Epidemics 5 (2013) 77–84

Contents lists available at SciVerse ScienceDirect

Epidemics

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

Food- and water-borne disease: Using case control studies to estimate the force of infection that accounts for primary, sporadic cases

G. Smith*

School of Veterinary Medicine, University of Pennsylvania, New Bolton Center, 382 West Street Road, Kennett Square, PA 19348, United States

ARTICLE INFO

Article history:
Received 2 July 2012
Received in revised form 22 February 2013
Accepted 29 April 2013
Available online 10 May 2013

Keywords:
Food-borne
Water-borne
Incidence rate
Force of infection
Odds ratio
Case control studies

ABSTRACT

Disease models which take explicit account of heterogeneities in the risk of infection offer significant advantages over models in which the risk of infection is assumed to be uniform across all hosts. However, estimating the incidence rate (force of infection) in the different at-risk (exposure) groups is no easy matter. Classically, epidemiologists differentiate groups of hosts with different infection-risks according to their exposure to putative explanatory risk factors. The importance of these risk factors is assessed by case-control studies, in which the measure of effect (the difference in disease occurrence between one population and another) is the odds ratio. This paper describes for the first time how – and under what circumstances – the incidence in these different exposure groups can be estimated from odds ratios derived from case control studies in which controls have been selected by density sampling. This new estimation technique can be applied to any transmission modality but is especially useful in the case of models of food- and water-borne disease for which the case control literature represents a vast and, as yet, untapped resource. The paper finishes with a worked example using one of the most common of all food- and water-borne pathogens, *Toxoplasma gondii*.

Introduction

Contemporary compartmental (SIR or SEIR) models of food- or water-borne infections have typically used a single parameter to represent the force of infection (incidence rate) attributable to an environmental source of infection (e.g. Liu et al., 2005; Brookhart et al., 2002; Eisenberg et al., 1998). Such models represent all hosts as being equally at risk of infection from this source and ignore the very rich epidemiological literature that demonstrates how much people (and animals) differ with respect to their likelihood of infection with food- and water-borne pathogens. It has long been appreciated that infectious disease models which take explicit account of heterogeneities in the risk of infection offer significant advantages over models in which the risk of infection is assumed to be uniform across all hosts (Anderson and May, 1991; Wallinga et al., 2006; Keeling and Rohani, 2008). However, estimating the incidence rate in the different at-risk groups is no easy matter especially in the case of food- and water-borne infections.

Epidemiologists differentiate groups of hosts with different infection-risks according to their exposure to putative explanatory risk factors. The importance of the risk factors is assessed using various measures of effect (the difference in disease occurrence between an exposed population and an unexposed population). For example, in a prospective cohort study of a food or water borne infection, the simplest measure of effect would be the relative risk (RR). If the incidence rates (instantaneous *per capita* rates of infection) in the exposed and unexposed groups are θ_1 and θ_2 respectively, then the relative risk is given by the risk of infection in the exposed cohort (c_1) divided by the risk of infection in the unexposed cohort (c_0), i.e.

$$\mathrm{RR} = \frac{c_1}{c_0} = \frac{1 - e^{-\theta_1 t}}{1 - e^{-\theta_0 t}} \tag{1}$$

where t is the duration of the study. Because the risk of infection in each cohort is measured directly it is straightforward to estimate the respective incidence rates ($\theta_1 = -\ln[1 - c_1]/t$ and $\theta_0 = -\ln[1 - c_0]/t$). Indeed, this is often claimed as one of the "advantages" of cohort studies in elementary epidemiology textbooks. However prospective cohort studies have a number of disadvantages: they are not suitable for rare diseases, they take a long time, and the magnitude of the measure of effect is a decreasing function of the duration of the study, which can make the strength of association between the risk factor and the infection rather difficult to evaluate For these reasons prospective cohort studies are found much less frequently in the literature than the more convenient case-control studies in which the measure of effect is the odds ratio (OR). There are specific circumstances in which the odds ratio is an unbiased estimator of the incidence rate

* Tel.: +1 610 925 6312; fax: +1 610 925 6830.
E-mail address: garys@vet.upenn.edu

1755-4365/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.epidem.2013.04.002

in the exposed group divided by the incidence rate in the unexposed group (Rothman, 1986; Pearce, 1993) but it is frequently emphasized that, even in such instances, while case-control study designs can be used to estimate the *ratio* of incidence rates, the absolute values of the individual incidence rates are unobtainable from the odds ratio. The purpose of this paper is to describe the conditions under which this admonition turns out to be wrong, and will explain how – and under what circumstances – the force of infection in different exposure groups can be estimated from the corresponding odds ratios. To the best of my knowledge, this methodology has not previously been described. The estimation method can be applied to any transmission modality but is especially useful in the case of models of food- and water-borne disease for which the case-control literature represents a vast and, as yet, untapped resource.

To anticipate the results a little, the methodology is valid only when dealing with endemic steady state infections that have been investigated with case-control studies that used density sampling. It also usually requires that there are independently derived estimates of (a) the equilibrium prevalence of infection and (b) the proportion of the host population exposed to the risk factors of interest. These assumptions are easily satisfied for many important food and water-borne infections. The methodology is not valid when dealing with epidemic infections – unless the infection is rare. The paper begins with a theoretical explication that justifies the assertion that there are circumstances in which it is possible to estimate the component incidence rates from an odds ratio, and concludes by showing how this can be applied to one of the most common of all food- and water-borne pathogens, *Toxoplasma gondii*.

Methods and results

Preamble

The method described below was developed primarily for use in the context of the transmission food- and water-borne infections, although it can be applied to any transmission modality. However, there are certain specific restrictions in the case of food- and water borne-disease that must be taken into account. These are dealt with first. Food- and water-borne infections in people are frequently the result of spill-over from the animal species populations in which the pathogen is maintained. Human cases are conventionally divided into those that arise as the result of infection from some environmental source (primary cases) and those that are more easily understood as the result of direct (or almost direct) horizontal transmission from other infected people (secondary cases). The latter are characteristic of institutional settings like long-term care facilities, nursing homes, and play groups. The primary cases are further subdivided into those that occur with no obvious spatial or temporal pattern (sporadic cases) and those that are interpretable as the result of some significant, identifiable point source contamination (a restaurant, for example). This paper is concerned with estimating the force of infection that leads to primary, sporadic cases.

Theoretical explication – odds ratios in case control studies of endemic food- and water-borne infections

We can estimate the force of infection (incidence rate) that accounts for sporadic cases of an endemic food- or water-borne infection in a group of hosts exposed to one or more risk factors from the odds ratios derived from case control studies provided the controls are selected by density sampling. In density sampling the controls are selected longitudinally throughout the course of the study as the cases are identified (Pearce, 1993). Nowadays, most case control studies involve density sampling (sometimes called 'risk set sampling' or 'sampling from the study base').

We proceed by creating a model for such a case control study. The important modeling assumptions are these: the system is at a dynamic equilibrium, and all "cases" are primary, "sporadic cases" – which is to say either there are no secondary cases or that secondary cases contribute negligibly to the main environmental source of infection and have been excluded during subject selection. Finally, we assume the force of infection attributable to a defined risk factor is constant. We also acknowledge the assumptions implicit in case control study design: all incident cases are recognized as cases and there is no bias (Miettinen, 1976). Given these conditions, the odds ratio (OR) derived from the data collected in the case control study is an unbiased estimator of the incidence rate in the exposed group divided by the incidence rate in the unexposed group. This is not a new observation (Rothman, 1986; Pearce, 1993) but, as this paper will demonstrate, given certain additional information it is possible to use this odds ratio to find the respective numerical values of the incidence rate in the exposed group and the incidence rate in the unexposed group.

With regard to primary, sporadic cases of a food or water borne infection, the usual SIR model for an endemic situation (considered at equilibrium) applies. The model is modified slightly to recognize heterogeneity with respect to exposure to different risk factors. In a case control study involving n specified risk factors we would have an estimated odds ratio (OR_x, $x = 1,\ldots, n$) for each risk factor (RF_x). Hosts would be divided into m exposure groups according to how many risk factors were considered. For example, if the study considered two risk factors there would be four exposure groups (those exposed to neither RF_1 nor RF_2, those exposed to RF_1, those exposed to RF_2, and those exposed to both RF_1 and RF_2). We designate these exposure groups by the subscript, h. In the example just given, $h = 0,\ldots, 3$ respectively. The transmission dynamics of each exposure group is most simply represented by

$$\begin{aligned}\frac{dS_h}{dt} &= \mu N p_h - \theta_h S_h - \mu S_h\\ \frac{dI_h}{dt} &= \theta_h S_h - \mu I_h - \gamma I_h\\ \frac{dR_h}{dt} &= \gamma I_h - \mu R_h\end{aligned} \tag{2}$$

Here θ_h is the force of infection (incidence rate) in exposure group h, N is size of the entire population (all exposure groups combined), μ is the instantaneous *per capita* birth and death rates, γ is the recovery rate and p_h is the proportion of N in exposure group h.

In case control studies of endemic disease, we collect incident cases. That is we attempt to identify all the cases arising during the interval of observation, t, and subsequently decide which were exposed to the risk factor of interest and which were unexposed. At the same time, we sample a fraction (k) of the non-cases (controls) and similarly assign members of this sample to the exposed or unexposed categories. In a study involving just a single risk factor (and therefore just two exposure groups ($h = 0$ or 1)) we would calculate the odds ratio as follows:

$A + C$ = number of incident cases identified. We find later that A subjects were in the exposed group and C subjects were in the unexposed group.
$B + D$ = the number of non-cases (controls) obtained by sampling a fraction, k, of all non-cases. We find later that B subjects were in the exposed group and D subjects were in the unexposed group.

The odds ratio (OR) is estimated as

$$\mathrm{OR} = \frac{AD}{BC} \tag{3}$$