Infectious Diseases of Humans

Dynamics and Control

ROY M. ANDERSON

Linacre Professor,
University of Oxford

and

ROBERT M. MAY

Royal Society Research Professor,
University of Oxford

1991

OXFORD
UNIVERSITY PRESS

with infections that are indirectly transmitted by arthropod vectors (Chapter 14). Both these chapters contain a significant amount of new material.

Macroparasitic infections are examined in Part II. This section starts with Chapter 15, a discussion of life cycles, general biology, and observed epidemiological patterns. This is followed by the development of a basic model to describe the transmission of macroparasitic infection in Chapter 16 (statics) and Chapter 17 (dynamics). Various summary statistics are introduced, and—extending previous analyses—the problems of control by chemotherapy and vaccination are discussed in these two chapters. The topic of acquired immunity is developed in Chapter 18, with a review of experimental facts and pertinent models. In particular, Chapter 18 presents some new ideas about the balance between age-related exposure to infection and the acquisition of immunity, and how this balance may determine the patterns of infection observed in human communities. Chapter 19 analyses a variety of refinements and complications introduced by heterogeneities in transmission. Problems associated with indirect transmission of infection via intermediate host species are addressed in Chapter 20, which combines reviews of existing work with some new ideas and analysis. Chapter 21 gives an evangelical account of the contributions to our understanding of the population dynamics of macroparasites that have been, and can be, made by experimental studies of host models in the laboratory. Chapter 22 deals with the evolution of drug resistance by macroparasitic populations, presenting some new work bearing on this important practical problem. This chapter also develops some interesting parallels between the evolution of drug resistance and the selective effects exerted on parasites by the host's immune system.

The book ends with a discussion in Chapter 23 of broader issues. These include the role of infectious disease agents as regulators of the abundance of their host population, coevolution amongst host and parasite populations, the importance of an ecological approach to epidemiological study, and future research needs and directions.

2

A framework for discussing the population biology of infectious diseases

By focusing on the overall population biology of associations between hosts and parasites, and emphasizing broad themes that are common to most systems, we seek to provide a framework within which a vast amount of information about parasitic infections may be organized in an orderly way. We aim to codify similarities and differences among the various viral, bacterial, protozoan, fungal, and helminth parasites, identifying the ecologically based patterns of relationships among epidemiological parameters such as transmission rates, virulence, life span of the parasite within the host, and so on.

To this end, the present chapter outlines some basic concepts. We begin by making a distinction between 'microparasites' and 'macroparasites'; this rough dichotomy cuts across conventional taxonomic lines to focus on the population biology of the parasite. Next we discuss the idea of the basic reproductive rate of the parasite, and then go on to outline some ideas about threshold host densities and about modes of transmission.

2.1 Microparasites and macroparasites

Microparasites may be thought of as those parasites which have direct reproduction—usually at very high rates—within the host (Anderson and May 1979*b*). They tend to be characterized by small size and a short generation time. Hosts that recover from infection usually acquire immunity against reinfection for some time, and often for life. Although there are important exceptions, the duration of infection is typically short relative to the expected life span of the host. This feature, combined with acquired immunity, means that for individual hosts microparasitic infections are typically of a transient nature. Most viral and bacterial parasites, and (in a more equivocal way) many protozoan and fungal parasites, fall broadly into the microparasitic category.

For such infective agents, it makes sense to divide the host population into relatively few classes of individuals: susceptible, infected, recovered-and-immune. Such a *compartmental model* for the dynamic interaction between parasitic and host populations is depicted schematically in Fig. 2.1. Our operational definition of a microparasite is, indeed, an organism whose population biology can to a sensible first approximation be described by some such compartmental model.

Greater detail and realism can be achieved by adding more compartments or categories to the model (for example, a class of latent, but not yet infectious, individuals). The essential feature of these compartmental models, however, is

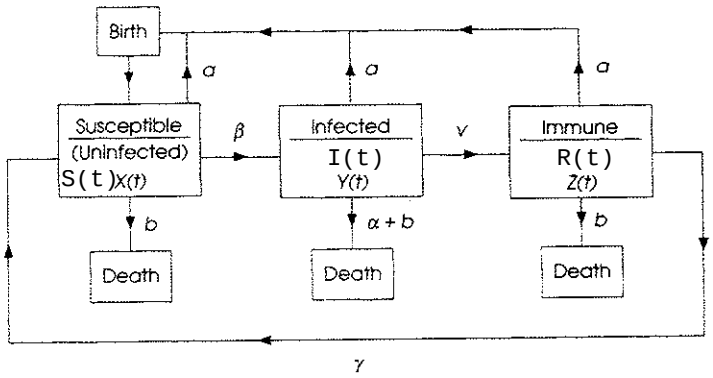

Fig. 2.1. Schematic representation of the flow of hosts between susceptible ($X(t)$), infected ($Y(t)$) and immune ($Z(t)$) classes, which records the dynamic interaction between a directly transmitted microparasite and its host population. In this diagram hosts reproduce at a per capita rate a and die at a per capita rate b. The infected hosts experience an additional death rate α, induced by microparasitic infection. The average durations of stay in the infected and immune classes are denoted by $1/v$ and $1/\gamma$, respectively. The transmission coefficient which determines the rate at which new infections arise as a consequence of mixing between the susceptible and infected individuals is defined by β.

that little or no account is taken of the degree of severity of the infection (i.e. the abundance of the parasite within the host); individuals either 'have measles' or they do not. In other words, the reality of infected individuals with differing nutritional, environmental, or genetic status is replaced by the simplified abstraction of some average 'infected' or 'immune' individual.

In addition to the distinction between infected and immune hosts, it is often desirable to distinguish between infection and disease. Thus, for example, in the literature concerned with microparasitic infections of humans, the period from the point of infection to the appearance of symptoms of disease is termed the incubation period. The duration of symptoms of disease, as illustrated diagrammatically in Fig. 2.2, is not necessarily synchronous with the period during which an infected host is infectious to susceptible individuals. Furthermore, a host may be infected but not yet infectious. The period from the point of infection to the beginning of the state of infectiousness is termed the latent period. With respect to the ecology of parasite transmission the sum of the average latent and average infectious periods is referred to as the average generation time of the infection (Fig. 2.2).

Most of the epidemiological and demographic parameters in Fig. 2.1—host birth and death rates, disease-induced death rates, recovery rates, rates of loss of immunity—can be measured directly by appropriate studies. The transmission rate, however, combines many biological, social, and environmental factors, and (as discussed in the next chapter) is thus rarely amenable to direct

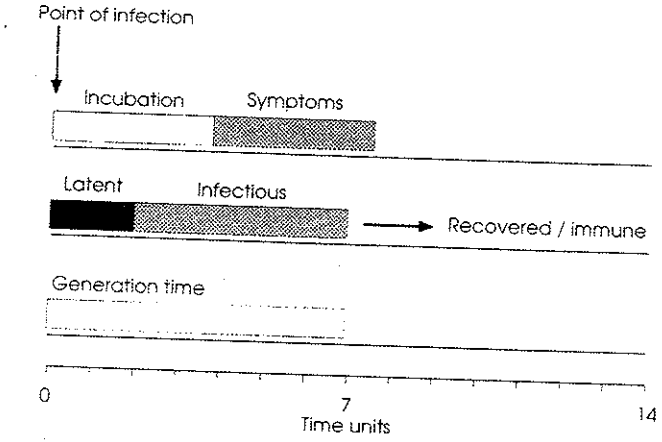

Fig. 2.2. Diagrammatic illustration of the relationship between the incubation, latent, and infectious periods for a hypothetical microparasitic infection. Note that the infectious period and the duration of symptoms of disease are not necessarily synchronous.

measurement. Indeed, it will often be that the best way to assess transmission rates is to infer them indirectly from data on population-level processes, as discussed below.

Macroparasites may be thought of as those having no direct reproduction within the definitive host (Anderson and May 1979*b*). This category embraces most parasitic helminths and arthropods. Macroparasites are typically larger and have much longer generation times than microparasites, with the generation time often being an appreciable fraction of the host life span. When an immune response is elicited, it usually depends on the past and present number of parasites harboured by the host, and it tends to be of a relatively short duration once parasites are removed (for example, by chemotherapy) from the host. Thus macroparasitic infections are typically of a persistent nature, with hosts being continually reinfected.

For macroparasites, the various factors characterizing the interaction—egg output per female parasite, pathogenic effects on the host, evocation of an immune response in the host, parasite death rates, and so on—can all depend on the number of parasites in a given host. Mathematically, this means that the relatively simple compartmental models must be replaced by more complicated models which take full account of the distribution of parasites among the hosts. Such a *distributional model* is shown schematically in Fig. 2.3. Operationally, our definition of a macroparasite is one whose population biology requires such a full description of the distribution of parasites among hosts.

Macroparasites are rarely, if ever, distributed in an independently random way among their hosts, but rather show an aggregated or 'clumped' distribution with often a minority of the host population harbouring the majority of the

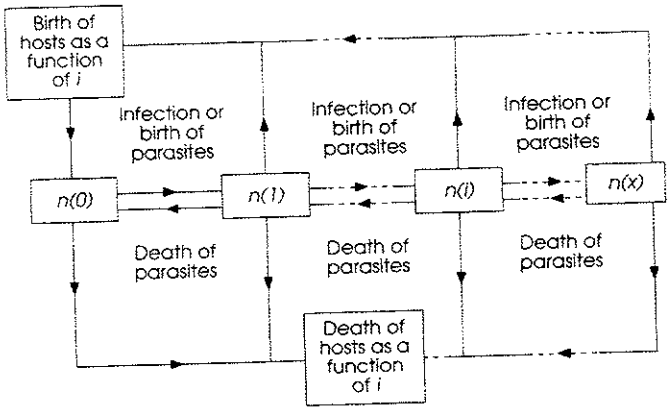

Fig. 2.3. Diagrammatic flow chart for a directly transmitted macroparasitic infection, based on a model with compartments for the number of hosts, $n(i)$, harbouring i parasites ($i = 0, 1, 2, \ldots$). The birth and death rates of the host population are denoted as functions of parasite burden.

parasite population. It is not uncommon to find 80 per cent or more of the macroparasites contained in 20 per cent or fewer of their human hosts.

From a public health viewpoint, it is significant that for many macroparasites a distinction can be made between infection (harbouring one or more parasites) and disease (harbouring a parasite burden large enough to produce pathogenic symptoms, or even host death). For a canonical microparasitic infection, such as smallpox, it is reasonable to assume that a given host does, or does not, 'have smallpox'; for a macroparasitic infection, such as hookworm or *Ascaris* or schistosomiasis, there is a real distinction between being infected with one or two worms, and carrying a worm burden large enough to cause illness.

The division into microparasites and macroparasites, whether made on biological or mathematical grounds, is necessarily a rough one. The distinction essentially corresponds to the extremes of a continuum, Many parasites are not easily forced into this dichotomous scheme. A lot of protozoan parasites, for instance, may to a good approximation have their epidemiology described by the compartmental models characteristic of microparasites, while on the other hand their patterns of persistence within the host population (with hosts continually being reinfected) are characteristic of macroparasites. Thus, as discussed later, an account of the degree of immunity that can naturally be elicited against malaria requires a model intermediate between the two extremes.

In short, the paradigmatic notions of microparasite and macroparasite are deliberate oversimplifications, which aim to emphasize the population dynamics of host–parasite associations and to de-emphasize conventional taxonomic categorizations. Realistic refinements can be grafted on, layer by layer, as we proceed from basic understanding to detailed application.

2.2 Basic reproductive rate of a parasite

The basic reproductive rate, R_0, is essentially the average number of successful offspring that a parasite is intrinsically capable of producing (Macdonald 1952; Dietz 1975, 1976; Yorke *et al.* 1979; May and Anderson 1979; Anderson 1981*b*; Anderson and May 1982*d*). It is, in effect, Fisher's (1930) 'net reproductive value' for the parasite. This concept is central to any discussion of the overall population biology of an organism. Clearly a parasitic species must have $R_0 > 1$ if it is to be capable of invading, and establishing itself within, a host population.

For a microparasite (represented by a compartmental model), R_0 is more precisely defined as the average number of secondary infections produced when one infected individual is introduced into a host population where everyone is susceptible.

When such a microparasitic infection becomes established in a host population, the fraction remaining susceptible decreases. Eventually an equilibrium may be attained, with the rate at which susceptible individuals are infected being balanced against a rate at which newly susceptible individuals appear (usually by birth, but possibly also by immigration or by loss of immunity). At equilibrium, each infection will on average produce exactly one secondary infection; that is, at equilibrium the effective reproductive rate of the parasite is $R = 1$. If we assume the host population is homogeneously mixed, in the sense that, on average, all hosts have intrinsically similar epidemiological properties (independent of age, genetic make-up, social habits, geographical location, etc.), then the number of secondary infections produced by an infected individual will be linearly proportional to the probability that any one random contact is with a susceptible individual. In this event, the effective reproductive rate, R, is equal to the basic rate, R_0, discounted by x, the fraction of the host population that is susceptible: $R = R_0 x$. Thus, under the rough approximation of treating the host population as homogeneously mixed, for a microparasite the equilibrium condition $R = 1$ leads to an important relation between R_0 and the fraction, x^*, of the host population that is susceptible at equilibrium:

$$R_0 x^* = 1. \tag{2.1}$$

Equation (2.1) has applications which will be discussed later. Beyond this, it has general interest for ecologists. For one thing, it is notoriously difficult to assess the intrinsic reproductive capacity, R_0, of any species of organism (even humans). It is also a problem for ecologists to determine exactly what density-dependent mechanisms operate to hold the reproductive rates of natural populations below their intrinsic capacities—capacities which could, if realized, blanket the world with that species. Equation (2.1) resolves both these problems for microparasites: because x^* can be found from serological or other data on age-specific susceptibility, the elusive quantity R_0 can be calculated (see Table 4.1, p. 70); and the density-dependent process holding R below R_0 is

simply the removal of susceptibles by immunity, following infection (which corresponds, in essence, to a very simple form of predation upon hosts by the parasite). Thus eqn (2.1) and the surrounding discussion illustrate the concepts of basic reproductive rates and density-dependent regulation of populations more clearly and quantitatively than the examples conventionally used in introductory biology textbooks and courses.

Unfortunately, eqn (2.1) depends on the assumption of homogeneous mixing. When age-related differences in transmission rate, and other inhomogeneities are acknowledged, some of the simplicity of eqn (2.1), and possibly also of the definition of R_0, can be lost. But the ideas presented here still underpin the more detailed discussion of such complications in Chapters 8–12.

For a macroparasite (represented by a distributional model), R_0 is the average number of female offspring produced throughout the lifetime of a mature female parasite, which themselves achieve reproductive maturity in the absence of density-dependent constraints.

The factors governing equilibrium in a host–macroparasite system are more complex than those in host–microparasite systems. In the absence of any density-dependent constraints, a macroparasite with $R_0 > 1$ could attain arbitrarily high population levels, as hosts with high parasite burdens continued to put out large numbers of eggs, leading to yet higher parasite burdens per host. In reality, various kinds of density-dependent processes intervene to halt such an exponential run-away: egg output per parasite declines as the number of parasites in a host increases; hosts with high burdens may be less likely to acquire further infections, for a diversity of possible reasons; parasite death rates may increase as the number of parasites in a host increases; the overall transmission rate may saturate to some upper limit when the parasite population is large; and a high burden may simply kill the host. Precisely which of these density-dependent factors, individually or in combination, will be primarily responsible for establishing equilibrium is likely to vary from one host–parasite association to the next. Examples are pursued in detail in Part II.

So far, we have dealt with the differing details of the definition of R_0 for microparasites and macroparasites, with the way density-dependent effects keep effective reproductive rates below R_0, and with ways of inferring the magnitude of R_0 from observations around equilibrium. There are, however, situations where we can observe a parasite in the initial phases of population growth, where density-dependent limitations have not yet come to be significant. Such situations arise, for instance, when a measles epidemic starts to spread among an island population who have not experienced measles for a generation or more (so that essentially all are susceptible), or when a new infection such as the human immunodeficiency virus (HIV, the aetiological agent of AIDS) begins its exponential rise among the population at risk, or when hookworm, *Ascaris*, or other helminths begin to re-establish themselves after being removed from a population by chemotherapy.

The initial, exponential rise in the proportion of hosts who are infected,

$P(t)$, depends on the rate at which new infectives are being produced, Λ:

$$P(t) = P(0)\exp(\Lambda t). \tag{2.2}$$

For both microparasites and macroparasites, Λ depends on R_0 and on the parasite lifespan, D:

$$\Lambda = (R_0 - 1)/D. \tag{2.3}$$

For microparasites D is essentially the duration of infectiousness, while for macroparasites it is the average life expectancy of adult parasites. The result (2.3) is established in detail for microparasites and macroparasites, separately in Chapters 6 and 16, respectively. The result, however, may be explained intuitively: each infection produces R_0 new infections in its lifetime, of duration D, and then dies; thus $R_0 - 1$ net infections are added in an interval D, corresponding to $(R_0 - 1)/D$ per unit time. As for eqn (2.1), eqn (2.3) depends on the assumption of homogeneous mixing and more complicated expressions are obtained once various kinds of inhomogeneities are taken into account. Nevertheless, the above discussion indicates how those rare instances where parasite populations are growing free from density-dependent checks can be used to estimate R_0, provided D is known.

2.3 Threshold host densities

In many simple circumstances, R_0 may be taken to be linearly proportional to the total number or density of hosts that are candidates for infection, whence

$$R_0 = N/N_T. \tag{2.4}$$

Here N is the host population size, and the proportionality constant (which subsumes all manner of biological, social, and environmental aspects of transmission) has been written as $1/N_T$. The condition $R_0 > 1$ for establishment of the parasite thus translates into the requirement that the host population, N, exceeds a certain threshold magnitude, N_T:

$$N > N_T. \tag{2.5}$$

More generally, R_0 is likely to be some non-linear function of N, $R_0 = f(N)$. As discussed in Chapter 9, data for measles and pertussis in Britain, for example, can be seen to obey a power-law relation $R_0 = (N/N_T)^\nu$, with $\nu < 1$. Likewise for macroparasites various effects can lead to R_0 increasing less fast than linearly with increasing host density. The criterion $R_0 > 1$, however, will still usually lead to a threshold condition, eqn (2.5); the essential idea of a threshold host density retains its validity.

An important class of exceptions are those infections that are transmitted by intimate contact with a defined group of people, as happens for sexually transmitted diseases, infections associated with sharing of needles by drug abusers, and the like. Here R_0 depends on the average rate at which new

partners are acquired, which usually has no direct dependence on N; doubling the population size does not affect sexual habits, except perhaps indirectly through social changes precipitated by greater crowding. Sexually transmitted parasites, which often produce long-lasting infections and do not induce acquired immunity in recovered hosts, can be admirably adapted to persist in low-density populations of promiscuous hosts; here a threshold average number of sexual partners replaces the threshold host density of eqn (2.5).

So far, our discussion of threshold densities has focused on the condition $R_0 > 1$. If populations were always large enough for us to ignore stochastic and other effects associated with the inconvenient fact that humans are born, infected, and die in integer units (rather than in a continuous and deterministic fashion), we would need go no further. But there are at least two effects that can lead to threshold densities in practice being higher than estimated from $R_0 > 1$ alone.

The first such effect has to do with the necessity for the equilibrium number of infected individuals to be, on average, large enough that statistical fluctuations are unlikely to break the chain that maintains the infection. In the simple case of a microparasitic infection in a homogeneously mixed population of susceptibles, infecteds, and immunes, the number of latent plus infected hosts at equilibrium, $H^* + Y^*$, is roughly (see Chapter 6)

$$H^* + Y^* = (1 - 1/R_0)[(D + D')/L]N. \qquad (2.6)$$

Here, as before, D is the average duration of infectiousness, D' is the average duration of latency, L is the average life expectancy of a host, and N is the total number of hosts. We require that $H^* + Y^*$ should be significantly greater than unity, so that statistical fluctuations are not likely to carry the number of infected hosts to zero and extinguish the infection from the population. This means that N must be significantly greater than $L/(D + D')$. As L is measured in decades, and D and D' typically in days or weeks, this can require N to be of order 10^5 or more (which in some cases may be larger than the N_T of eqn (2.5)). We think it is useful to call the above effect *endemic fade-out*.

Even more severe constraints can be put on population size by the severe oscillations in numbers of susceptibles and infectives that characterize the introduction of a microparasite into a previously unexposed population. As discussed more fully in Chapter 6, the number of infected individuals falls to very low levels in the wake of the first epidemic. Stochastic effects may well lead to 'fade-out' of the infection during this epidemic phase of parasite establishment, thus preventing the infection from damping to the more easily maintained endemic state. We call this phenomenon *epidemic fade-out*. On the one hand, the number of infections in the trough following the epidemic introduction will be substantially less than given by eqn (2.6). On the other hand, the parasite need only survive in a few such troughs, and in this respect needs a narrower margin to cushion it against stochastic extinction during these relatively short episodes than it does over the long haul in the endemic state.

Whether such stochastic fade-out in the epidemic phase puts more, or less, stringent requirements on N than does stochastic extinction in the endemic state (via eqn (2.6)) will depend on the details for a specific microparasite.

In short, the threshold density may be governed by deterministic factors ($R_0 > 1$), or by one or other of a variety of stochastic considerations. Whatever the determining factor, a direct assessment of the threshold density is usually difficult. Some useful generalizations can nevertheless be made. Many microparasitic infections, such as smallpox or measles, are of very short duration and have relatively low transmission efficiencies; that is, the transmission stages are short-lived in the external environment and fairly direct contact is required to produce infection. In this event, a large population of hosts will be needed before R_0 can exceed unity, and thus the threshold density for the host population, N_T, will be large. In other words, such infections require a large population of hosts in order to give birth to new susceptibles at a rate that keeps pace with their loss by infection. Specifically, it has been estimated that N_T for maintaining measles in human communities is around 300 000 individuals or more (Bartlett 1957; Black 1966).

Conversely, the reproductive life span of many macroparasites within a host is an appreciable fraction of the host's life, and transmission pathways are often quite efficient, involving intermediate vector hosts or long-lived transmission stages. Threshold host densities for the maintenance of macroparasitic populations can thus be small.

These notions have some general implications that may be of interest to animal ecologists. In so far as directly transmitted microparasites typically require high host densities in order to persist, they should more commonly be associated with animals which exhibit herd or schooling behaviour, or which breed in large colonies. A certain amount of anecdotal support for these ideas comes form the observed abundance of such infections within modern human societies, large herds of ungulates, breeding colonies of seabirds, and communities of social insects. Humans get microparasitic infections, such as colds, while relatively solitary carnivores such as tigers are afflicted more by macroparasites, such as intestinal worms. But there is a need for comparative studies in which data are compiled systematically to test these ecological ideas.

2.4 Direct and indirect transmission

The above ideas about the essential character of the parasite and about its overall reproductive rate are valid, independent of the details of the transmission process. We conclude this chapter by considering some of the different ways parasites may be transmitted, which are summarized schematically in Fig. 2.4.

For *direct* transmission, depicted in Fig. 2.4(a), the transmission stages of the parasite pass directly from one host to the next. Some parasites, including many viruses and bacteria, pass by direct contact between hosts or in vapour droplets.

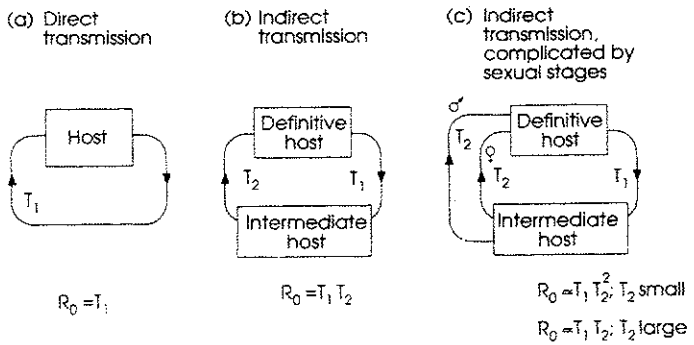

Fig. 2.4. Diagrammatic representation of direct and indirect transmission and the complications introduced by the sexual stages of macroparasitic organisms. The quantities T_1 and T_2 denote summary transmission parameters for the flow of parasites from definitive host to intermediate host (T_1) and intermediate host to definitive host (T_2). (See text for details.).

Other parasites, such as hookworm and many other helminths, have free-living transmission stages. These transmission stages may be capable of surviving for a long time in the external environment, as is the case, for example, for anthrax and the nematode *Ascaris*; in this event, the transmission rate depends not simply on the number of current infections, but rather on a summation over past infections (appropriately discounted as one goes further back in time).

As indicated in Fig. 2.4(a), for direct transmission the various factors are concatenated into an overall transmission factor T_1, which gives R_0.

When transmission is *indirect*, as illustrated in Fig. 2.4(b) and (c), the parasite (micro or macro) passes through one or more species of intermediate host in order to complete its life cycle. In simple cases, as indicated in Fig. 2.4(b), the basic reproductive rate will be a product of the factors involved in transmission from the definitive to the intermediate host, T_1, and those from the intermediate back to the definitive host, T_2: $R_0 = T_1 T_2$.

The dynamics of such indirect transmission systems can be messier than comparable systems with direct transmission, because changes in the levels of infection in the definitive host population can lead to changes in the levels of infection among intermediate hosts, and so on. Often, however, the characteristic time-scales for population change in intermediate host populations (typically arthropods) are significantly shorter than those for definitive, human host populations, so that intermediate vectors may be regarded as effectively at the equilibrium appropriate to the prevailing host state.

The buffering effect of a reservoir of intermediate hosts can often mean that $R_0 > 1$ even though the definitive host population is relatively small; the intermediate host population can compensate for the relative paucity of definitive hosts in the overall combination $T_1 T_2$ that gives R_0.

Figure 2.4(c) illustrates an additional complication that arises for the many macroparasites with a sexual stage in the definitive host. In this case, as first emphasized by Macdonald (1965) for the schistosome flukes, it is necessary to have a mated pair of adult macroparasites in the definitive host. If the average level of transmission from intermediate to definitive host, T_2, is low, then the probability of having one mated pair scales approximately as T_2^2, and $R_0 \simeq T_1 T_2^2$. Conversely, at high transmission levels, a host is likely to acquire several parasites, in which case these complications are not important and $R_0 \simeq T_1 T_2$. These effects, and their possible implications for control strategies, are discussed more fully in Chapter 20.

2.5 Summary

A framework for studying host–parasite associations is outlined, placing emphasis on population biology rather than on taxonomy. The broad categories of microparasites and macroparasites are distinguished on biological grounds, and defined in terms of the structure of the mathematical models used to study them (compartmental and distributional, respectively). The basic reproductive rate, R_0, of a parasite is defined, and some of the density-dependent processes which in practice constrain reproductive rates below R_0 are sketched. We show how threshold host densities for maintenance of infection may arise deterministically (from $R_0 > 1$) or stochastically (from either epidemic or endemic fade-out) and we speculate on some possible ecological implications. The chapter ends with a brief review of various kinds of direct and indirect transmission of infection, noting the implications for R_0 and threshold densities.