximation. Bootstrap confidence intervals do not require a normal approximation (Efron and Tibshirani 1993) and should be considered for use in the analysis. Clayton and Hills (1993) provide a readable book on statistical models in epidemiology.

SUMMARY

Because of the fundamental role of transmission of the infectious agent and dependent happenings, epidemiologic measures of interest in infectious disease epidemiology include the transmission probability, the contact rate, infectiousness, the basic reproductive number, R_0, as well as direct and indirect effect measures. The key dependent happening relation is that the incidence rate is a function of the contact rate, the transmission probability, and the prevalence of infectives in the population. The dependent happening relation helps distinguish risk factors for susceptibility from risk factors for exposure to infection. Measures such as the transmission probability that condition on contact between infectives and susceptibles are called conditional parameters. Measures of disease frequency that do not, such as incidence rate and incidence proportion, are unconditional measures. Association and causal effects differ under most circumstances. Study designs in infectious disease epidemiology include several that enable estimation of the transmission probability ratio. These generally include information on contacts between individuals or within small transmission units. In estimation of indirect and overall effects of an intervention program, the unit of analysis is the population. The dynamics of infection and transmission units within a population need to be taken into account when designing and interpreting studies in infectious disease epidemiology.

Dr. Halloran was partially supported in writing this chapter by NIH grants RO1-AI32042 and RO1-AI40846.

REFERENCES

Addy CL, Longini IM, and Haber MS. A generalized stochastic model for the analysis of infectious disease final size data. Biometrics. 47:961–974, 1991.

Andersen PK, Borgan O, Gill RD, and Keiding N. Statistical Models Based on Counting Processes. New York: Springer-Verlag, 1993.

Breslow NE, and Cain KC. Logistic regression for two-stage case-control data. Biometrika. 75:11–20, 1988.

Breslow NE, and Holubkov R. Maximum likelihood estimation of logistic regression parameters under two-phase, outcome dependent sampling. J. R. Stat. Soc. B. 59:447–461, 1997.

Brookmeyer R, and Gail MH. Biases in prevalent cohorts. Biometrics. 43:739–749, 1987.

Cain KC, and Breslow NE. Logistic regression analysis and efficient design for two-stage studies. Am. J. Epidemiol. 128:1198–1206, 1988.

Carroll RJ, and Wand WP. Semiparametric estimation in logistic measurement error models. J. R. Stat. Soc. B. 53:573–585, 1991.

Cisse B, Aaby P, Simondon F, et al. Role of schools in the transmission of measles in rural Senegal: implicationsn for measles control in developing countries. Am. J. Epidemiol. 149:295–301, 1999.

Clayton D, and Hills M. Statistical Models in Epidemiology. Oxford: Oxford University Press, 1993.

Cox DR. Regression models and life-tables (with discussion). J. R. Stat. Soc. 30(Series B):284–289, 1972.

Cox DR. Planning of Experiments. New York: John Wiley & Sons, 1958.

Cox DR, and Oakes D. Analysis of Survival Data. London: Chapman and Hall, 1984.

Dantes HG, Koopman JS, Addy CL, et al. Dengue epidemics on the Pacific Coast of Mexico. Int. J. Epidemiol. 17:178–186, 1988.

Datta S, Halloran ME, and Longini IM. Augmented HIV vaccine trial designs for estimating reduction in infectiousness and protective efficacy. Stat. Med. 17:185–200, 1998.

Diggle PJ, Liang K-Y, and Zeger SL. Analysis of Longitudinal Data. Oxford: Oxford University Press, 1994.

Donner A. Some aspects of the design and analysis of cluster randomization trials. Appl. Stat. 47:95–114, 1998.

Durham LK, Longini, IM, Halloran ME, et al. Estimation of vaccine efficacy in the presence of waning: application to cholera vaccines. Am. J. Epidemiol. 147:948–959, 1998.

Efron B, and Tibshirani RJ. An Introduction to the Bootstrap. New York: Chapman and Hall, 1993.

Evans AS. Causation and disease: The Henle-Koch postulates revisited. Yale J. Biol. Med. 49:175–195, 1976.

Fine PEM, Clarkson JA, and Miller E. The efficacy of pertussis vaccines under conditions of household exposure: further analysis of the 1978–80 PHLS-ERL study in 21 area health authorities in England. Int. J. Epidemiol. 17(3):635–642, 1988.

Flanders WD, and Greenland S. Analytic methods for two-stage case-control studies and other stratified designs. Stat. Med. 10:739–747, 1991.

Freeman J, and Hutchison GB. Prevalence, incidence, and duration. Am. J. Epidemiol. 112:707–723, 1980.

Gail MH. Adjusting for covariates that have the same distribution in exposed and unexposed cohorts. In: Moolgavkar SH, and Prentice RL, eds. Modern Statistical Methods. New York: Wiley, 1986: 3–18.

Gail MH. The effect of pooling across strata in perfectly balanced studies. Biometrics. 44: 151–162, 1988.

Gail MH, Tan WY, and Piantadosi S. The effect of omitting covariates on tests for no treatment effect in randomized clinical trials. Biometrika. 75:57–64, 1988.

Gail MH, Wieand S, and Piantadosi S. Biased estimates of treatment effect in randomized experiments with non-linear regressions and omitted covariates. Biometrika. 71:431–444, 1984.

Glynn JR, Vynnycky E, and Fine PEM. Influence of sampling on estimates of clustering and recent transmission of mycobacterium tuberculosis derived from DNA fingerprinting techniques. Am. J. Epidemiol. 149:366–371, 1999.

Golm GT, Halloran ME, and Longini IM. Semiparametric models for mismeasured exposure information in vaccine trials. Stat. Med. 17:2335–2352, 1998.

Golm GT, Halloran ME, and Longini IM. Semiparametric methods for multiple exposure mismeasurement and a bivariate outcome in HIV vaccine trials. Biometrics. 55:94–101, 1999.

Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. Am. J. Epidemiol. 125:761–768, 1987.

Greenland S, and Robins JM. Identifiability, exchangeability, and epidemiologic confounding. Int. J. Epidemiol. 15:412–418, 1986.

Greenland S, Robins JM, and Pearl J. Confounding and collapsibility in causal inference. Stat. Sci. 14:29–46, 1999.

Greenland S, and Thomas DC. On the need for the rare disease assumption in case-control studies. Am. J. Epidemiol. 116(3):547–553, 1982.

Grenfell BT, and Anderson RM. The estimation of age-related rates of infection from case notifications and serological data. J. Hyg. Cam. 95:419–436, 1985.

Haber M, Longini IM, and Cotsonis GA. Models for the statistical analysis of infectious disease data. Biometrics. 44:163–173, 1988.

Halloran ME, and Longini IM. Using validation sets for outcomes and exposure to infection in vaccine field studies. Am J Epidemiol (in press) 2001.

Halloran ME, Longini IM, Struchiner CJ, Haber MJ, and Brunet RC. Exposure efficacy and change in contact rates in evaluating prophylactic HIV vaccines in the field. Stat. Med. 13:357–377, 1994.

Halloran ME, and Struchiner CJ. Study designs for dependent happenings. Epidemiology. 2:331–338, 1991.

Halloran ME, and Struchiner CJ. Causal inference for infectious diseases. Epidemiology. 6:142–151, 1995.

Halloran ME, Struchiner CJ, and Longini IM. Study designs for different efficacy and effectiveness aspects of vaccination. Am. J. Epidemiol. 146:789–803, 1997.

Hau CH, Hien TT, Tien NTK, et al. Prevalence of enteric hepatitis A and E viruses in the Mekong River Delta region of Vietnam. Am. J. Trop. Med. Hyg. 60:277–280, 1999.

Hayes RJ, Mosha F, Nicoll A, et al. A community trial of the impact of improved sexually transmitted disease treatment on HIV epidemic in rural Tanzania: 1. Design. AIDS. 9:919–926, 1995.

Hayes RJ, Alexander NDE, Bennett S, Cousens SN. Design and analysis issues in cluster-randomized trials of interventions against infectious diseases. Stat. Methods Med. Res. 9:95–116, 2000.

Holland PW. Statistics and causal inference. J. Am. Stat. Assoc. 81:945–960, 1986.

Hooper RR, Reynolds GH, Jones OG, et al. Cohort study of venereal disease. I: The risk of gonorrhea transmission from infected women to men. Am. J. Epidemiol. 108:136–144, 1978.

Hudgens MG, Longini IM, Halloran ME, et al. Estimating the transmission probability of human immunodeficiency virus in injecting drug users in Thailand. Appl. Stat. 50:(Part 1), 2001.

Keiding N. Age-specific incidence and prevalence: A statistical perspective. J. R. Stat. Soc. A. 154(3):371–412, 1991.

Kemper JT. Error sources in the evaluation of secondary attack rates. Am. J. Epidemiol. 112: 457–464, 1980.

Klar N, Gyorkos T, and Donner A. Cluster randomization trials in tropical medicine: a case study. Trans. R. Soc. Trop. Med. Hyg. 89: 454–459, 1995.

Koepsell TD, Wagner EH, Cheadle AC, et al. Selected methodological issues in evaluating community-based health promotion and disease prevention programs. Annu. Rev. Public Health. 13:31–57, 1992.

Koopman JS, Longini IM, Jacquez JA, et al. Assessing risk factors for transmission of infection. Am. J. Epidemiol. 133:1199–1209, 1991.

Liang KY, and Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 73:13–22, 1986.

Lipsitch M. Vaccination against colonizing bacteria with multiple serotypes. Proc. Natl. Acad. Sci. USA. 94:6571–6576, 1997.

Longini IM, Datta S, and Halloran ME. Measuring vaccine efficacy for both susceptibility to infection and reduction in infectiousness for prophylactic HIV-1 vaccines. J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 13: 440–447, 1996.

Longini IM, and Halloran ME. A frailty mixture model for estimating vaccine efficacy. Appl. Stat. 45:165–173, 1996.

Longini IM, Hudgens MG, Halloran ME, and Sagatelian K. A Markov model for measuring vaccine efficacy for both susceptibility to infection and reduction in infectiousness for prophylactic HIV-1 vaccines. Stat. Med. 18: 53–68, 1999.

Longini IM, and Koopman JS. Household and community transmission parameters from final distributions of infections in households. Biometrics. 38(1):115–126, 1982.

Longini IM, Koopman JS, Haber M, and Cotsonis GA. Statistical inference for infectious diseases: Risk-specified household and community transmission parameters. Am. J. Epidemiol. 128(4):845–859, 1988.

Magder L, and Brookmeyer R. Analysis of infectious disease data from partners studies with unknown source of infection. Biometrics. 49:1110–1116, 1993.

Murray DM. Design and Analysis of Group-Randomized Trials. New York: Oxford University Press, 1998.

Oakes M. Statistical Inference. Chestnut Hill: Epidemiology Resources, Inc., 1990.

O'Neill, POD, Balding DJ, Becker NG, Eerola M, and Mollison D. Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. Appl. Stat. 49:517–542, 2000.

Orenstein WA, Bernier RH, and Hinman AR. Assessing vaccine efficacy in the field: further observations. Epidemiol. Rev. 10:212–241, 1988.

Pepe MS, and Fleming TR. A nonparametric method for dealing with mismeasured covariate data. J. Am. Stat. Assoc. 86:108–113, 1991.

Prentice R, and Sheppard L. Aggregate data studies of disease risk factors. Biometrika. 82: 113–125, 1995.

Rampey AH, Longini IM, Haber MJ, and Monto AS. A discrete-time model for the statistical analysis of infectious disease incidence data. Biometrics. 48:117–128, 1992.

Reilly M, and Pepe MS. A mean score method for missing and auxiliary covariate data in regression models. Biometrika. 82:299–314, 1995.

Rhodes PH, Halloran ME, and Longini IM. Counting process models for differentiating exposure to infection and susceptibility. J. R. Stat. Soc. B. 58:751–762, 1996.

Robins JM, Hsieh F, and Newey W. Semiparametric efficient estimation of a conditional density with missing or mismeasured covariates. J. R. Stat. Soc. Ser. B. 57:409–424, 1995.

Robins JM, Rotnitzky A, and Scharfstein DO. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: Halloran ME and DA Berry, eds. Statistics in Epidemiology, Environment and Clinical Trials. New York: Springer-Verlag, 1999.

Robins JM, Rotnitzky A, and Zhao LP. Estimation of regression coefficients when some regressors are not always observed. J. Am. Stat. Assoc. 89:846–866, 1994.

Rodrigues L, and Smith P. Case-control approach to vaccine evaluation. Epidemiol. Rev. 21: 56–72, 1999.

Rosenbaum P. Observational Studies. Berlin: Springer-Verlag, 1995.

Ross R. 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.

Rothman K, and Greenland S. Modern Epidemiology. Philadelphia: Lippincott-Raven, 1998.

Rubin DB. Bayesian inference for causal effects. The role of randomization. Ann. Stat. 7: 34–58, 1978.

Rubin DB. Comment: Neyman (1923) and causal inference in experiments and observational studies. Stat. Sci. 5:472–480, 1990.

Schoenfeld D. Partial residuals for the proportional hazards regression model. Biometrika. 69:239–241, 1982.

Small PM, Hopewell PC, Singh SP, et al. The epidemiology of tuberculosis in San Francisco: A population-based study using conventional and molecular methods. N. Engl. J. Med. 330:1703–1709, 1994.

Smith PG. Retrospective assessment of the effectiveness of BCG vaccination against tuberculosis using the case-control method. Tubercle. 62:23–35, 1982.

Smith PG. Evaluating interventions against tropical diseases. Int. J. Epidemiol. 16(2):159–166, 1987.

Smith PG, Rodrigues LC, and Fine PEM. Assessment of the protective efficacy of vaccines against common diseases using case-control and cohort studies. Int. J. Epidemiol. 13(1): 87–93, 1984.

Struchiner CJ, Halloran ME, Robins JM, and Spielman A. The behavior of common measures of association used to assess a vaccination program under complex disease transmission patterns—a computer simulation study of malaria vaccines. Int. J. Epidemiol. 19:187–196, 1990.

Sugiyama H. Some statistical contributions to the health sciences. Osaka City Med. J. 6:141–158, 1960.

Vaupel JW, Manton KG, and Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography. 16:439–454, 1979.

Wacholder S. Practical considerations in choosing between the case-cohort and nested case-control design. Epidemiology. 2:155–158, 1991.

Zhao LP, and Lipsitz S. Designs and analysis of two-stage studies. Stat. Med. 11:769–782, 1991.