Transactions of the Illinois State Academy of Science
(2000), Volume 93, #4, pp. 293-300

received 3/6/00
accepted 10/2/00

Methods for Calculating Two Incidence Estimators, Relative Change in Prevalence and Apparent Incidence, from Sequential Measures of Prevalence

John Mellen
Department of Biology
Bemidji State University
Bemidji, MN 56601

ABSTRACT

Investigating the dynamics of host-parasite systems requires an understanding of incidence rate, the rate of colonization of uninfected susceptible hosts. Perceived difficulties in collecting field information have confounded the use of incidence rate to describe metapopulation dynamics of naturally occurring parasite endemics in populations of wild hosts. Herein a rationale is presented for calculating two discrete time representations of the incidence concept, relative change in prevalence, and apparent incidence from sequential changes in prevalence in host-parasite systems conforming to models of simple epidemics. These methods do not require *a priori* knowledge of the infection status and may be applied to populations of wild hosts.

Keywords: incidence rate, prevalence, recovery rate, simple epidemics

INTRODUCTION

Understanding the population ecology of helminth parasites requires a knowledge of the rate at which helminths colonize their hosts. Unlike populations of free-living animals whose life histories are frequently described in terms of births and deaths, the dynamics of most helminth populations in their definitive hosts must be characterized by colonization and death (Esch and Fernandez 1993). In general, the offspring of one helminth generation are shed into the environment from where, after a more or less complicated developmental sequence depending upon the parasite species considered, infective larva must colonize a host and metamorphose into adult worms. The ephemeral nature of the availability of suitable hosts also contributes to the importance of colonization in the population ecology of helminths. Helminth infrapopulations, all the individuals of one parasite species within an individual host, may be extinguished through the development of a host immune response as well as by death of the host, while births in the host population generate new vacant habitat. Over time the maintenance of helminth metapopulations is recognized by parasitologists as being critically dependent upon the establishment of new infrapopulations through the colonization of susceptible hosts.

Ecological interactions within and between infrapopulations are considered also to be critically dependent upon the colonization of host individuals by parasites (Simberloff and Moore 1997). Many questions about the relationship between ecological phenomena and colonization rates have been generated by the evolutionary theory of parasites (i.e. Price 1980, 1987; Esch and Fernandez 1993). Examples of these include: (1) is infrapopulation size correlated with the magnitude of colonization rates; (2) is colonization rate associated with metapopulation size; (3) is colonization rate correlated with host specificity; and (4) are the colonization rates of common species high enough to allow interactions between parasite species to influence parasite community structure? Clearly, the estimation of the colonization rate of helminth parasites is related to a large number of areas of investigation in parasite population ecology. Unfortunately, the quantitative description of parasite colonization rate in natural settings and evaluation of its relationship to ecological expectations has remained problematic because of perceived difficulties in estimating the colonization rate of populations of wild hosts by parasitic helminths.

Within parasitology the term colonization rate is comparable to the epidemiological term incidence. Conceptually, epidemiologists use incidence to describe the propagation rate of cases of infections within a host population (Mausner and Bahn 1974). Margolis et al. (1982), serving as an *ad hoc* committee to the American Society of Parasitology on the use ecological terminology in parasitology, recommended adopting the term incidence for similar use in parasite ecology. For use in parasite ecology, they defined incidence as the proportion of new cases of parasitic infection among uninfected hosts during a period of time. This definition made a useful distinction between incidence and prevalence, the proportion of the host population infected with a parasite species, but provided little assistance in applying incidence to parasite ecology. Parasitologists' frustration with their inability to estimate incidence was expressed in the commentary of the *ad hoc* committee, “Incidence is not likely to be applicable when studying populations of feral [*wild*] animals because the number of uninfected individuals at the beginning of the time period is rarely known.” The committee's emphasis of prerequisite knowledge of the infection status of individuals directed parasitologists' attention toward impractical solutions.

The legacy of the comments of the *ad hoc* committee concerning incidence includes the belief among parasitologists that extraordinary collection techniques (mark and recapture) and case identification methods (techniques that don't kill the host) are required to estimate incidence. Recent authors have repeated that belief over several editions of a popular parasitology text (i.e. Roberts and Janovy 1999) indicating that the belief is becoming established as a fundamental of parasitology training in the United States. This report concerns an algebraic approach to using prevalence information to obtain estimates of incidence. The method is suitable for use in studies of parasitic infections in wild populations and does not require monitoring special individual hosts.

METHODS

The Relationship Between Incidence and Changing Prevalence

Seasonal irruptions of helminth infections in wild animals are commonly witnessed as the rise and fall in a sequence of estimates of prevalence. Such patterns (i.e. Dronen 1978) provide an indication of the timing and outcome of temporal changes in underlying processes of infection in terms of changes in the proportion of the host population that is

infected. In a sequence of two prevalence estimates, the prevalence at a subsequent time period depends upon an initial prevalence, P, and the change in prevalence, ΔP, during the sampling period. Stated algebraically

$$P_{t+1} = P_t + \Delta P \qquad \textit{Eq. 1}$$

Information concerning incidence rate is contained in ΔP and can be demonstrated using an extremely simple epidemic model (Figure 1). The model represents a host population divided into two states, susceptible, S, individuals and infected, I, individuals. It is termed a susceptible-infected or SI epidemic model (Poole 1974). In the SI model the transition of members of the host population from one state to the other depends upon incidence rate, *i*, and recovery rate, *r*, the proportion of infected hosts that return to the susceptible state.

When the sizes of the infection states of the population are considered as proportions, I is the proportion of the population in the infected state. Hence, I equals P, the prevalence of infection in the host population. It follows that the uninfected proportion of the host population, S, equals 1-P. A change in prevalence, ΔP, is equal to the difference between the proportion of the host population becoming infected, S*i*, and the proportion of the host population recovering from infection, I*r*. Substitution of the terms of the simple epidemic model into *Eq. 1* yields:

$$P = P_{t+1} - P_t = S_t\, i - I_t\, r \qquad \textit{Eq. 2}$$

Expressing the term I as P yields the equivalent

$$P = P_{t+1} - P_t = S_t\, i - P_t\, r$$

Equation 2 provides the framework for estimating incidence from changes in prevalence of infection in wild populations of host. Because *Eq. 2* is based on proportions of the population in each epidemiologic state, estimating incidence, at least in principle, is not dependent upon following special individuals through time. Obtaining values for *i* and *r,* requires the application of common techniques for solving equations with two unknowns.

Obtaining an Estimate of Incidence from Sequential Prevalence Data

When parasite infections conform to a simple epidemic model, estimates of both incidence rate, *i,* and recovery rate, *r,* may be obtained from sequential estimates of change in prevalence. These estimates enable the creation of a system of two equations which may be solved for estimates of *i* and *r*, that will satisfy both equations. The equations of interest represent two changes in prevalence developed from three consecutive samples.

$$\begin{aligned} P_{t+1} - P_n &= S_t\, i - P_t\, r \\ P_{t+2} - P_{t+1} &= S_{t+1}\, i - P_{t+1}\, r \end{aligned} \qquad \textit{Eqs. 3}$$

If it can be assumed that *i* and *r* remain constant between three consecutive samplings (limitations are discussed below) these parameters can be estimated using a technique of

matrix algebra termed Cramer's rule (Kline et al. 1959). Where only two sequential estimates of prevalence are available, the midpoint value of prevalence can be found by interpolation and used as the third prevalence estimate. The computations are straight forward and easily undertaken by hand or programmed into a spreadsheet. For example, assume three consecutive estimates of prevalence are P_1=47%, P_2 =65%, and P_3=75%. Then *Eqs. 3* may be assigned variables so that

$$P_2 - P_1 = S_1\, i - P_1 r \text{ becomes } 18 = 53i - 47r$$

and

$$P_3 - P_2 = S_2\, i - P_2 r \text{ becomes } 10 = 35i - 65r$$

This leaves the rate constants to be determined through the application of Cramer's Rule as follows:

$$i_{t,t+2} = \frac{\begin{vmatrix} \Delta P_{1,2} & P_1 \\ \Delta P_{2,3} & P_2 \end{vmatrix}}{\begin{vmatrix} S_1 & P_1 \\ S_2 & P_2 \end{vmatrix}}$$

and by expansion

$$i_{t,t+2} = ((18 \text{ x } -65) - (-47 \text{ x } 10))/((53 \text{ x } -65) - (-47 \text{ x } 35)) = -700/-1800 = 0.39$$

The numerator is obtained from the determinant of the system. The denominator is obtained by replacing the coefficient of the variable for which the system is being solved with the appropriate ΔP. So, recovery rate is similarly found as

$$r_{t,t+2} = ((53 \text{ x } 10) - (18 \text{ x } 35))/ ((53 \text{ x } -65) - (-47 \text{ x } 35)) = -100/-1800 = 0.05$$

The application of Cramer's Rule has limiting assumptions. The determinant of the system must not equal zero. For the model considered herein, this will occur if the change in prevalence equals zero, even if during such an equilibrium *i* and *r* are known by other means to be non-zero. Another assumption of Cramer's Rule is that the rates *i* and *r* remain constant over the sampling period for which estimates are sought. When interpolation is not used to find a midpoint prevalence, the midpoint prevalence cannot be a minimum or a maximum where that assumption would be unreasonable. Because the parameters obtained are discrete time estimates of a phenomenon that occurs in continuous time, it may be appropriate to refer to estimates obtained by this method as apparent incidence rate and apparent recovery rate.

Unique solutions for both *i* and *r* in *Eq. 2* cannot be determined from one change in prevalence. However, incidence rate may be determined from one change in prevalence if recovery rate is zero. The use of sentinel (tracer hosts) in parasitology represents such a special case. In this methodology, uninfected susceptible hosts are placed at risk to infection, and the change in their infection status after exposure is determined. The proportion of infections after exposure represent the incidence rate. With respect to *Eq. 2*, when $r = 0$ the term I*r* is dropped and incidence may be found as:

$$i = (P_{t+1} - P_t) / S_t \qquad \textit{Eq 4.}$$

Using contrived data where $P_1 = 47\%$, $P_2 = 65\%$,

$$i = (65\text{-}47) / 53 = 18/53 = .34$$

The underlying assumption of this estimate of incidence is that there are no hosts recovering from infection during the time bounded by the two prevalence estimates; an original prevalence of zero *is not* required for the use of tracer hosts. In circumstances where recovery rate is absent or considered to be insignificantly low relative to the sampling interval, this change in prevalence relative to the uninfected proportion of the population might be a useful approximator of the colonization process. Serologic surveys may provide such a circumstance, although sero-positivity is not necessarily an indication of infection. While serologic changes may not be permanent, the change from sero-positivity to sero-negativity is likely to meet the criteria of being slow relative to the sampling interval. The term relative change in prevalence is an appropriate name for the representation in *Eq 4* if the recovery rate can be assumed very small but cannot be known to be zero. If applied to circumstances where recovery occurs, this estimator of incidence is subject to a systematic bias proportional to both the magnitude of prevalence and the non-zero recovery rate.

DISCUSSION

Progress in studying the population ecology of parasitic helminths has been hampered by a failure to understand the relationship of the concept of incidence to theoretical epidemiology. This has led to the growing belief that estimating incidence rate requires conditions that cannot easily be met during investigations of helminth parasites in wild populations (Margolis et al. 1982, Roberts and Janovy 1999). As shown above, sequential changes in prevalence of infection, which can be obtained for populations of wild hosts, contain incidence information that may be possible to extract with the aid of simple epidemic models.

Models of simple epidemics have existed for more than 75 years (i.e. Lotka 1924, reprinted as Lotka 1956). Martini (1921, as cited by Lotka 1956) considered a variation on the simple epidemic model in the first quarter of the 20th century. It is worth stating that opportunities for noticing the relationship between changes in prevalence and the rate parameters of simple epidemic models have existed since at least that time. However, epidemiology focuses on human disease where incidence of disease is frequently estimated from physician and laboratory reports of new cases of disease. Perhaps because of this lack of need, using simple epidemic models to determine incidence rate from prevalence data gained no interest.

The historic association of parasitology with medicine and the adoption of epidemiologic terminology by parasitologists has obscured the relationship of their work with that of ecologists of free-living animals. The representation of the change in prevalence in *Eq 2* is clearly a form of the so-called patch metapopulation model originally introduced by Levins (1969). The epidemiologic terms used by parasitologists parallel terminology

used in metapopulation models; incidence rate is very similar to patch colonization rate, and recovery rate is similar to patch extinction rate. Knowledge of this similarity and an ability to estimate incidence and recovery rates for natural infections in wild animals will help parasite population ecologists to more fully participate in the field of metapopulation studies.

Obtaining estimates of incidence using the method suggested above must be done with some care. When applied to prevalence trajectories, the solution of *Eqs. 3* will generate estimates of both *i* and *r* that provide apparently good fits to the prevalence trajectory on which they are based, regardless of whether or not the SI model appropriately reflects the nature of the host-parasite system. Consequently, apparent goodness of fit of predicted values of prevalence to empirical measurements of prevalence is a necessary but insufficient criterion for acceptance of incidence estimators.

The biological goodness of incidence estimators must be considered in view of available information on the life-history of the host-parasite system under study. If the SI model were inappropriate for the system under study, so too would be incidence estimators obtained through its use. Susceptibility to infection in the host population may vary with age (size), gender, or other factors that might stratify the population relative to susceptibility making a simple epidemic model unacceptable. As in the familiar Lincoln-Peterson estimate of population abundance, measurement of changes in proportional representation in a population may be influenced by changes in population size. Where host births, deaths, or migration are significant, the SI model may again be unacceptable.

Sampling protocols used to obtain sequential estimates of prevalence must use the same care expected in all longitudinal studies of parasite occurrence. Again, the methods must consider the life-history of the specific host-parasite system under study. As the time between samplings becomes large the assumptions of constancy of *i* and *r*, and the absence of changes in the host population during sampling intervals become less tenable.

In conclusion, parasite ecologists have worked under the impression that *a priori* knowledge of the infection status of special individuals and special circumstances are needed to estimate incidence rate (Margolis et al. 1982, Roberts and Janovy, 1999). This has discouraged the measurement of incidence rate and, consequently, the study of metapopulation dynamics of helminth parasites in wild populations of hosts. At least in principle, two approximations of incidence may be obtained from a series of prevalence estimates. These estimates are suitable for naturally occurring epidemics in wild host populations that conform to simple epidemic models.

ACKNOWLEDGMENTS

Dr. John C. Holmes, University of Alberta, and Dr. Dennis Mick, Carroll College are thanked for their criticism during the early development of these ideas.

LITERATURE CITED

Esch, G.W. and C.W. Fernandez, 1993. A Functional Biology of Parasitism: Ecological and Evolutionary Implications. Chapman and Hall, London, 337 pp.

Dronen, N.O. 1978. Host-parasite population dynamics of *Haematoloechus coloradensis* Cort, 1915 (Digenea: Plagiorchidae). The American Midland Naturalist, 99 (2): 330-349.

Kline, W.E.; R. Oestrele; L.M. Willson. 1959. Foundations of Advanced Mathematics. American Book Company, New York.

Levins, R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. Bulletin of the Entomological Society of America. 15: 237-240.

Lotka, A.J. 1956. Elements of Mathematical Biology. Dover Publications, New York. 465 pp. (published in 1924 as Elements of Physical Biology)

Margolis, L., G.W. Esch, J.C., Holmes, A.M. Kuris, G.A. Schad. 1982. The use of ecological terms in parasitology (Report of an ad hoc committee of the American Society of Parasitology). Journal of Parasitology, 68 (1), 131-133.

Mausner, J.S. and A.K. Bahn 1974. Epidemiology, An Introductory Text. W.B. Saunders. Philadelphia. 377 pp.

Poole, R. W. 1974. An Introduction to Quantitative Ecology. McGraw-Hill, New York. 532 pp.

Price, P.W. 1980. Evolutionary Biology of Parasites. Princeton University Press. Princeton.

Price, P.W. 1987. Evolution in parasite communities. International Journal for Parasitology, 17, 209-214.

Roberts, L.S. and J. Janovy, Jr. 1999. Gerald Schmidt & Larry S. Roberts' Foundations of Parasitology. 6th Edition McGraw-Hill.

Simberloff, D. and J. Moore, 1997. Community ecology of parasites and free-living animals. In D.H. Clayton and J. Moore. Host Parasite Evolution: General Principles and Avian Models. Oxford University Press. Oxford. 473 pp.

Figure 1. Graphic representation of a Susceptible-Infected (SI) epidemic model. P=prevalence, the proportion of infected hosts.