6.14 Mathematical models of transmission and control

Roy Anderson and D. James Nokes

Introduction

The aim of this chapter is to show how simple mathematical models of the transmission of infectious agents within human communities can help to aid the interpretation of observed epidemiological trends, to guide the collection of data towards further understanding, and to help in the design of programmes for the control of infection and disease. A central theme is to improve understanding of the interplay between the variables that determine the typical course of infection within an individual and the variables that control the pattern of infection and disease within communities of people. This theme hinges on an understanding of the basic similarities and differences between different infections in terms of the number of population variables (and consequent equations) needed for a sensible characterization of the system, the typical relations between the various rate parameters (such as birth, death, recovery, and transmission rates), and the form of expression that captures the essence of the transmission process.

Model construction, whether mathematical, verbal, or diagrammatic, is in principle the conceptual reduction of a complex biological or population-based process into a more simple idealized and easily understandable sequence of events. Consequently, the use of mathematical modelling as a descriptive and interpretative tool is a very common exercise in scientific study. Its use, therefore, in epidemiological study should not be viewed as intrinsically difficult or beyond the comprehension of those trained in medical or biological disciplines. The reductionist approach, inherent in model construction, which helps to define processes clearly and identify the most important components of a system, is employed in many areas of public health research and practice. The following situations, for example, are all likely to involve, at the very least, the implicit use of models to simplify and aid understanding: the assessment of the cause and severity of sporadic epidemics of *Salmonella* or hepatitis A virus food poisoning or Legionnaires' disease; the cost-effectiveness analysis of various measures used to combat an infection within a hospital, within a community, countrywide, or globally; or the identification of the factors that control the maintenance of an endemic infection within a community.

Most epidemiological problems, by definition, are concerned with the study of populations and so involve quantitative scores of, for example, abundances and rates of spread. Thus it is invariably necessary to convert any descriptive model of process into a more formal mathematical framework so that we work with numbers and not words. The use of a more formal structure enables us to incorporate quantitative estimates of abundances or rates, derived from experiment or field observations, into the model and to make predictions of the likely behaviour of the system under varying conditions, particularly when we are concerned with the introduction or alteration of measures to control infection or disease.

It is the step of translation from verbal or diagrammatic description into a formal mathematical framework that arouses the deepest suspicions amongst medical or public health workers. Quite naturally this response is in part a consequence of the use of, what is to many, a strange symbolism to describe familiar verbal or conceptual identities. It must be remembered, however, that mathematics is the most precise language we have available for scientific study and once a problem is formulated in mathematical terms many techniques are available to pursue the logical consequences of the stated assumptions. The clear and unambiguous statement of assumptions is of course a particular attribute of mathematical, as opposed to verbal, description. Excessive use of symbolism or formal methods of analysis can confuse as opposed to clarify and it must be admitted that some sections of the mathematical epidemiological literature have drifted from their original moorings and sail free from the constraints of data or relevance. But to jump from this observation to the belief that mathematical models have nothing to contribute in practice to the design of public health programmes is a mistake. Sensibly used, mathematical models are no more and no less than tools for thinking about things in a precise way.

The second area of suspicion, aside from symbolism, concerns simplification. A frequent criticism of mathematical work in epidemiology is that model formulation involves too many simplifying assumptions despite known biological complexity. This is often true, and needs to be remedied, but it is in part a consequence of the infancy of the discipline and, in some cases, a result of inadequate quantitative understanding of a particular problem. There are, however, two important counter-arguments to the criticism of simplification. Firstly and most importantly, it is often the case in biological study that a few processes dominate the generation of observed pattern despite the fact that many more can, to a lesser degree, influence the outcome. The identification of the dominant processes is an important facet of model construction and, what is termed, sensitivity analysis. The second point concerns scientific method. The process of understanding the consequences of a series of simple assumptions and building upon this by slowly adding complexity is directly analogous to the laboratory scientist's approach of carefully controlling most variables and allowing a few to vary in a planned design. Carefully building complexity on a simple framework can greatly facilitate our

Anderson R, Nokes DJ. Mathematical models of transmission and control (ch. 6.14). In: Detels R, McEwen J, Beaglehole R and Tanaka H.(eds). Oxford Textbook of Public Health. New York: Oxford University Press. Fourth edition Volume 2. 2002

understanding of the major factors that influence or control a particular process or pattern.

The chapter is organized as follows. The second section following this introduction provides a brief review of the historical development of mathematical epidemiology and outlines the types of infection that will be considered in latter sections. The third section addresses the problems of model construction, design, and application. The fourth section examines the major concepts in quantitative epidemiology that have been derived from mathematical study, such as threshold host densities for the persistence of an infection, the basic reproductive rate, and herd immunity. In the fifth section methods are explored by which to obtain some of the basic epidemiological parameters from empirical observation. The sixth section turns to applied problems and considers the use of models in the design of control strategies for infection and disease, and the final section is reserved for concluding thoughts. Throughout, mathematical details are kept to a bare minimum and the reader interested in technical details of model construction and analysis is referred to papers in specialist journals.

Historical perspective

The application of mathematics to the study of infectious disease appears to have been initiated by Daniel Bernoulli in 1760 when he used a mathematical method to evaluate the effectiveness of the techniques of variolation against smallpox (Bernoulli 1760). Further interest did not occur until the middle of the nineteenth century when, in 1840, William Farr effectively fitted a normal curve to smoothed quarterly data on deaths from smallpox in England and Wales over the period 1837 to 1839 (Farr 1840). This empirical approach was further developed by John Brownlee (1906) who considered in detail the 'geometry' of epidemic curves. The origins of modern mathematical epidemiology owe much to the work of Hamer, Ross, Soper, Kermack, and McKendrick who, in different ways, began to formulate specific theories about the transmission of infectious disease in simple but precise mathematical statements and to investigate the properties of the resulting models (Ross 1911; Kermack and McKendrick 1927; Soper 1929). The work of Hamer (1906), Ross (1911), Soper (1929), and Kermack and McKendrick (1927) led to one of the cornerstones of modern mathematical epidemiology via the hypothesis that the course of an epidemic depends on the rate of contact between susceptible and infectious individuals. This led to the so-called 'mass-action' principle in which the net rate of spread of infection is assumed to be proportional to the density of susceptible people multiplied by the density of infectious individuals. In turn this principle generated the celebrated threshold theory according to which the introduction of a few infectious individuals into a community of susceptibles will not give rise to an epidemic outbreak unless the density or number of susceptibles is above a certain critical value (see the review by Fine (1993)).

Since these early beginnings the growth in the literature has been very rapid and reviews have been published by Bailey (1975), Becker (1979), Anderson and May (1985*c*, 1991), Dietz (1987), and Scott and Smith (1994). In more recent work there has been an emphasis on the application of control theory to epidemic models (Wickwire 1977), the study of the spatial spread of the disease (Cliff *et al.* 1993), the investigation of the mechanisms underlying recurrent epidemic behaviour (Anderson and May 1982), the importance of heterogeneity in transmission (Anderson and May 1985*a*), the formulation of stochastic (probabilistic) models (Ball 1983), the formulation of models for indirectly transmitted infections with complex lifecycles (Anderson and May 1985*b*; Rogers 1988), the study of sexually transmitted infections such as gonorrhoea and HIV (Hethcote and Yorke 1984; Anderson *et al.* 1986; May and Anderson 1987), and the development of models for infectious agent transmission in developing countries with positive net human population growth rates (Anderson *et al.* 1988; McLean and Anderson 1988*a*,*b*). Such theoretical work is beginning to play a role in the formulation of public health policy (Babad *et al.* 1995) and the design of control programmes (Nokes and Anderson 1991) but there is a need in future work for greater emphasis on data-oriented studies that link theory with observation.

In the following sections we attempt to give a flavour of recent work and to distil the major conclusions that have emerged in particular areas. We have deliberately chosen to concentrate on directly transmitted viral and bacterial infections that constitute the major infectious diseases of children in developed countries and, as a consequence of the recent pandemic of AIDS, sexually transmitted infections. Our reasons are simply that the mathematical models are more highly developed in these fields by comparison with others (e.g. vector-borne infections), that theory has close contact with empirical epidemiological data in these areas, and that model structure is somewhat simpler than for other infections such as metazoan parasites.

Model construction

Definition of terms

Epidemiology

Epidemiology as a subject is concerned with the study of the 'behaviour' of an infection or disease within a population or populations of hosts (= humans). 'Behaviour' refers to observed patterns such as the incidence (the rate at which new cases arise or are reported) of infection or disease. Examples of 'behaviour' are epidemics (a rise and subsequent fall in incidence) and endemicity (the stable maintenance of infection within the human community). The aim of the discipline is to determine the underlying processes and understand the interactions between them, that generate observed patterns (e.g. the rate of spread of infection and the pattern of susceptibility to infection). Epidemiology is a quantitative discipline that draws on statistical techniques for parameter estimation and mathematical methods for delineating the dynamic changes that occur through time, across age classes, or over different spatial locations. The discipline also makes use of modern molecular (e.g. DNA probes and polymerase chain reaction) and immunological (measures of the abundances of antibodies specific to an infectious agent's antigens) techniques for the detection and quantification of current and past infection or disease.

Populations

The definition and description of the host and parasite populations is of obvious importance in epidemiological studies. A population is an

assemblage of organisms of the same species (or genetic type etc.) which occupy a defined point or points in the plane created by the dimensions of space and time. The basic unit of such populations is the individual organism (i.e. parasite or human host). Populations may be divided (= stratified) into a series of categories or classes, the members of which possess a unifying character or characters such as age, sex, or their stage of development. Such subdivisions may be made on spatial or temporal criteria to distinguish a local population from a larger assemblage. The boundaries in space, time, and genetic constitution between different populations are often vague, but it is important to define what constitutes the 'study population' as clearly as possible.

The natural history of infection

Mathematical models are often used to depict the rate of spread or transmission of an infectious agent through a defined human community. For their formulation three broad classes of information are required.

1. The modes and rates of transmission of the agent.
2. The typical course of events within an individual following infection.
3. The demographic and social characteristics of the human community.

The mode of transmission (i.e. direct, indirect, horizontal, vertical, etc.) is of obvious importance (Table 1), but if there is more than one route the relative efficiency of each in determining overall transmission must be understood. When considering microparasitic infections (e.g. viruses, bacteria, and protozoa that multiply directly within the host) it is generally not possible to measure the pathogen abundance within the host (i.e. the burden or intensity of infection). However, following invasion it is important to obtain quantitative information on the typical durations of the latent and infectious periods of the infection and the incubation period of the disease it induces. As depicted in Fig. 1, the latent period is defined as the average period of time from the point of infection to the point when an individual becomes infectious to others, the infectious period denotes the average period over which an infected person is infectious to others, and the incubation period defines the average period from infection to the appearance of symptoms of disease. In practice all these periods are variable between individuals, depending on factors such as the size of the inoculum of the infectious agent that initiates infection, the genetic background of host and parasite, past experience of infections, and the nutritional status of the host. The use of an average is an economy of thought and where knowledge permits models should be based on distributed latent and infectious periods. In some instances the infectious period may be influenced by patient management practices such as the confinement of an infected person once symptoms of infection are diagnosed (e.g. measles and tuberculosis).

There are instances in the case of viral and bacterial infections when a knowledge of pathogen abundance within blood, excretions, secretions, and other tissues or organs of the host can be of importance in determining the infectivity of an infected person to susceptible contacts. A good example is provided by HIV-1. Current evidence suggests that the infectiousness of an infected person varies greatly over the long and variable incubation period of the disease AIDS that the virus induces (Fig. 2). It is believed on the basis of recorded fluctuations in HIV antigenaemia that a short period of high infectiousness occurs shortly after infection, followed by a long period of low to negligible infectiousness (perhaps many years) before infectiousness again increases as the infected patient develops symptoms of AIDS (Anderson and May 1988). In these cases rather complex models are required to mirror the natural history of infection (Anderson 1988).

The human immune response to infection, its ability to confer protection against reinfection, and the duration of this protection have important implications for model construction. For the majority of childhood viral infections the assumption of lifelong immunity following recovery appears to be correct. However, as one moves up a scale of parasite structural (antigenic) complexity from viruses to bacteria to protozoa in general the duration of acquired immunity

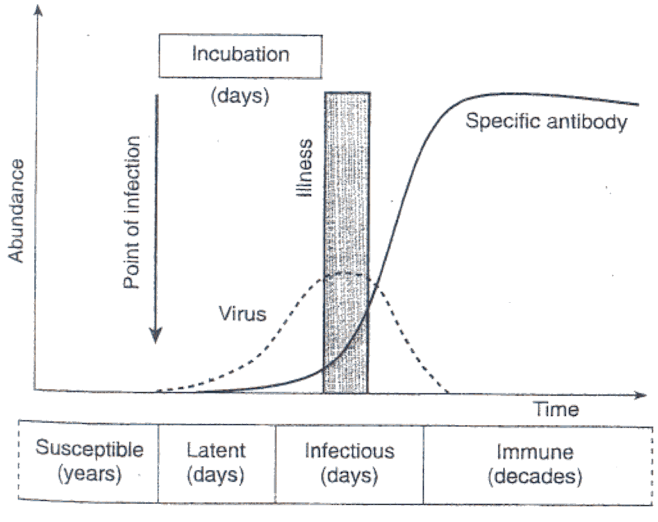


Fig. 1 Schematic representation of the typical time-course of an acute viral or bacterial infection in a host individual and the corresponding progression through infection classes (note the different time durations within each of these classes). (Source: Nokes and Anderson 1988.)

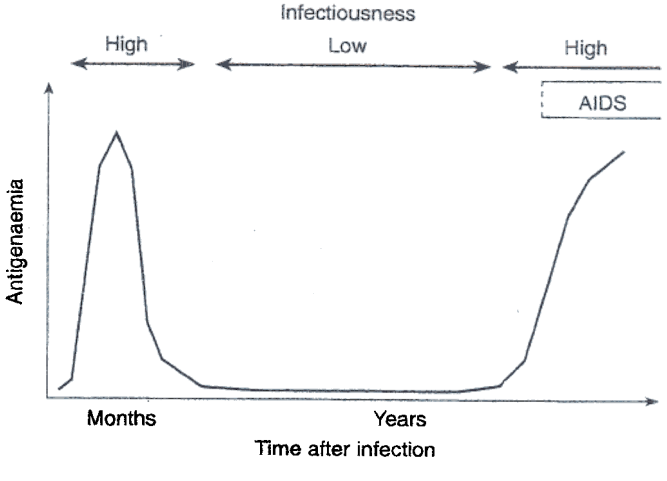


Fig. 2 Possible changes in HIV-1 concentration in the blood of an infected individual (antigenaemia) and in the associated degree of infectiousness during the long incubation period of AIDS

Table 1 Epidemiological classification of infectious diseases of public health importance in developed countries

Mode of transmission	Type of parasite	Examples (diseases or agents)
	Micro[b]	
	Viruses	Rubella, hepatitis B, cytomegalovirus, retroviruses
	Protozoa	*Toxoplasma gondii*
HORIZONTAL		
Direct		
Close contact	Micro	
	Viruses	Measles, mumps, rubella, Epstein–Barr virus, herpes simplex-1, respiratory syncytial virus, influenza-2, varicella, common cold
	Bacteria	Diphtheria, pertussis, bacterial meningitis
	Macro[c]	
	Nematodes	*Enterobius vermicularis* (pinworm)
Environmental	Micro	
	Viruses	Hepatitis A, polio, Coxsackie
	Bacteria	Tetanus, *Shigella*, *Salmonella*, typhoid, cholera, Legionnaires' disease
	Protozoa	*Giardia intestinalis*, amoebiasis
	Macro	
	Nematodes	Pinworm
Sexual	Micro	
	Viruses	Hepatitis B, HIV, herpes simplex-2, cytomegalovirus
	Bacteria	*Neisseria gonorrhoeae*, syphilis
	Protozoa	*Trichomonas vaginalis*
Not direct		
Via other host species (zoonoses)	Micro	
	Virus	Rabies
	Protozoa	*Toxoplasma gondii*
	Macro	
	Nematodes	*Toxocara species*
	Cestodes	*Taenia solium*, *T. saginata*, *Echinococcus granulosus* (hydatid)
Vector-borne[d]	Micro	
	Viruses	Hepatitis B, HIV, Venezuelan equine encephalitis
	Bacteria	*Yersinia* species (plague)
	Protozoa	*Plasmodium* species (malaria)

[a]Inclusive of transplacental and perinatal infection.

[b]Microparasites are those that multiply directly within the host individual, usually resulting in acute infections and subsequent durable immunity to reinfection.

[c]Macroparasites are larger parasites whose reproductive stages pass out of the host. Infection intensity is thus a process of accumulation, and can be measured as worm burden.

[d]Needle transmission is included.

decreases. For certain infections, such as gonorrhoea, acquired immunity is absent while for many protozoan infections it is of short duration (e.g. *Plasmodium* sp.). The inability to develop effective immunity is often related to the genetic diversity of the infectious agent population (antigenic diversity) such that infection with one genetic strain fails to protect against invasions by another (e.g. *Neisseria gonorrhoea*, *Neisseria meningitidis*, and influenza viruses). The question of immunity can be complicated by a degree of cross-immunity (non-specific in character) resulting from infection by dissimilar organisms (e.g. many bacterial infections of the respiratory tract).

Demographic and behavioural characteristics of the human community are usually important in the study of transmission dynamics. For infections that confer lifelong immunity on host recovery the rate of input by births of new susceptibles will influence the overall pattern of infection in a community. Similarly, the rate of transmission of 'close contact' infections (Table 1) will depend upon the degree of mixing between individuals and the density and age distribution of susceptibles and those infected. Heterogeneity in behaviour within a community is of particular importance in the study of sexually transmitted infections since rates of sexual partner change vary greatly between individuals (Johnson *et al.* 1992, 1994). More generally,

heterogeneity in any behaviour, whether sexual or social mixing, must be captured in model formulation.

It will be clear from the preceding comments that much quantitative detail about the natural history of infection must be understood for accurate model formulation. In many instances such detail is not available, but model formulation can greatly facilitate our knowledge of what needs to be understood to define the transmission dynamics of a given infection. With respect to many childhood viral and bacterial infections, such as measles, rubella, mumps, pertussis, and diphtheria, a great deal is understood about the natural history and, hence, much of the work on mathematical models has focused on those infections. Their direct route of transmission, their tendency to induce lifelong immunity, plus, in most cases, the availability of serological or virological techniques to detect past or current infection facilitates the acquisition of quantitative data.

Units of measurement

The unit of measurement employed in epidemiological study depends on the type of infection. The most basic unit is that of the individual parasite. As already discussed, in most cases this unit is not a practicable option for microparasitic organisms due to difficulties in detection and quantification (however, advances in molecular biology and biochemistry are generating new techniques which may be of value in the near future). As such, the most useful unit is that of the infected host which allows the human community to be stratified on the basis of whether individuals are susceptible, infected but not yet infectious (= latent or pre-patent), infectious, and recovered (= immune in the case of many viral infections). Infection may be detected directly (e.g. DNA probes, virus, or bacterial culture) or indirectly by the presence of antibodies specific to pathogen antigens (serological and salivary tests). Seropositivity does not necessarily discriminate between infected and recovered individuals, but for many viral and bacterial infections serological surveys of a population, perhaps stratified by age, sex, and other variables carried out longitudinally (through time via cohort monitoring) or horizontally (across age classes) provide a key measure of transmission and the broad epidemiological characteristics of the infection.

What models describe

At any point in time a population may be classified by the density or number of susceptible, infected, and immune individuals. With the passing of time and concomitantly as individuals age, people may move from one infection class to the next. As such, with the recruitment of new susceptibles by birth and, in some cases, the loss of

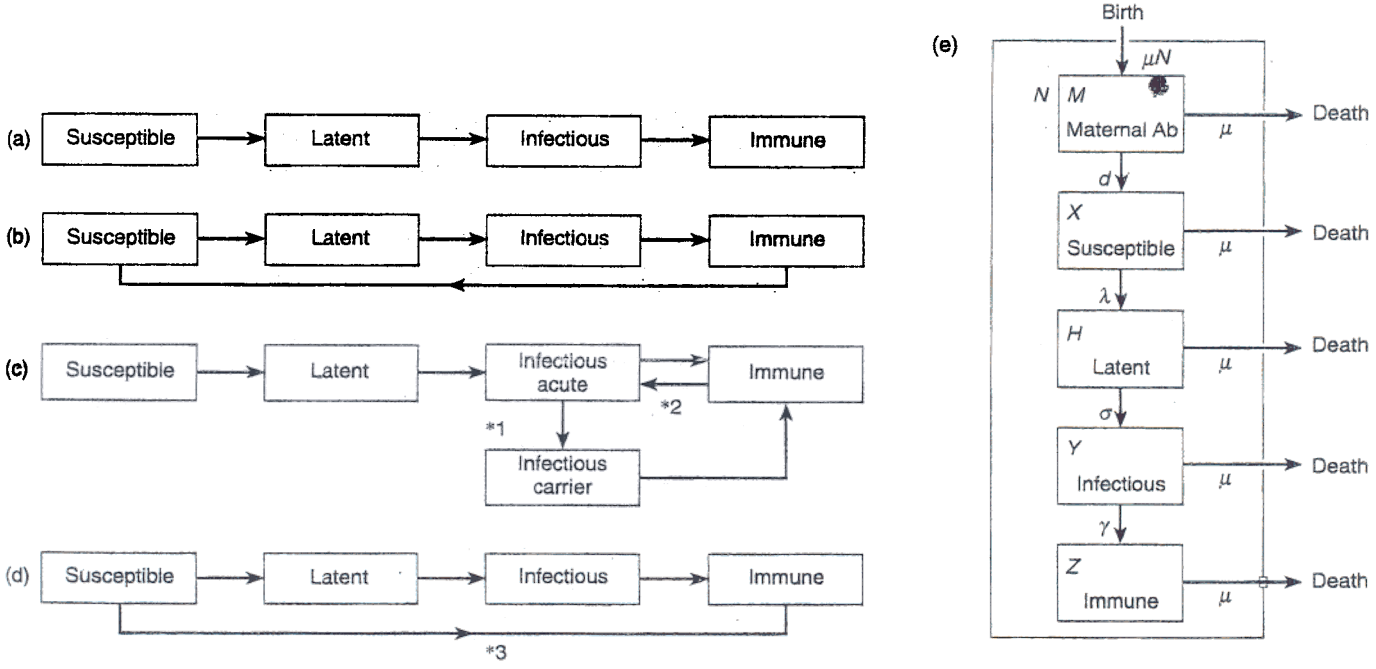


Fig. 3 Flow diagrams used to describe the movement of individuals within populations compartmentalized according to infection status to particular parasitic agents. (a) Simple model for infections inducing lasting immunity (e.g. measles, mumps, rubella, yellow fever, and poliomyelitis) or (b) in which immunity is transient and individuals subsequently return to the susceptible pool (e.g. *Neisseria gonorrhoea*, typhoid, cholera, *Trichomonas vaginalis*). (c) Many infections persist within the host for long periods of time, during which the infected individual may remain infectious (1), as is the case for carriers of hepatitis B virus, gonorrhoea, *Salmonella typhi*, and *Treponema pallidum* (syphilis), chronic tuberculosis patients, or during recrudescence of herpes viruses and malaria. The epidemiological importance of this characteristic is that it enables the perpetuation of such infections in low density communities (see discussion of the mass-action principle in the text). For other infections immunity is defence against disease but not asymptomatic reinfection (2) from which new infectious individuals arise (e.g. *Haemophilis influenzae* and *Neisseria meningitidis*). (d) Vaccination (3) has the effect of transferring individuals directly from the susceptible to the immune class. (e) More detailed description of the transmission dynamics of an acute microparasitic infection which explicitly accounts for births and deaths in the population. All neonates are born possessing maternally derived protective antibody. The net birth rate is assumed to equal the sum of the net death rates for each subpopulation (compartment), that is, births = μN, where $N = M + X + H + Y + Z$ = constant population size. The per capita rates defining movement between infection classes are described in the text.

immunity, the population structure is a dynamic process with individuals flowing from one class to the next. Mathematical models of transmission attempt to capture the dynamic nature of these changes in the form of difference (discrete time steps) or differential (continuous time) equations (Scott and Smith (1994) give a simple introduction). With respect to microparasitic infections where the population is stratified or compartmentalized by infection status, the resulting models are often referred to as compartmental models. The types and numbers of compartments will depend upon the type of infectious agent and the details of its natural or life history. A number of examples are recorded in Fig. 3 in the form of flow diagrams. These diagrams form a useful intermediary step between biological comprehension and mathematical formulation.

Population rates of flow

Following the introduction of an infection into a stable population the number or density of individuals within the various infection compartments will depend on the rates of flow between compartments such as infection and recovery rates. The size of a population in a specific compartment will depend on the magnitude of those rates that determine the entry and duration of stay. In general, the shorter the duration of stay (the higher the rate of leaving) in a particular compartment the smaller the size of the population in that category (the inverse relationship between 'standing crop' and 'rate of turnover'). If the infection attains a stable endemic equilibrium in the human community, the net input into each compartment will exactly balance the net output. The relative numbers in each compartment will be directly related to the duration of stay. Thus, for example, in the case of endemic measles in a developed country where immunity is lifelong (many decades), individuals remain in the susceptible class for an average of 4 to 5 years and in the latent and infectious classes for a few days (say 7 days on average in each). As such, most people are in the immune class, followed by the susceptible class, and few individuals at any point in time are in the latent and infectious classes. Figure 4 provides a diagrammatic representation of this point.

A formal demonstration of the influence of rates of flow (or durations of stay) on the proportion of susceptibles, those infected, and immunes in a population is made possible by the translation of the flow diagram of movement between compartments (Fig. 3) into a set of coupled differential equations. Typically, these describe the rates of change with respect to time (or age or both) of the densities of infants with maternally derived immunity (due to maternal antibodies), susceptibles, infecteds not yet infectious, infectious individuals, and immunes, denoted respectively by $M(t)$, $X(t)$, $H(t)$, $Y(t)$, and $Z(t)$ at time t (Fig. 3(e)). In writing down these equations we need to define the rates of flow between compartments by a series of symbols. For example, in common notation δ (delta) defines the loss of maternally derived immunity, i.e. the average per person rate of loss of passive protection. The absolute rate of loss from or movement out of class M (Fig. 3) requires that the per capita rate (i.e. person/unit of time) be multiplied by the size of the M subpopulation, that is δM (which has units of persons/unit of time). If δ is the per capita rate of movement out of class M then the average duration of maternally derived immunity is $1/\delta$. These principles apply to the other rate terms shown in Fig. 3(e). Hence, using conventional symbols, β (beta) is the transmission coefficient that defines the probability of contact and infection transfer between a susceptible and infectious person, σ

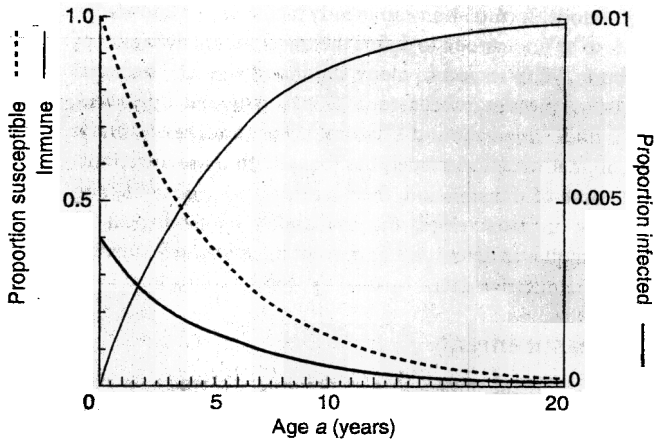


Fig. 4 The proportions of a population who are in the susceptible, infected (either latent or infectious), and immune classes for a typical childhood viral infection. In this example, which is based on measles, the force of infection, $\lambda = 0.2$ per year (corresponding to an average age at infection of 5 years) and the rate of movement from the latent class, σ, and recovery from infectiousness, γ, is 52 per year (corresponding to an average duration of stay in each of these infected classes of 1 week). Note that the proportion of the population in the infected classes is always much less than that in the susceptible or the immune classes (Fig. 1).

(sigma) defines the per capita rate of leaving the latent class (average latent period $1/\sigma$), γ (gamma) the per capita rate of leaving the infectious class (average infectious period $1/\gamma$), and μ (mu) the natural per capita mortality rate ($1/\mu$ is average life expectancy). For developed countries it is commonly assumed that population size is approximately constant such that net births exactly balance net deaths. Therefore the net death rate, μN (where N is the total population size $M + X + H + Y + Z$) is equated by births (hence the term mN for births in Fig. 3(e)). Additionally, it is assumed that infection does not induce an extra case mortality rate over and above natural mortality. With this notation we can define the equations for M, X, H, Y, and Z as

$$dM/dt = \mu N - (\delta + \mu)M \quad (1)$$

$$dX/dt = \delta M - (\beta Y + \mu)X \quad (2)$$

$$dH/dt\ \beta XY - (\sigma + \mu)H \quad (3)$$

$$dY/dt = \sigma H - (\gamma + \mu)Y \quad (4)$$

$$dZ/dt = \gamma Y - \mu Z. \quad (5)$$

In these equations, which constitute a simple model of infection transmission, $d\square/dt$ simply refers to the rate of change of the number or density of individuals in a class (M, X, H, Y, or Z) with respect to (over) time. The right-hand side of each of these equations then expresses precisely what the rate of change is. For a simple introduction to the definition, manipulation, and interpretation of such differential equations in the epidemiological context, the reader is referred to Scott and Smith (1994). The major assumptions incorporated in the above equations are that the net rate of infection βXY is proportional to the density of susceptibles multiplied by the density of infectious individuals, that individuals leave each compartment at a constant per capita rate (other than for the susceptible class, X, (see below)) because it is assumed that the per person rates δ, μ, σ, and γ, do not change over time, and that net births exactly balance net deaths (reasonably accurate for developed countries). To explore what

these assumptions imply in terms of the dynamics of transmission and the numbers or proportions of individuals in each class we need to solve these equations either analytically to obtain explicitly expressions for $M(t)$, $X(t)$, $H(t)$, $Y(t)$, and $Z(t)$ in terms of the rate parameters and the variable time (t) or numerically to generate projections of changes in the numbers in each compartment through time (Scott and Smith 1994). In the case of simple models we can often obtain exact analytical solutions as is the case for the equation for $M(t)$ in the model defined by eqns (1) to (5). The solution gives us the number of infants with maternally derived protection at time t, $M(t)$:

$$M(t) = \{(\mu N)/(\mu + \delta)[1 - e^{-(\mu + \delta)t}]\} + M(0)e^{-(\mu - \delta)t} \quad (6)$$

where $M(0)$ is the number protected at time $t = 0$.

More generally, the complexity of the life histories of many infections makes analytical solution difficult or impossible and numerical methods are required. Modern computers make light work of very complex systems of equations describing disease transmission and many software packages are available for the solution of sets of differential equations and now for model making. Nevertheless, in these cases some general analytical insights can be obtained by examining the equilibrium properties of the model which is done by setting the time derivatives (i.e. the d/dt) equal to zero, that is such that there are assumed to be no further changes in the number of individuals within each infection class because the flows into and out of any one category are equal. These equations can then be solved to determine the numbers at equilibrium (i.e. at stable endemicity) in each class (referred to as M^*, X^*, H^*, Y^*, and Z^*). For example, in the simple model of eqns (1) to (5) by simple algebraic manipulation we obtain

$$M^* = \mu N/(\delta + \mu) \quad (7)$$

$$X^* = (\sigma + \mu)(\gamma + \mu)/\beta\sigma \quad (8)$$

$$H^* = (\gamma + \mu)Y^*/\sigma \quad (9)$$

$$Y^* = (\delta M^* - \mu X^*)/\beta X^* \quad (10)$$

$$Z^* = \gamma Y^*/\mu \quad (11)$$

where N is the constant representing the total population size. These equilibrium solutions illustrate how the various rate parameters that determine flow between compartments influence the numbers of individuals in each compartment when the infection is at an endemic steady state. For example, based upon the assumptions in our model for an acute childhood infection (eqns (1)–(5)), we can suggest that at endemic equilibrium the number or density of individuals in the maternal antibody class, M^*, is directly dependent upon the net rate of births (where births equal deaths, μN) and inversely related to the rate of loss from the class ($\delta + \mu$) (where $1/(\delta + \mu)$ is the average duration in the maternal antibody protected class).

Parameter estimation

The preceding section provided a clear illustration of the numerous parameters that are necessary to define even the simplest model of direct transmission within a human community. To make the best use of a model it is desirable to have available estimates for each of the parameters for a given infection. Some, such as the demographic rates of birth and mortality and total population size, can be easily obtained via national census databases (usually finely stratified by age and sex in developed countries). Others, such as the average latent and infectious periods, must be determined either by clinical studies of the course of infection in individual patients (e.g. measures of change in viral abundance during the course of infection) or by detailed household studies of case-to-case transmission. Statistical methodology plays an important role in this instance since, as noted earlier, latent, infectious, and incubation periods are rarely constant from one individual to the next. Statistical estimation procedures have been developed to help derive summary statistics of these distributions (e.g. means and variances) (Bailey 1975).

Invariably, the most difficult parameter to estimate is the transmission coefficient β (see eqn (2)), which is a measure of the rate of contact between members of a population plus the likelihood of infection resulting from contact. In some cases, such as certain sexually transmitted infections (e.g. gonorrhoea), direct estimates can be obtained via contact tracing methods (Hethcote and Yorke 1984). More commonly, indirect methods must be employed, often themselves based on model formulation and analysis. A simple example employs the model defined in the previous section by eqns (1) to (5). We can define the component βY of the transmission term as the per capita rate at which susceptibles (X) acquire infection. This rate is commonly referred to as the 'force of infection' and denoted by the symbol λ (lambda). Analysis of the model reveals that the average age at which an individual typically acquires infection, A, is approximately related to the force of infection by the expression

$$A \simeq 1/\lambda. \quad (12)$$

Hence if we can estimate A from an age-stratified serological profile or from age-defined case-notification records (Box 1) we can, via eqn (12), estimate λ. More generally, this rate often varies with age and more complex methods of estimation must be employed given good age-stratified serological data (see Grenfell and Anderson 1985). Put in simple terms, if the proportion susceptible at age $a + 1$ is $x(a + 1)$ then the force of infection over the age interval $a \rightarrow a + 1$ (defined per unit of age) is simply

$$\lambda = -\ln[x(a + 1)/x(a)]. \quad (13)$$

With serological data finely stratified by age, under the assumption that the infection confers lifelong immunity upon recovery, eqn (13) can be used to estimate how λ changes with age in a given community. For most childhood viral and bacterial infections λ is a function of age, changing from low values in infant classes to high in child to young teenage classes back to low in adult age classes (Fig. 5). This is thought to reflect patterns of intimate contact via attendance at school and play activities.

Further complications may arise if rates of contact or transmission vary through time, perhaps due to seasonal factors such as the aggregation and dispersal of children at term and school holiday periods (Yorke *et al.* 1979; Anderson 1982; Bolker and Grenfell 1993). The problems of parameter estimation are considered in more detail in a later section.

Concepts in quantitative epidemiology

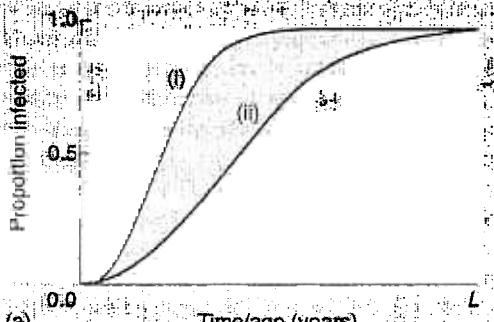
The incidence of infection and disease

Transmission by direct contact and the law of mass action

When close contact between infectious and susceptible individuals is necessary for transmission, the number of new cases in a population which arise in a unit of time (i.e. incidence of infection) is often

Box 1 Surveillance profiles

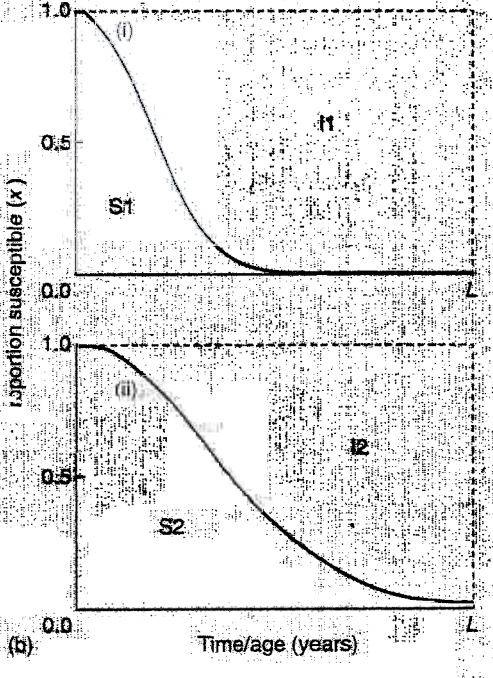
Two infections (i) and (ii) are at endemic equilibrium (i.e. roughly constant incidence in time) in a stationary host population (i.e. births are equal to deaths). The changes, with time or increasing age, in the proportion of the population that has experienced each infection may be estimated from longitudinal cohort or horizontal cross-sectional surveys (serological or case notifications) of individuals from birth to life expectancy L.



The steeper profile of (i) compared with (ii) is an indication that the basic reproductive potential R_0 of infection (i) exceeds that for infection (ii) such that

$R_0(i) > R_0(ii)$.

Assume that each infection induces lifelong immunity; then, from the above profiles, changes with age/time in the proportion x susceptible to each infection are as shown in figure (b).



The (equilibrium) proportion of the total population susceptible to infection (i) is

$x^*(i)$ = area S1/(area S1 area I1)

and for infection (ii) it is

$x^*(ii)$ area S2/(area S2 area I2).

Note that the equilibrium proportion susceptible to infection (i) (with the higher reproductive rate) is smaller than that for infection (ii) (with a lower rate of reproduction), i.e.

$x^*(i) < x^*(ii)$.

The relationship between these two epidemiological parameters may be usefully expressed as

$R_0(i) = 1/x^*(i)$

and

$R_0(ii) = 1/x^*(ii)$.

Summing the proportion susceptible, $x(a)$, in the above graphs for each age class from age 0 years (time 0) to L years, we can determine the average age at infection, A:

$x(a) = x(0) + x(1) + x(2) + \ldots + x(L) = A = S$

from which it can be seen that

$A(i) < A(ii)$.

Note also that

$R_0(i) = L/A(i) = L/S1$

and

$R_0(ii) = L/A(ii) = L/S2$.

(See also eqns (12) and (13), and Anderson and May (1983), for estimation of the force of infection from surveillance profiles.)

Summary examples

Assume infection (i) is measles and infection (ii) is rubella in the United Kingdom with average lifespan L of 75 years. If S1 = 5 and S2 = 10, then $A(i) = 5$ years and $A(ii) = 10$ years, and

$x^*(i) = 5/75 = 0.066'$ $x^*(ii) = 10/75 = 0.133'$.

Therefore

$R_0(i) = 1/0.066' = 15$

or

$R_0(i) = 75/5 = 15$

and

$R_0(ii) = 1/0.133' = 7.5$

or

$R_0(ii) = 75/10 = 7.5$.

The implications of this difference in the basic reproductive rate of infection to the proportion of the population that must be vaccinated in order to eradicate each infection can be seen in Fig. 1.1.

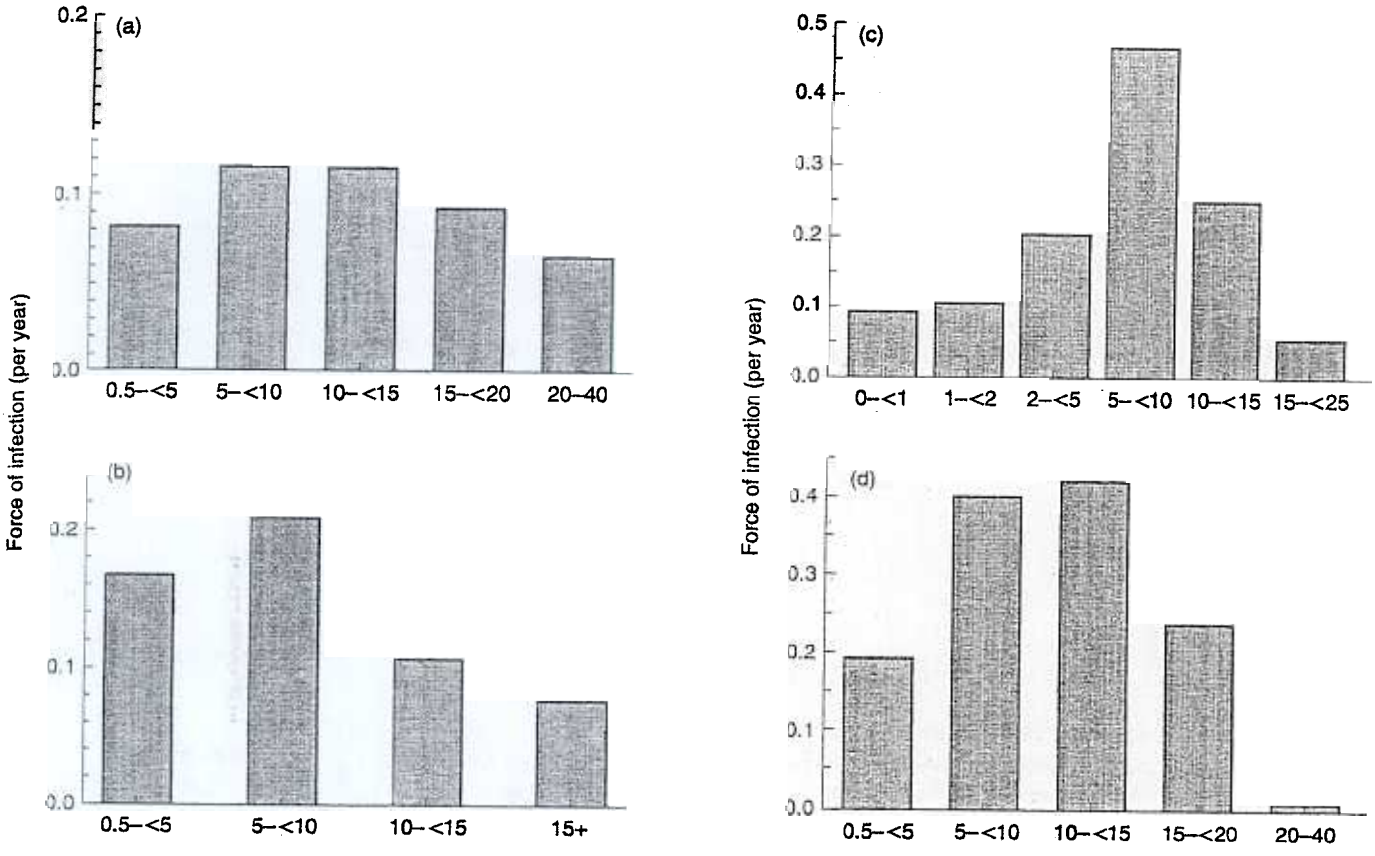


Fig. 5 Examples of the age-dependent nature of the per susceptible rate of transmission for common childhood viral and bacterial infections. Graphs (a) and (b) derive from horizontal cross-sectional serological surveys in the United Kingdom, of rubella (Nokes *et al.* 1986) and mumps (Anderson *et al.* 1987) respectively. Graphs (c) and (d) provide estimates based on case-notification data for England and Wales for whooping cough (Anderson and May 1985*b*) and measles (Grenfell and Anderson 1985) respectively.

assumed to be approximately given by the density (or number) of susceptibles, X, multiplied by the density (or number) of infectious persons, Y, multiplied by the probability of an effective (infectious) contact between an infectious person and a susceptible, β (i.e. βXY). This relationship is commonly referred to as the 'law of mass action' by analogy with particles colliding within an ideal gas system (Box 2). The basic assumption implicit in this concept is that the population mixes in a random manner (often referred to as homogeneous mixing). The term βXY which describes net transmission is the major non-linear expression in most compartmental models of directly transmitted viral and bacterial infections. It is, of course, a crude approximation of what actually occurs in human communities and more realistic refinements of this assumption are discussed in later sections. However, it provides a convenient point of departure for model construction and analysis.

The transmission coefficient β

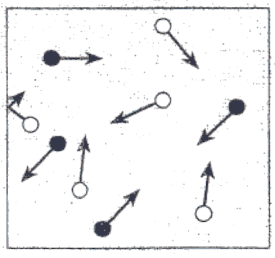
The probability of transmission, β, is made up of two components, namely the rate at which contacts occur between susceptible and infectious persons and the likelihood that transmission will result from a contact. Consequently β is dependent on sociological and behavioural factors within the host population (i.e. rate of mixing) and the biological properties that determine the infectiousness of an infected person and the susceptibility of an uninfected individual. These biological properties involve factors such as the virulence of the infectious agent and the genetic background plus the nutritional status of the human host.

Incidence estimates

The incidence of infection, I, can be measured by direct observation of new cases, such as notifications of measles or pertussis. Unfortunately, however, measures of incidence tell us nothing about the respective densities of susceptibles or infectious people, nor the magnitude of the transmission coefficient β. It is common practice in epidemiology for I to be expressed as the number of cases per unit of population (usually 100 000 people in a defined class, such as age or sex) over a defined period of time such as 1 year (e.g. 5/100 000 per annum). Such measures are often referred to as attack rates (**AR**). However, they are a rather poor measure of the intensity of transmission within a population since they take no account of the proportion of the community (or age or sex class) that is susceptible to infection (Box 3). A better measure of the rate at which susceptibles acquire infection is provided by a parameter termed 'the force of infection' commonly denoted by the symbol λ. It simply defines the probability that a susceptible individual will acquire infection over a short period of time (i.e. a per susceptible (= per capita) rate of infection) and, in the terminology of the mass-action principle, is defined as $\lambda = \beta Y$. Here β might be thought of as the force of infection for one infectious person

Box 2 The law of mass action and the incidence of infection

Imagine susceptible and infectious individuals behaving as ideal gas particles within a closed system with no immigration or emigration and occupying a defined space, where X is the number of particles of one gas (i.e. susceptibles), Y is the number of particles of a second gas (i.e. infectious people), and β is the collision coefficient for the formation of molecules of a new gas from one molecule each of the original gases (i.e. new cases of infection) (figure (a)).

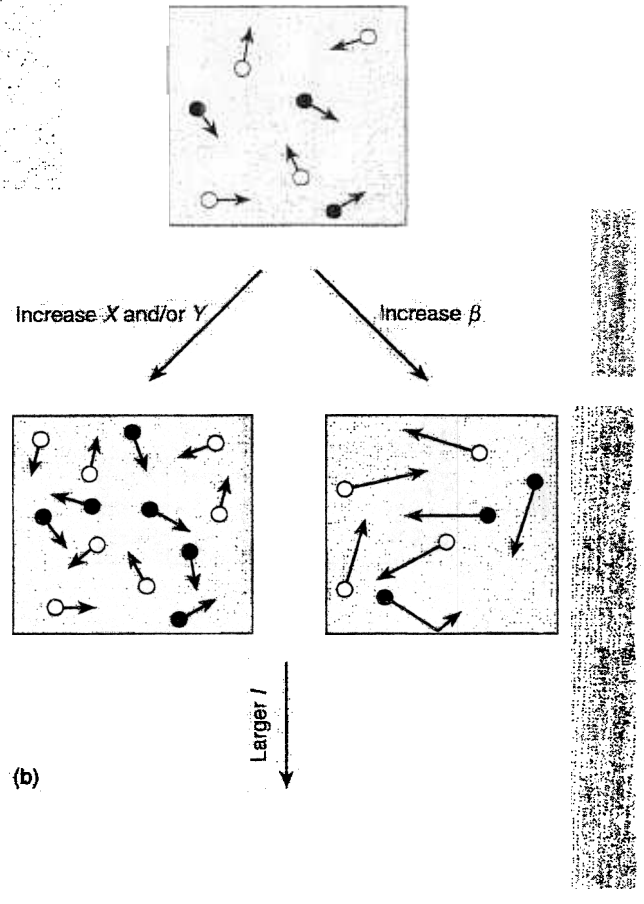


(a)

Gas particles (individuals) are mixing in a homogeneous manner such that collisions (contacts) occur at random. The law of mass action states that the net rate of production of new molecules (i.e. cases), I, is simply

$$I = \beta XY.$$

The coefficient β is a measure of (i) the rate at which collisions (contacts) occur and (ii) the probability that the repellent forces of the gas particles can be overcome to produce new molecules, or, in the case of infection, the likelihood that a contact between a susceptible and an infectious person results in the transmission of infection. Under these assumptions, the incidence of infection will be increased by larger numbers (or densities) of infectious and susceptible persons and/or high probabilities (β) of transmission (figure (b)).



in a community. Estimates of this rate λ can be derived from age-stratified serological profiles or case notifications (Anderson and May 1983, 1991; Grenfell and Anderson 1985) (Fig. 5).

Validity of the mass-action principle

Despite the simplicity of the notion of homogeneous mixing implicit in the mass-action principle of transmission, the predictions of simple compartmental models based on this assumption often mirror observed epidemiological patterns surprisingly well (Anderson and May 1982). In part, this is a consequence of increased travel, movement, and mixing within many societies in developed countries. Measles epidemics, for example, are often synchronous in England and Wales, with a clear distinction in all parts of these regions between years of high incidence and years of low incidence (Fig. 6). However, the less able an infection is to spread through a particular population (lower R_0, see below) then the more important are slight deviations from homogeneous mixing, resulting in a lower degree of synchrony of epidemics in a country (Fig. 7). The assumption is most appropriate for infections which are spread by close contact between individuals such as respiratory infections transmitted by contaminated droplets and nasopharyngeal secretions. In such cases, the survival of the infectious agent in the external environment is of very short duration (i.e. minutes). As such, there is no significant reservoir of infectious stages to maintain transmission in the absence of infectious persons.

Many kinds of heterogeneities can invalidate the mass-action principle and much attention in recent years has been devoted to their inclusion in compartmental models. The major sources are heterogeneities arising from age-related factors, that determine contact and mixing patterns (i.e. 'who mixes with whom') and spatial factors such as differences in population densities in urban and rural areas of a country (Anderson and May 1984, 1991; May and Anderson 1984). Such sources of heterogeneity are very important in the design of control policies based, for example, on mass vaccination and models have been developed to assess their impact.

Heterogeneity in behaviour is of particular importance in the study of sexually transmitted infections such as gonorrhoea and HIV. One of the major determinants of the rate of spread of such infections is the distribution of the rate of sexual-partner change within a defined community (Fig. 8). These distributions are typically highly heterogeneous in character (i.e. the variance in the rate of partner change is much greater in value than the mean rate of partner change) where most people have few different sexual partners in a lifetime (or over a defined period of time) and a few have very many. The activities of individuals in the 'tail' of the distribution (the highly sexually active)

Box 3 Interpreting attack rates

Care should be exercised when interpreting attack rates in the absence of information on the proportion of individuals within the population who are immune as a result of previous infection (assuming we are considering an infection such as measles that induces lasting immunity on recovery). A simple illustrative example is given below based on case notification for measles.

Age (years)	Attack rate per head of population in that age class	Percentage immune in the age class	Modified attack rate based on infection per head of the susceptible population
2	180/100 000	10	
10	20/100 000	90	

At a first glance at column 2, the attack rate suggests that infants aged 2 years have a much greater chance of acquiring infection than children aged 10 years. However, if we adjust the denominator of the attack rate from per head of population in that age class to per head of susceptible population in the age class, we see from the fourth column that the rate of infection is identical in both age classes.

are clearly important for the persistence and spread of infection since those with many partners are both more likely to acquire infection and more likely to transmit it to others.

Simple theory based on compartmental models of the transmission of infections such as gonorrhoea and HIV assumes that the net rate at which infection is spread in, for example, a male homosexual community is determined by the proportion of infectious persons

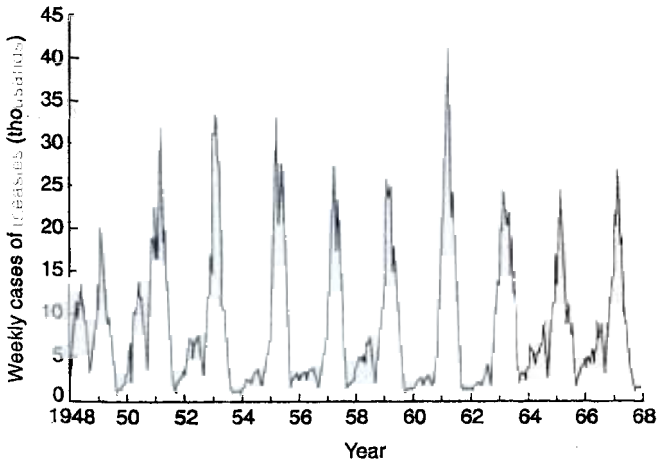


Fig. 6 The number of cases of measles reported each week in England and Wales between 1948 and 1968. (Source: Office of Population Census and Surveys, London.)

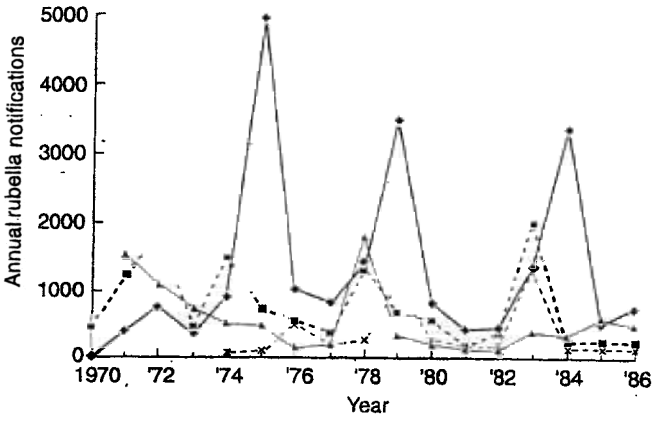


Fig. 7 Annual rubella case notifications reported by four city health authorities in England: Leeds (♦), Bristol (■), Manchester (▲), and Newcastle (×). The dominant interepidemic period is roughly 4 to 5 years with peak incidence often slightly out of phase between cities (compare with Fig. 6). (Source: Communicable Disease Surveillance Centre, London.)

(Y/N where N is the total size of the sexually active population) multiplied by the density of susceptibles (X) multiplied by a transmission coefficient β. This coefficient is defined as the probability that a sexual contact (per partner) results in transmission, B, multiplied by the effective rate of sexual-partner change, c (which determines contacts). If the population mixed homogeneously, this effective rate would simply be the mean rate of sexual-partner change, m. When great heterogeneity in rates of partner change is present within a population the effective rate must be defined in terms of this variability as well as the mean rate of activity. If we assume that the population is divided into classes with different rates of partner change and that partners are chosen (from any class) in proportion to their representation in the population multiplied by the rate of partner change in each group (an assumption of 'proportional mixing' (Anderson *et al.* 1986; May and Anderson 1987; Garnett *et al.* 1992; Gupta and Anderson 1992)), then the effective rate of partner change, c, is given by

$$c = m + (s^2/m) \tag{14}$$

where m is the mean rate of partner change and s^2 is the variance in the rate. The importance of variability in contact is clear from this simple equation. For example, suppose the mean rate per year is unity but the variance is five times greater. If we assumed that homogeneous mixing occurred our estimate of the effective rate would be 1, but if we take account of heterogeneity the effective rate is six times as large. The influence of the small proportion of highly sexually active individuals on the overall transmission rate is very significant.

Transmission thresholds and the basic reproductive rate of infection

The basic reproductive rate of infection R_0

A key measure of the transmissibility of an infectious agent is provided by a parameter termed the basic reproductive rate (or, also in the literature, basic reproduction number or ratio) and denoted by the symbol R_0. It measures the average number of secondary cases of infection generated by one primary case in a susceptible population.

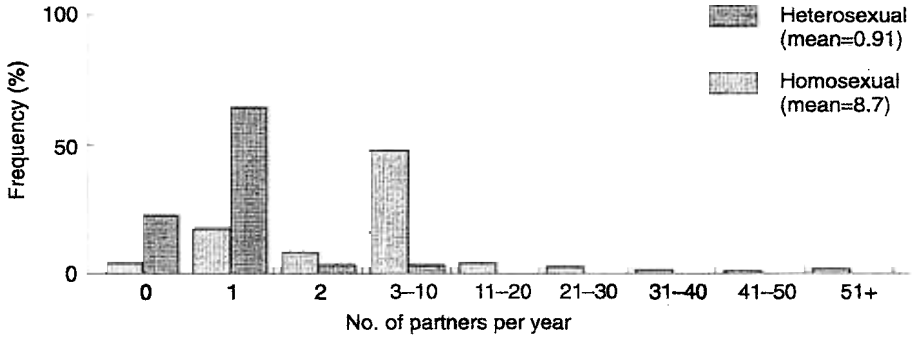


Fig. 8 Variation in the numbers of different sexual partners per year revealed from surveys of the male homosexual and the heterosexual communities in the United Kingdom, 1986 (Anderson 1988). The skewed distribution observed in each instance (an indication that although the majority of individuals have few partners, a few have very many), and the mean rate of sexual-partner change (indicated), are both of significance to the perpetuation and rate of spread of sexually transmitted diseases in the community.

Its value is defined by the number of susceptibles present with which the primary case can come into contact (X) multiplied by the length of time the primary case is infectious to others, D, multiplied by the transmission coefficient, β (rate of effective mixing):

$$R_0 = \beta XD. \tag{15}$$

Note that R_0 is a dimensionless quantity (i.e. the units of measurement cancel out) that defines the potential to produce secondary cases (in a totally susceptible population) per generation time (i.e. the average duration of the infection).

The basic reproductive rate is of major epidemiological significance since the condition $R_0 = 1$ defines a transmission threshold below which the generation of secondary cases is insufficient to maintain the infection within the human community. For values above unity the infection will trigger an epidemic and, with a continual input of susceptibles, will result in endemic persistence. A further quantity of interest is the effective reproductive rate R which defines the generation of secondary cases in a population which contains susceptibles and immunes (as opposed to just susceptible individuals). If the prevalence or incidence of infection is stable through time, the effective reproductive rate R must equal unity in value a situation in which each primary case gives rise, on average, to a single secondary infectious individual.

Factors that influence R_0

The simple expression $R_0 = \beta XD$ (appropriate for directly transmitted infections under the mass-action assumption) provides a framework for assessing how different epidemiological factors influence transmission success. Clearly, high transmission coefficients, long periods of infectiousness, and high densities of susceptibles enhance the generation of secondary cases. Note that its value depends not only on the properties that define the course of infection in an individual (i.e. the duration of infectiousness, D), but also on attributes of the host population such as the density of susceptibles, X, and the component of β that determines the rate of contact or mixing. A good example of the influence of population level characteristics is provided by the rate of transmission of the measles virus in urban centres in developed and developing countries. The more rapid rise in the proportion of children who have experienced infection, with age, in developing countries by comparison with developed regions is in part a consequence of higher population densities and poorer living conditions (McLean and Anderson 1988*a*).

Principles of control

The threshold condition for persistence of an infection, defined by $R_0 = 1$, captures the essence of the problem of control. To eradicate an infection we must reduce the value of the basic reproductive rate below unity. Similarly, to reduce incidence the value of R_0 must be reduced below the level that pertains prior to the introduction of control measures. Reductions can be achieved by reducing the infectious period D by, for example, the isolation of infectious persons (perhaps recognized by clinical symptoms of disease), reducing the number or density of susceptibles, usually by immunization, and by altering the social and behavioural factors that determine transmission such as improving living conditions to reduce overcrowding (in the case of sexually transmitted infections, education can serve to reduce rates of sexual-partner change or promote the use of condoms to lower the probability of transmission).

The threshold density of susceptibles

It is clear from the definition of R_0 given above that to maintain the value of the basic reproductive rate above unity the density of susceptibles in the population must exceed a critical value. More precisely, this critical level X_T is (for the mass-action assumption) obtained by setting $R_0 = 1$ in eqn (15) and rearranging:

$$X_T = 1/\beta D. \tag{16}$$

The aim of mass vaccination, aside from protecting the individual, is to lower the density of susceptible people in the population. If eradication is the aim of control then the density of susceptibles must be reduced to less than X_T in value.

Critical community size

The magnitude of R_0 and, concomitantly, the size of the threshold density of susceptibles determines whether or not an epidemic of an infection will occur when introduced into a given community. In practice, however, for infections that induce lasting immunity in those who recover, the long-term endemic persistence of infection will depend on the renewal of the supply of susceptibles by new births or, to a lesser extent, by immigration. As such, the net birth rate in a

community, which is itself dependent on the total population size (or density), will influence the likelihood of persistence. There is, therefore, a critical community size for the endemic persistence of a given infection. In certain island communities, immigration of susceptibles and infectious individuals may also play a role in the long-term persistence of a given infection (Black 1966; Anderson and May 1986, 1991). These factors are of growing significance as rates of population movement increase as a result of, for example, improved air transport services. Table 2 provides an example of the relationship between community size and the likelihood of the endemic persistence of the measles virus.

The concepts of a threshold density of susceptibles and a critical community size are most relevant for directly transmitted viral and bacterial infections that induce lasting (= lifelong) immunity. The production of long-lived infective stages or the use of vectors (such as mosquitoes) lessen the importance of the human population density for the persistence of an infection. In the case of sexually transmitted infections, simple models suggest that there is no critical density of susceptibles for persistence since the magnitude of R_0 can be approximately given by

$$R_0 = BcD \tag{17}$$

where c is the effective rate of sexual-partner change, D is the average duration of infectiousness, and B is the transmission probability per partner contact (Anderson *et al.* 1986). This is simple to arrive at theoretically. If, as stated earlier, the incidence of cases of a sexually transmitted disease is defined as

$$I = BcXY/N \tag{18}$$

then following the introduction of a single infectious person ($Y = 1$), infectious over a period D, into a totally susceptible population ($N = X$), the number of secondary cases will be represented by eqn (17).

The dependence upon the number of susceptibles is lost. Biologically this is more difficult to grasp, but it does seem reasonable that the rate of sexual-partner change should be more important to the potential for spread of a sexually transmitted disease than the number of susceptibles in the population.

Table 2 Island community size and endemic persistence of measles

	Population size (units of 100 000)	Percentage of months in which no cases were reported
Hawai	5.50	0
Fiji	3.46	36
Iceland	1.60	39
Samoa	1.18	72
Solomon	1.10	68
Fr. Polynesia	0.75	92
New Caledonia	0.68	68
Guam	0.63	20
Tonga	0.57	88
New Hebrides	0.52	70
Gilbert and Ellice	0.40	85
Greenland	0.28	76
Bermuda	0.41	49
Faroe	0.34	68
Cook	0.16	94
Niue	0.05	95
Nauru	0.03	95
St Helena	0.05	96
Falkland	0.02	100

Source: Anderson (1982*b*).

Regulation of infection within human communities

The regulation (i.e. modulation or control) of the incidence or prevalence of a particular infection within a human community is largely determined by the level of herd immunity (i.e. the proportion of the population immune to infection) and the net rate of input of new susceptible individuals. A simple example serves to illustrate this point. Consider a closed population with no inflow or outflow of susceptible, infected, or immune individuals. If the densities of susceptibles, infecteds, and immunes at time t are defined by $X(t)$, $Y(t)$, and $Z(t)$ respectively, then under the mass-action assumption of transmission the rates of change in the densities with respect to time can be captured by three coupled differential equations:

$$dX/dt = -\beta XY \tag{19}$$

$$dY/dt = \beta XY - \gamma Y \tag{20}$$

$$dZ/dt = \gamma Y. \tag{21}$$

It is assumed here that there is no latent period of infection (individuals are infectious once infected), that the average duration of infectiousness is given by $D = 1/\gamma$ where γ is the rate of recovery from infection, that immunity is lifelong, and that no losses occur due to mortality. If we start with a totally susceptible population and introduce a few infecteds, the occurrence of an epidemic will depend upon the magnitude of the basic reproductive rate R_0 ($R_0 = \beta XD$) and, concomitantly, whether or not the density of susceptibles exceeds the critical threshold value X_T ($X_T = 1/\beta D$) (Fig. 9). Assuming that R_0 is greater than unity then an epidemic will occur, but as time progresses the density of susceptibles will decline ($X \rightarrow Y \rightarrow Z$) until the effective reproductive rate R is less than unity (i.e. susceptible numbers fall below the threshold X_T) and the infection dies out.

For the persistence of the infection one of two things must happen. Firstly, suppose susceptibles are continually introduced into the population at a net rate bN where b is the per capita birth rate and that natural mortalities occur in each class at a per capita rate μ. For simplicity we further assume that net births exactly balance net deaths ($bN = \mu N$) to maintain the total population at a constant size. With these assumptions and provided that $R_0 \geq 1$, we find that the infection persists in the population (Fig. 10(a)) with an endemic equilibrium density of susceptibles again equal to X_T and equilibrium densities of infecteds, Y and immunes, Z, given by

$$Y^* = [\mu/(\mu + \gamma)](N - X_T) \tag{22}$$

$$Z^* = (\gamma/\mu)Y^*. \tag{23}$$

Secondly, suppose that there are no new births and no mortality but that immunity is of short duration such that individuals leave the immune class Z to regain the susceptible class X at a per capita rate α

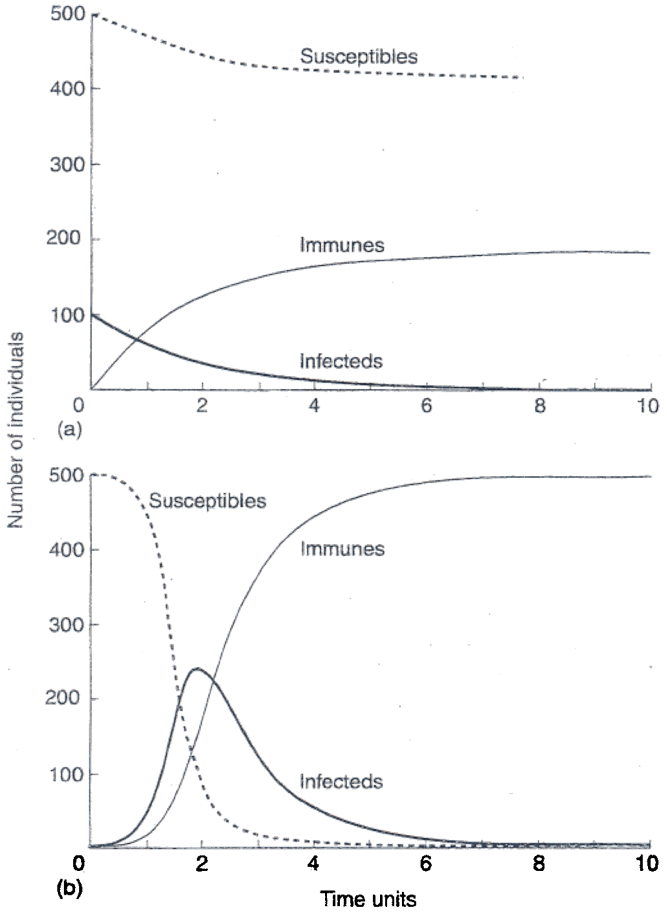


Fig. 9 Conditions for an epidemic. (a) Host density (susceptibles) below the threshold level (at time 0, $X = 500$, $Y = 100$, $Z = 0$, $\beta = 0.0001$, $\gamma = 1$, $R_0 = 0.005$, XT 10000). (b) Host density above the threshold level (at time 0, $X = 500$, $Y = 1$, $Z = 0$, $\beta = 0.01$, $\gamma = 1$, $R_0 = 5$, $XT = 100$).

(alpha) where $1/\alpha$ is the average duration of immunity. We again find that the infection can persist (Fig. 10(b)) (provided that $R_0 \geq 1$) with equilibrium densities of infecteds and immunes of

$$Y^* = [\alpha/(\alpha + \gamma)](N - X_T) \quad (24)$$

$$Z^* = (\gamma/\alpha)Y^*. \quad (25)$$

Note that the faster the loss of immunity (α large) the higher the equilibrium density of infecteds and the lower Z^*.

These two examples show how the net input of susceptibles and the degree of herd immunity (as controlled by the duration of immunity to reinfection following recovery) influence the likelihood that an infection will persist endemically after the initial epidemic has swept through a susceptible population following the introduction of an infection. In these simple models of the transmission of direct-contact infections, the density of infecteds tends to exhibit oscillatory behaviour after the introduction of infection due to the rise and fall in the density of susceptibles taking the effective reproductive rate above and below unity in value. These oscillations are seen to damp down, settling to the equilibrium values given analytically (e.g. eqns (22)–(25)). This propensity to exhibit oscillatory behaviour is more apparent if the infection is of short duration such that infection prevalence is sensitive to the availability of susceptibles and induces long-lasting immunity since it takes some time, under these circumstances, for new births or loss of immunity to replenish the supply of susceptibles such that R is again above unity in value. Maintenance of these oscillations over the longer period would require a force to be applied periodically—in reality this might derive from seasonal changes in mixing rates as a result of school opening and closing. In Fig. 10 it should be noted that the numbers infected, Y, are always increasing when susceptibles, X, exceed the threshold X_T (thus $R < 1$) and are always on the decrease when $X < X_T$ (when $R < 1$). Hence, infection is being driven by the availability of susceptibles.

Other factors that can promote long-term persistence include the production of infective stages that are able to survive for long periods

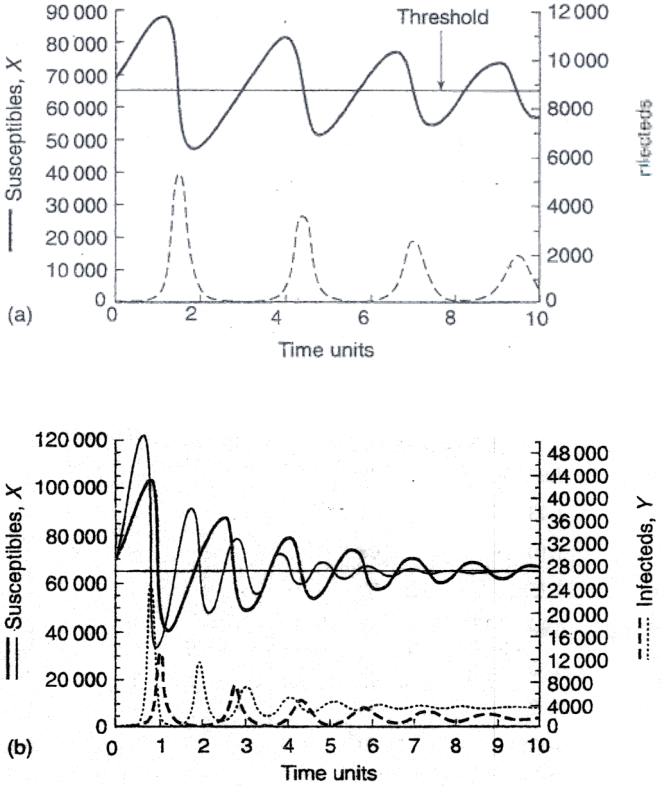


Fig. 10 Conditions for the persistence of an infection in a community. In each case solid curves represent susceptible numbers and dashed lines are infected. (a) Renewal of susceptibles by births (initial conditions: $X = 70\,000$, $Y = 1$, $Z = 930\,000$, $\beta = 0.0004$, $\gamma = 26$, $\mu = 0.02$, $\alpha = 0$, $R_0 = \beta N/(\mu + \gamma) \cong 15$, $X = 65\,050$). Notice that the numbers of susceptibles oscillates above and below the threshold susceptible number, X_T (marked) and that each epidemic starts when susceptible numbers exceed the threshold, X_T, and subsequently decays as susceptibles fall below the threshold, X_T. Oscillations of X and Y (and Z, not shown) gradually damp over time towards the predicted equilibrium values X^*, Y^*, and Z^* (see text).
(b) Renewal of susceptibles through waning immunity, at rate α of 0.05 (thick lines) or 0.1 (thin lines) (corresponding average durations of immunity are 20 and 10 units of time respectively) (other initial settings as for (a) above except for no mortality, i.e. $\mu = 0$). Notice that for the two different rates of loss of immunity the equilibrium susceptible numbers are the same ($X^* = X_T$) since waning immunity has no impact on R_0. However, a higher rate of loss of immunity does result in an increase in numbers of infecteds at equilibrium, Y^*.

in the external environment, sexual transmission, vertical transmission, from mother to unborn offspring, vector transmission, and the carrier state in which some individuals (for genetic or other reasons) atypically harbour the infection for long periods of time (see Table 1 and Fig. 3 for examples).

Herd immunity and mass vaccination

When an infection is persisting endemically in a community such that the net rate at which new cases of infection arise is approximately equal to the net rate at which individuals recover and acquire immunity, the effective reproductive rate R is equal to unity in value. This is known as endemic equilibrium. In practice, for many common viral and bacterial infections the incidence of infection fluctuates both on seasonal and longer-term cycles. The effective reproductive rate therefore fluctuates below and above unity in value as the incidence and density of susceptibles change (see Figs 6 and 10). However, the average value over a series of incidence cycles (both seasonal and longer term) will be approximately equal to unity in the absence of control intervention or changing social and demographic patterns. The effective reproductive rate is reduced below the basic reproductive rate in relation to the fraction of contacts that are with susceptible individuals $x = X/N$, i.e. by the simple equation

$$R = R_0 x. \tag{26}$$

At equilibrium when R is on average unity, the proportion susceptible represents a threshold, x^*, below which infection rates would decline (see Fig. 10). Thus from eqn (26) we see that at equilibrium this proportion susceptible x^* is equal to the reciprocal of the basic reproductive rate R_0 (i.e. $R_0 \approx 1/x^*$). The magnitude of x^* (and therefore R_0) can be determined from cross-sectional serological surveys given data on the age structure of the population. If x_i is the proportion susceptible in age class i and pi is the proportion of the population in the same age class then

$$= \sum_{i=1}^{n} x_i p_i \tag{27}$$

in a population with n age classes. This assumes that the serological profile is unchanging over time (Box 1).

To block transmission and eliminate an infection it is necessary to raise the level of herd immunity by mass vaccination such that the magnitude of the effective reproductive rate is less than unity in value. If x^* is the threshold susceptible proportion then $1 - x^*$, which we call p_c, represents the herd immunity threshold. Vaccinating a proportion of the population $p > p_c$ will lead to elimination of the infection. Therefore this quantity is a critical level for mass vaccination and since, from eqn (26), $x^* = 1/R_0$, p_c may be related to the basic reproductive rate in the following way:

$$p_c = 1 - 1/R_0. \tag{28}$$

The relationship between p_c and R_0 is depicted diagrammatically in Fig. 11: the larger the magnitude of the infection's transmission potential (as measured by R_0) the greater the proportion of the population that must be immunized to block transmission. Note that it is not necessary to vaccinate everyone in the community to prevent the spread of infection. The principle of herd immunity implies the indirect protection of the individual conferred by the protection

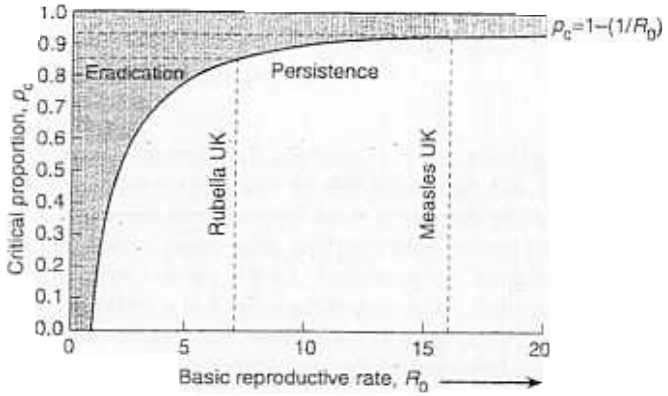


Fig. 11 Relationship between the proportion of the population vaccinated at or near birth and the likelihood of an infection persisting or, alternatively, being eliminated. Infectious agents with high basic reproductive rates in defined communities will be more difficult to control by mass vaccination as illustrated by the example of measles and rubella in the United Kingdom. (Source: Nokes and Anderson 1988.)

(= vaccination) of the population. The mechanism underlying this concept is that of the critical density of susceptibles required to maintain the magnitude of the reproductive rate above unity in value.

Age at vaccination

In general, immunization programmes are introduced by focusing on cohorts of children such that the level of immunization coverage is built up over many years of routine vaccination, that is as children pass some age gateway. In these circumstances, p_c of eqn (28) must be interpreted as the proportion of each cohort vaccinated as soon after birth as is practically feasible, taking account of the need to immunize after the decay in maternally derived specific antibody. For most viral infections the average duration of protection against infection provided by maternal antibodies is approximately 6 months. Clearly, it will take many years of cohort immunization to achieve the desired level of artificially induced herd immunity. A further complication is that it is often the case that the average age at vaccination is higher than what is epidemiologically ideal, resulting from the desire to link vaccination with a delivery opportunity (such as first attendance at school) or variation in the age of delivery resulting perhaps from inefficiency in the co-ordination system or motivation of the population. In this case, simple mathematical models suggest that the level of vaccination coverage required to eradicate the infection under a policy which vaccinates (with a vaccine with 100 per cent efficacy) at an average age of V years is

$$p > [1 + (V/L)]/[1 + (A/L)] \tag{29}$$

where L is human life expectancy and A is the average age at which the infection was acquired prior to the introduction of vaccination (Anderson and May 1983). It is clear from this expression that transmission cannot be interrupted unless the average age at vaccination, V, is less than the average age at infection, A, prior to control.

Imperfect vaccines

Various forms of vaccine failure can be specified. At the time of delivery only a proportion of individuals may respond by generating protective immunity postimmunization. This has been called vaccine

'take' (McLean and Blower 1993). In addition, a proportion of those who initially 'take' still may not be able to fend off an infection if exposed. This might be thought of as an exposure–dose-dependent phenomenon and a vaccine exhibiting this effect might be said to provide only a 'degree' of protection. Finally, vaccine-induced immunity may wane with the passing of time such that a previously protected individual once again becomes susceptible. Therefore a vaccine may only confer protection for a particular duration.

The impact of these three vaccine failings is to reduce the effectiveness or impact of a specified level of vaccination coverage and therefore increase the level of coverage required to achieve elimination of the infection. This new required vaccination proportion, $p\hat{}$, of an imperfect vaccine, may be related to the critical proportion that needs to be effectively vaccinated for elimination, p_c, in the form

$$p\hat{} = p_c/\varphi \tag{30}$$

where φ is the effective vaccine efficacy defined as

$$\varphi = \omega_1\omega_2\,[\mu/(\mu + \omega_3)] \tag{31}$$

(McLean and Blower 1993). Here ω_1 is 'take', ω_2 is 'degree', $1/\omega_3$ is 'duration' (i.e. ω_3 is the rate of waning vaccine-induced immunity), and μ is the death rate. The effects of 'take' and 'degree' are clearly going to be in direct proportion to their magnitude. For example, if a vaccine is being used to interrupt transmission of an infection with an R_0 value of 10 (for which, from eqn (28), the critical proportion to be effectively immunized, p_c, is 0.9) but only 90 per cent of those vaccinated respond (i.e. $\omega_1 = 0.9$), then the new proportion needing to be vaccinated is 0.9/0.9 = 1.0, i.e. 100 per cent coverage. The effect of a vaccine waning over time is less obvious but may be seen from Fig. 12. Here vaccine impact f due to the waning immunity effect (for a vaccine which has perfect 'take' and 'degree') is related to the rate of loss of vaccine-induced immunity, expressed as the time taken for 10 per cent of those vaccinated to lose protection. We can see from this graph that even a slight waning of immunity may cause a very significant reduction in the impact of a vaccine, for example if 10 per cent of those effectively vaccinated at birth lose their immunity by age 30 years

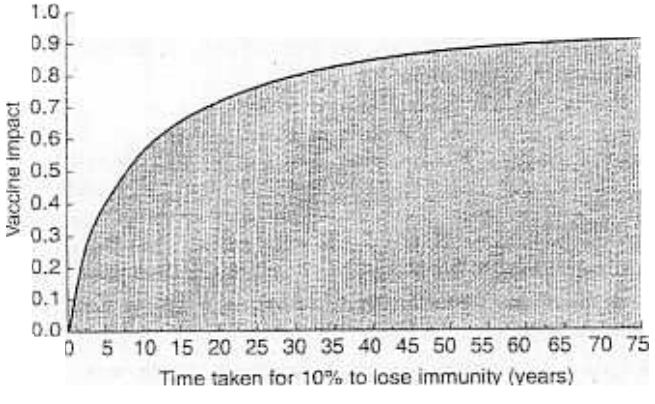


Fig. 12 The impact of an imperfect vaccine. The time taken for 10 per cent of individuals to lose their immunity after vaccination is related to the impact of a vaccine, $\varphi = \mu/(\mu + \omega)$ (the vaccine is assumed to have perfect 'take' and 'degree'). Here, the proportion whose immunity has waned in τ years, p_τ, is related to the rate of loss of immunity by the expression $p_\tau = \exp(-\omega\tau)$.

vaccine impact is reduced by 20 per cent (i.e. $\varphi = 0.8$ from Fig. 12) and in our above example of an infection with $R_0 = 10$, 100 per cent vaccination in infants would not be sufficient to eliminate transmission.

As a final note of caution, the components that make up the term vaccine impact, φ, have a compounding effect since they are multiplied by one another. Thus even if each individually is of little significance, the compounded effect on impact may still be very significant.

The prevalence of infection and the basic reproductive rate

A further epidemiological feature arising from the existence of a critical density of susceptibles to maintain infection concerns the relationship between the magnitude of the basic reproductive rate and the prevalence of infection in a population in which the infectious agent persists endemically. As depicted in Fig. 13 simple models predict that the relationship is non-linear such that a marked reduction in the endemic prevalence or incidence will only occur as the transmission potential is reduced to an extent where it approaches the threshold level $R_0 = 1$. The practical implication is that we should not expect the decline in the incidence of infection induced by mass vaccination to be directly proportional to the level of vaccination coverage. The greatest changes are predicted to occur when coverage attains high levels.

Interepidemic period T

Many viral and bacterial infections that induce lasting immunity to reinfection and which have high transmission potentials (R_0 large) tend to exhibit oscillatory fluctuations in incidence. A good example is that of measles which in the United Kingdom prior to mass

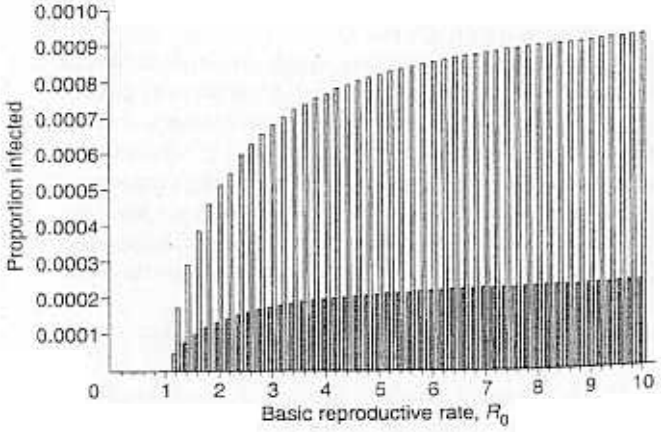


Fig. 13 Predicted changes in the equilibrium proportion of a population infected (i.e. the stable endemic prevalence of an infection) as the transmission potential of the microparasitic agent varies. For infections where there is no loss of immunity, the level of the plateau of prevalence is dependent upon the rate of input of new susceptibles (i.e. the birth rate, b) and the duration of infectiousness, $1/\gamma$. In the figure, $b = 1/75$ per year and $\gamma = 52$ per year (i.e. a 1-week infectious period) (closed bars) or $\gamma = 13$ per year (i.e. a 4-week period) (open bars). An important point to observe is that the greatest changes in the proportion infected occur over the first few increments of R_0 (irrespective of the magnitude of b or γ.)

vaccination oscillated on a seasonal basis (owing to the aggregation and desegregation of children for school term and holiday periods) and a longer-term 2-year cycle with years of high incidence separated by years of low incidence (Anderson *et al.* 1984) (Fig. 6). Time-series analyses reveal that these longer-term cycles for infection such as measles, mumps, rubella, and pertussis are not due to chance fluctuations but arise as a result of the dynamic interaction between the net rates of acquisition of infection and immunity on recovery.

Simple models based on the mass-action assumption suggest that the interepidemic period, T, of the longer-term cycles is determined by the generation time of the infection, k, defined as the sum of the latent and infectious periods and the transmission potential of the infection inversely measured by the average age at infection, A, where

$$T = 2\pi(AK)^{1/2}. \quad (32)$$

This simple prediction matches well the observation for a variety of common childhood infections prior to mass vaccination (i.e. the 2-year cycles of measles, the 3-year cycles of mumps, the 4- to 5-year cycles of rubella, and the 3- to 4-year cycles of pertussis). Non-seasonal oscillation arises as a consequence of the exhaustion of a supply of susceptibles, as an epidemic passes through a population, plus the time lag that arises before new births replenish the pool to trigger the next epidemic. As such the interepidemic period is also influenced by the birth rate of the community (which influences the average age of infection, A, in eqn (32)). For example, in developing countries such as Kenya with high birth rates, measles tends to cycle on a 1-year time scale in urban centres as opposed to the 2-year cycle in the United Kingdom prior to control (McLean and Anderson 1988*b*).

Parameter estimation

Survey data

Survey data on the incidence or prevalence of infection (past or current) can be obtained in a variety of ways. Longitudinal (= through time) data can be acquired by monitoring a cohort of people through time and recording infection as it occurs. Horizontal (= one point in time) – cross-sectional (= across age and sex classes) can be acquired by a survey at one point in time or over a short interval of time, by the examination of different age classes within the population. Such surveys are of most use when based on serological examinations to determine the proportion of individuals in a given age class who have antibodies specific to the antigens of a particular infectious agent. These cross-sectional serological profiles reflect the proportion in each age class who have, at some time in the past, experienced infection. Case-notification data stratified by age and sex, and recorded over a set interval of time such as 1 year, can be accumulated to indicate what proportion of the cases occurs by any given age. This may then be used to infer changes in the proportion who experience infection as a function of age. Such data are clearly less reliable than serological information since they are dependent on a lack of bias in reporting efficiency by age class. Bias is to be expected if the seriousness of the disease induced by infection changes with age (e.g. rubella in women and mumps in men) or where the incidence of subclinical (i.e. undetectable) infections is age dependent.

An alternative to the use of serum for the detection of specific antibodies to infectious agents is saliva. More specifically, when looking for systemic antibodies (e.g. immunoglobulin G and M) the fluid which collects around the gums and under the tongue (as distinct from salivary gland secretions) is rich in serum antibodies. This is known as gingivocrevicular exudate or secretion. The disadvantage of using salivary fluid for antibody detection is the low concentration of immunoglobulin it contains relative to serum. Immunoglobulin G in whole saliva is approximately 1000-fold less concentrated than in serum, although in crevicular fluid it may only be five-fold more dilute (Mortimer and Parry 1991). In recent years highly sensitive assays have been developed to overcome the dilution problems and, accompanied by developments in devices for the collection of crevicular fluid samples, have now been successfully employed in the detection of antibodies to a variety of infections, including measles, rubella, and mumps (Perry *et al.* 1993; Brown *et al.* 1994) human parvovirus, and hepatitis B virus (core antibodies) (Parry *et al.* 1989), and HIV (Holmstrom *et al.* 1990; Behets *et al.* 1991; Van Den Akker *et al.* 1992).

The advantages of using saliva over serum are numerous and associated largely with the collection procedure. For example, sampling is non-invasive and is more acceptable which will assist in response level, the collection process is easier and can be carried out by non-technical personnel, and there is lower risk to both subject and investigator (Mortimer and Parry 1991). Surveys based on saliva collection offer great potential in the fields of epidemiology and surveillance, including the measurement of population immunity in the evaluation of the impact of vaccination programmes of infection prevalence in assessing the rate of spread of infections, such as HIV, through communities. The opportunity for longitudinal surveillance will be beneficial to studies of spatial and temporal patterns of disease spread and salivary diagnosis will become increasingly useful in outbreak investigation and control.

When conducting surveys a number of points should be borne in mind. Firstly, sample sizes should be as large as practically possible, finely stratified by age (preferably infants to elderly people). How large will depend upon what we wish the accuracy or power (see Sokal and Rohlf 1981) of subsequent analyses to be, but 25 to 50 per yearly age class is a rough working estimate. Secondly, the incidence of infection may oscillate on a seasonal or longer-term basis. As such, it is good practice to carry out surveys that span epidemic and interepidemic years. Thirdly, systematic changes through time may occur in a given population due to social, behavioural, economic, or other changes. Examples include the observed reduction in the incidence of hepatitis A in northern European countries over the past few decades due to improved standards of hygiene and the rise in the incidence of gonorrhoea in certain developed countries during the 1960s and 1970s due to changes in sexual behaviour (e.g. increased rates of sexual-partner change). Basic reproductive rates and rates of infection may therefore change through time irrespective of the impact of control measures.

The basic reproductive rate of infection

Estimating individually the component parameters that determine the magnitude of the basic reproductive rate, R_0, is fraught with many problems. In the case of directly transmitted viral and bacterial infections, we require a knowledge of the transmission coefficient, β, the density of susceptibles, X, and the average duration of infectiousness. In practice it is often easier to use indirect methods to arrive at estimates of R_0 employing serological data finely stratified by age. As discussed earlier, the rate of decay with age in the proportion susceptible to infection provides measures of the age-dependent forces

of infection ($\lambda(a)$). These in turn can be used to obtain an estimate of the average age, A, at which an individual typically acquires infection. Mathematical models can be used to define a relationship between the magnitude of R_0 and the average age at infection. In the simplest case the relationship is of the form

$$R_0 = Q/A \tag{33}$$

where Q denotes the reciprocal of the net birth rate of the community. In developed countries where net births are approximately equal to net deaths the quality Q is equal to the average life expectancy (from birth) L (Anderson and May 1985*a*). More generally, if maternally derived antibodies provide protection for an average of F years R_0 is related to A by the expression

$$R_0 = Q/(A - F). \tag{34}$$

A simple example of the use of this equation is provided by the transmission of the measles virus in the United Kingdom prior to the introduction of mass vaccination. In this case the values of A, L, and F were 5 years, 75 years, and 0.5 years respectively, leading to an R_0 estimate of between 16 and 17. The inverse relationship between R_0 and A makes good intuitive sense—infections with high transmission potentials will tend to have low average ages at infection and vice versa. These notions are depicted diagrammatically in Box 1, and Table 3 lists some estimates of R_0, A, L, and the critical level of vaccination coverage to block transmission, p_c, for a variety of common infectious agents in defined localities.

An alternative method to that outlined above is based on the prediction of simple models that the magnitude of R_0 is related to the fraction of the population susceptible to infection, x^* when the infection has attained its endemic equilibrium. The relationship is simply

$$R_0 = 1/x^* \tag{35}$$

and arises from the fact that at equilibrium the effective reproductive rate is equal to unity in value (see eqn 26). Note that eqns (33) and (35) imply that the average age at infection, A, is inversely related to the equilibrium fraction of susceptibles in a population, x^* required to ensure each primary case gives rise on average to at least one secondary case (Box 1). In general, however, the method based on estimating the average age at infection is the better one given good age-stratified serological data.

Latent and infectious periods

Two sources of data are available to estimate latent and infectious periods. The first derives from clinical, virological, and immunological

Table 3 Epidemiological parameters for a variety of childhood infections in developed countries in the absence of mass vaccination

Infection	Average age at infection A (years)	Location and date	Data type	Life expectancy L (years)	R_0[a]	p_c (%)
	5.0	England & Wales, 1948–68	Case notifications	70	15.6	94
	5.5	USA, large families, 1957	Serology	70	14.0	93
	8.0	USA, small families, 1957	Serology	70	9.3	89
Whooping cough	4.5	England & Wales, 1944–78[b]	Case notifications	70	17.5	94
	4.9	USA, urban, 1908–17	Case notifications	60	13.6	93
	6.5	USA, rural, 1908–17	Case notifications	60	10.0	90
Chickenpox	8.6	USA, urban, 1913–17	Case notifications	60	7.4	86
	6.8	USA, urban, 1943	Case notifications	70	11.1	91
Mumps	7.0	UK, urban, 1977	Serology	75	11.5	91
	5.7	Netherlands, urban, 1980	Serology	75	14.4	93
	9.9	USA, urban, 1943	Case notifications	70	7.4	86
Diphtheria	10.4	USA, 1912–28	Case notifications	60	6.1	84
Rubella	10.8	England, urban, 1980–84[c]	Serology	75	7.3	86
	10.2	GDR, 1972	Serology	70	7.2	86
Scarlet fever	8.0	USA, urban, 1908–17	Case notifications	60	8.0	88
	12.3	USA, rural, 1918–19	Case notifications	60	5.1	80

Parameter definitions given in text (data from a variety of sources).

[a] $R_0 = L/(A - F)$ where F is duration of maternally derived protection, assumed to last for 6 months in all cases. Note that no consideration of age-dependent forces of infection is given (see text).

[b] Encroaches on to vaccination era.

[c] Male serology—only females vaccinated under selective immunization policy.

studies of the course of infection in individual patients. For some common microparasitic infections, the presence of the infectious agent in host tissues, excretions, and secretions can be directly assessed. Durations of antigenaemia in body fluids and secretions or of infective particles in specific cells will, in many instances, reflect the period over which an infected person is infectious to others (although this is, of course, not always the case as, for example, with the latent herpesviruses).

Alternatively, statistical methods can be employed in the study of transmission within small groups of individuals. The classic data on measles, collected by Hope Simpson (1952) in the Cirencester area of England during the years 1946 to 1952, record the distribution of the observed time interval between two cases of measles in 219 families with two children under the age of 15 years. The bulk of these observations represent case-to-case transmissions within a family. However, in a small number of families, where the observed interval is only a few days it may be assumed that these cases are double primaries, both children having been simultaneously infected from some outside source. Statistical methods, based on chain binomial models, can be used to derive estimates of the latent, infectious, and incubation periods (Bailey 1973). A rough guide to these periods for various common viral and bacterial infections is presented in Table 4. Some of these estimates are based on detailed analyses of case to case data while others are more speculative.

Sexually transmitted infections

Rather different problems in parameter estimation, to those outlined above, are presented by sexually transmitted infections. By way of an illustration and given the topicality of the infection, we focus on HIV.

The characteristics of most sexually transmitted diseases cause their epidemiology to differ from that of common childhood viral and bacterial infections. Firstly, the rate at which new infections are produced does not appear to be closely correlated with population density. Secondly, the carrier phenomenon in which certain individuals harbour asymptomatic infection is often important. Thirdly, many sexually transmitted diseases induce little or no acquired immunity on recovery. Fourthly, net transmission depends on the degree of heterogeneity in sexual activity prevailing in the population and the degree to which individuals in one sexual activity class (perhaps defined in terms of the rate of sexual-partner change) mix with those in the same and in different classes (i.e. 'who has sex with whom').

The basic reproductive rate, R_0, in its simplest form is determined by the transmission probability, B, multiplied by the effective rate of sexual-partner change, c, multiplied by the average duration of infectiousness, D. Heterogeneity in sexual activity is a major influence on the magnitude of transmission success. Recent national surveys of sexual attitudes and lifestyles suggest that most people have few different sexual partners and a few have many (Anderson and May 1988; ACSF 1992; Johnson *et al.* 1994). The distributions of reported numbers of sex partners per defined period of time therefore tend to be skewed with a long right-hand tail where a few individuals report many partners (Fig. 14). As pointed out earlier, under these circumstances the variance in partner numbers, s^2, is much greater in value than the mean, m, and the effective rate of sexual partner change, c, is defined as $c = m + (s^2/m)$ (as in eqn 14). It follows that those with high rates of sexual partner change play a disproportionate role (relative to their proportional representation in a community) in the spread of infection. In the case of HIV each component of R_0 is difficult to measure due to the sensitivity and the practical difficulties associated with the study of sexual behaviour and the long and variable incubation period of the disease AIDS induced by the infection. Over the long incubation period infectiousness appears to vary widely for an individual and between individuals.

As a consequence some indirect measure of transmission potential is required. Mathematical models of transmission suggest that the doubling time t_d (the average time over which the number of cases of infection doubles) of an epidemic of HIV in a defined risk group (e.g. male homosexuals), during the early stages of the epidemic, is related to the magnitude of R_0 by the equation

$$t_d = D\ln(2)/(R_0 - 1) \tag{36}$$

where D denotes the average duration of infectiousness. Current estimates of the incubation period of HIV suggest a mean period of

Table 4 Average duration of infection classes for a variety of microparasites

Infectious disease	Latent period $1/\sigma$ (days)	Infectious period $1/\gamma$ (days)	Incubation period[a] (days)
Measles	6–9	6–7	11–14
Chickenpox	8–12	10–11	13–17
Rubella	7–14	11–12	16–20
Hepatitis A	13–17	19–22	30–37
Mumps	12–18	4–8	12–26
Polio	1–3	14–20	7–12
Smallpox	8–11	2–3	10–12
Influenza	1–3	2–3	1–3
Scarlet fever	1–2	14–21	2–3
Whooping cough	6–7	21–23	7–10
Diphtheria	2–5	14–21	2–5

[a]Time to appearance of symptoms.

Source: Anderson (1982*b*).

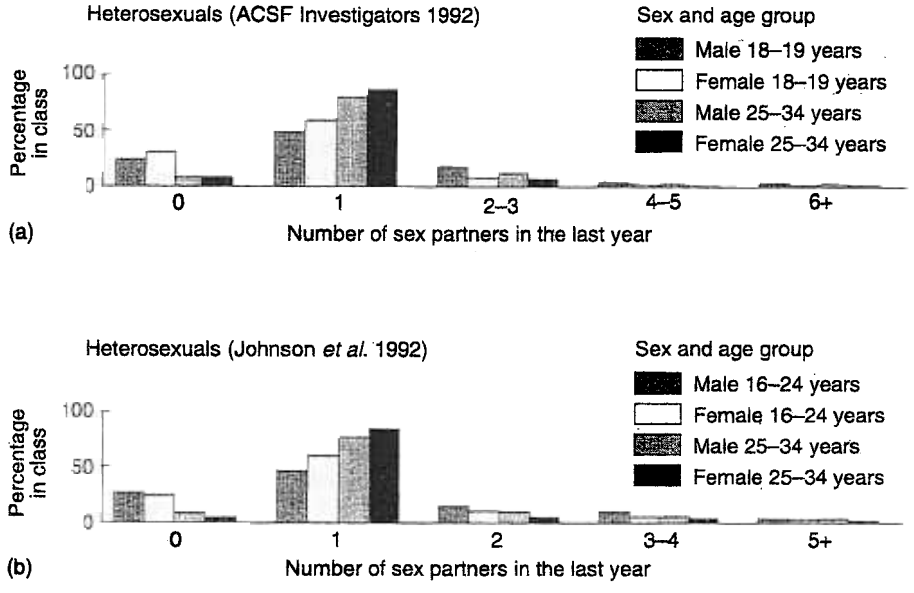


Fig. 14 Frequency distributions of the reported number of different sexual partners over the past year in two surveys in (a) France (ACSF 1992) and (b) Britain (Johnson *et al.* 1992) of sexual attitudes and lifestyles, stratified by age and sex. The similarities in the results of the two surveys are striking.

approximately 8 years. It is probable that the average infectious period is much shorter perhaps of the order of 2 years or so (however, this is uncertain at present; see Anderson and May 1988). If we assume the value of D lies between 2 and 6 years eqn (32) gives estimates of R_0 in the range of 2.7 to 6, given an observed doubling time of around 10 months in male homosexual communities in the United States during the early 1980s (May and Anderson 1987). Of course, this method of estimation is very crude, but it provides a rough guide to the degree to which sexual habits must change in order to reduce the magnitude of R_0 below unity in value (i.e. by a factor of 3 to 6).

More generally, certain of the parameters that determine the magnitude of R_0 may vary between the sexes. This is certainly the case for gonorrhoea (Hethcote and Yorke 1984) and it may be true for HIV. In these circumstances, when considering transmission via heterosexual contact, the basic reproductive rate adopts the form

$$R_0 = (B_1 B_2 c_1 c_2 D_1 D_2) \tag{37}$$

where the subscripts 1 and 2 denote males and females respectively.

Further complications arise in the definition of the case reproductive number, R_0, when we take into account the pattern of mixing between different strata of the sexually active population. For example, in light of the data presented in Fig. 14 concerning heterogeneity in reported rates of sexual partner acquisition per year in France and Britain, it seems sensible to stratify the population by the rate of sexual partner change into low-, medium-, and high-'activity' classes. The magnitude of any epidemics of a sexually transmitted disease and the endemic level of infection in a community will depend on the degree to which the small number of people with high rates of sexual-partner change mix with the medium- and low-activity classes. If mixing is random across activity classes the infection will be widely disseminated in the community. However, if mixing is highly assortative (i.e. like with like) the infection will tend to be restricted to the small proportion of individuals in the high-activity class (the so-called 'core' group) with a few cases in the other classes. The prevailing pattern of mixing is therefore of great importance in determining the prevalence of an sexually transmitted disease and the degree to which it is disseminated in a defined community. Recent studies of mixing patterns based on contact tracing via sexually transmitted disease clinics suggest that mixing is more assortative than random in character (Garnett and Anderson 1993). Once mixing is taken into account it is necessary to redefine transmission success in terms of the number of secondary cases of infection in group i generated by contact with infectives in group j, R_{0ij}, where

$$R_{0ij} = p_{ij} BcD. \tag{38}$$

Here p_{ij} is the probability that a susceptible in group i has a sexual contact with someone in group j.

Again, more generally, the population is structured by other variables such as age, ethnicity, area of residence, and educational attainment. Here again, behavioural studies suggest a degree of assortative mixing with respect to the choice of sexual partner—except in contact with commercial sex workers.

Models and the design of control programmes

Mathematical models can be of help in defining the targets for a control programme, in interpreting observed epidemiological changes under the impact of control, and in discriminating between different approaches (Nokes and Anderson 1987, 1988, 1991, 1992, 1993; Garnett *et al.* 1992; Gupta and Anderson 1992). In this section we consider two themes, namely, the design of mass vaccination

programmes to control childhood viral and bacterial infections and education to induce changes in sexual behaviour to control sexually transmitted diseases.

Impact of mass vaccination

In practical terms, the level of vaccination coverage in a given community or country is determined by a variety of economic and logistical factors (developing countries) or motivational and legislative issues (industrialized countries). However, models can define the ideal goal of a given programme. We have already outlined the relationship between the critical level of vaccination coverage required to block transmission, pc and various epidemiological (R_0), demographic (net birth rate and life expectancy, Q and L), and logistical (V, the average age at vaccination) parameters (see eqns (28) and (29) and Table 3) and vaccine properties (see eqns (30) and (31)). In many instances, the high transmission potentials of common childhood viral and bacterial infections imply very high levels of infant vaccination coverage if transmission is to be interrupted. If vaccine efficacy is less than 100 per cent (e.g. the current pertussis vaccines), then problems may arise in attaining these targets even if legislation enforces vaccination of all children before entry to school (as in the United States). Models emphasize the point that to obtain the best effects very high levels of coverage should be aimed at with vaccination at as young an age as is practically feasible given the complications presented by the presence of maternally derived antibodies in infants.

Aside from defining targets for vaccination coverage, models can assist in interpreting the impact of a given programme on epidemiological parameters such as the incidence of infection, the average age at infection, and the interepidemic period. In a later part of this section we consider the principles underlying an alternative approach to mass vaccine intervention, that of pulsed immunization across age cohorts, which has recently met with such success in controlling polio and measles in Central and South America.

Incidence of infection

Immunization has the direct effect of reducing the number of cases of infection as a result of the protection of the vaccinated individuals ($X \rightarrow Z$, see Fig. 3(d)). Since this reduces the number of infectious persons in the vaccinated population, an indirect effect is a reduction in the net rate of transmission of the virus or bacterium. This is the principle of herd immunity, where susceptibles gain protection from the vaccinated proportion of the population. Provided the infection is able to persist endemically (i.e. the level of coverage is less than that required for eradication), models suggest that the equilibrium proportion of susceptibles in the population will remain constant irrespective of the level of coverage below the critical point for eradication. This prediction is illustrated diagrammatically in Fig. 15. The level of coverage simply reduces the proportion of seropositive individuals who have acquired immunity via infection as opposed to via vaccination. As mentioned earlier (see Fig. 13) the manner in which the incidence declines as the level of coverage rises is non-linear in form with the most dramatic reductions occurring as the proportion vaccinated approaches the critical point for the interruption of transmission. As the level of coverage approaches the critical point the proportion of immune persons who possess vaccine-induced immunity approaches unity.

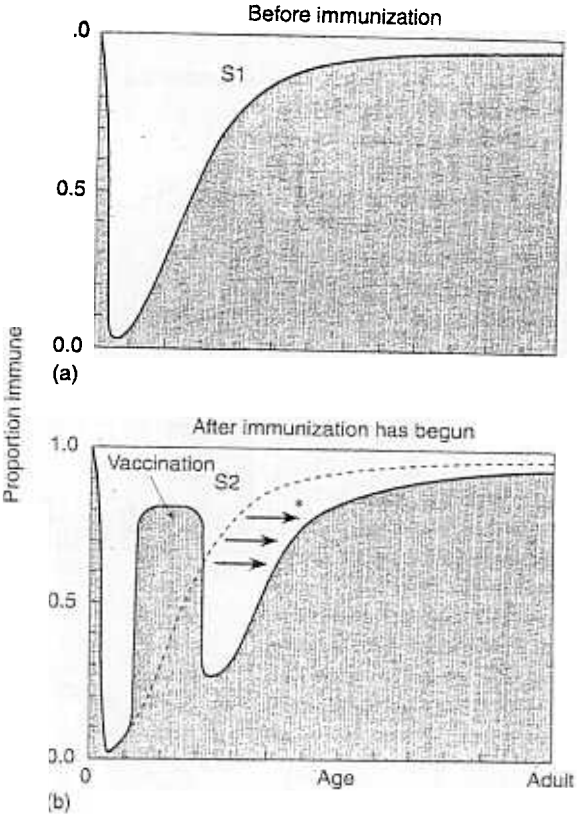


Fig. 15 Diagrammatic representation of the predicted impact of mass immunization (against a typical childhood viral or bacterial infection) on the age distribution of susceptibility in a population. Before immunization (a) there is a 'valley' of susceptibles (S1) in the young age classes. Attempts to fill in this valley by vaccination (b) reduces the rate of transmission of the infection thus lowering the probability of unvaccinated individuals being infected. As a consequence there is an upward shift in the ages of susceptibles (*) from that pertaining before vaccination (dotted line). Two points are important: (i) the number or proportion of susceptibles after immunization has begun (area S2) is roughly unchanged from that which existed before immunization (area S1) and (ii) the average age of susceptibles increases. (Source: Nokes and Anderson 1988.)

The average age at infection

As a direct result of reducing the net rate of transmission, vaccination acts to increase the average age at which susceptibles acquire infection over that pertaining prior to control (i.e. by reducing the probability of coming into contact with an infectious person). Observation now bears out the expectation of an increased average age of susceptibles and of infection as a result of mass vaccination programmes. The example in Fig. 16 shows the prevaccination (1982) serological profile (or distribution of susceptibles by age) for rubella in Finland (Fig. 16(a)) and the profile (for males only) in 1986, 4 years after mass infant measles, mumps, and rubella (MMR) vaccine was introduced (Fig. 16(b)) (Ukkonen and Von Bonsdorf 1988). The similarity with Fig. 15 is striking. Also shown in Fig. 16 is the changing distribution of diagnosed rubella cases, with a marked increase in the average age. Later we discuss how this change in the age distribution of the incidence of infection can influence the incidence of disease arising

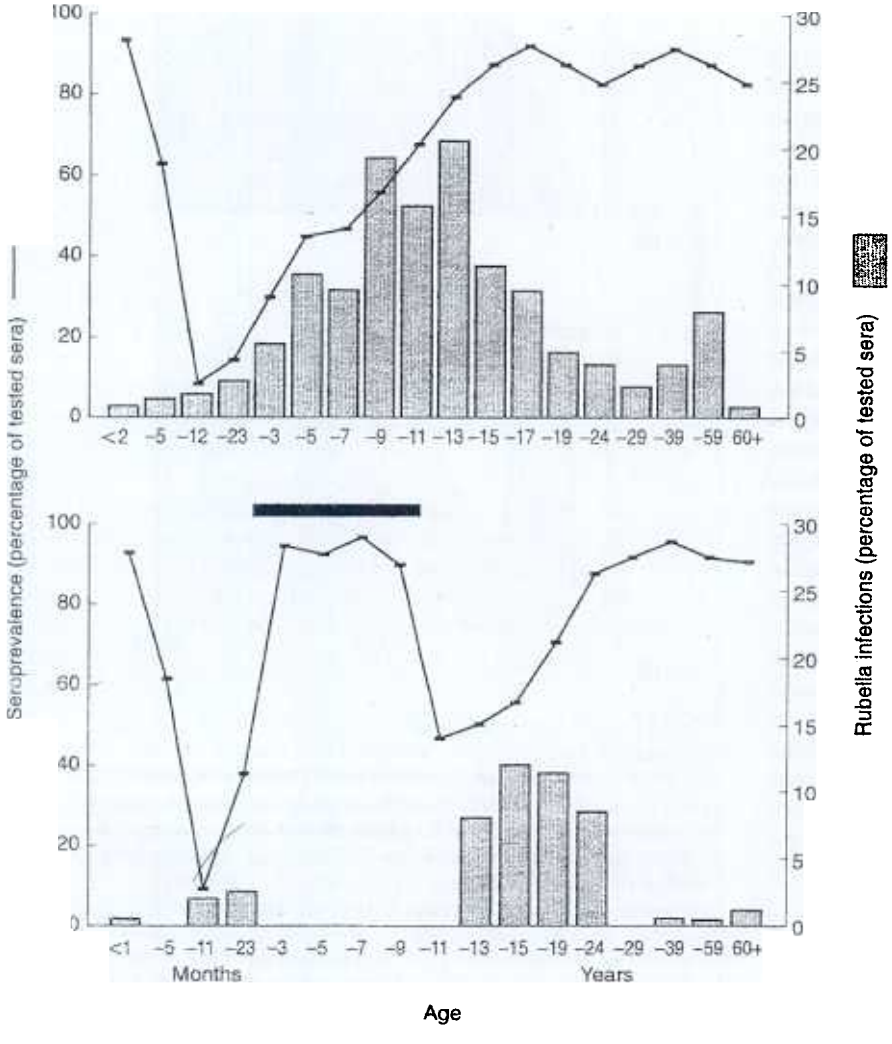


Fig. 16 The observed impact of mass immunization against rubella in Finland (Ukkonen and Von Bonsdorf 1988). (a) The prevaccination (1982) age seroprevalence of specific rubella antibodies (line) with the age distribution of diagnosed cases (bars). (b) Four years after mass infant MMR vaccination was introduced with the age range affected shown by the solid bar (data for males only) (other details as for (a)). (Source: Nokes and Anderson 1993.)

from infection if older people differ in their vulnerability to complications and concomitant morbidity when compared with younger people.

Interepidemic period

Simple models also predicted that a reduction in the transmission rate in a vaccinated population will act to lengthen the interepidemic period over that pertaining prior to control (Anderson and May 1983). This may be shown easily using our model in Fig. 10 if a proportion of all individuals entering the population are vaccinated at the time of birth (starting from time unit 5 onwards), resulting in an increase in the time taken for susceptibles to build up to threshold numbers and, hence, an increase in the interval between epidemics (Fig. 17). This pattern has been observed in various vaccinated communities (Fig. 18).

Cautionary notes

The changes in epidemiological patterns of infection induced by vaccination are not always beneficial. An increased interepidemic period, for example, can induce complacency in the community with respect to the need to maintain high levels of vaccination coverage. Motivating parents to ensure that their children are vaccinated during long periods of low incidence (the troughs in the epidemic cycle) can be problematic particularly if there is some small but measurable risk associated with vaccination. At the start of a mass immunization programme the probability of serious disease arising from vaccination is usually orders of magnitude smaller than the risk of serious disease arising from natural infection. As the point of eradication is approached, the relative magnitudes of these two probabilities must inevitably be reversed. The optimum strategy for the individual (not to be vaccinated) therefore becomes at odds with the needs of society (to maintain her immunity) (Nokes and Anderson 1991). This issue—

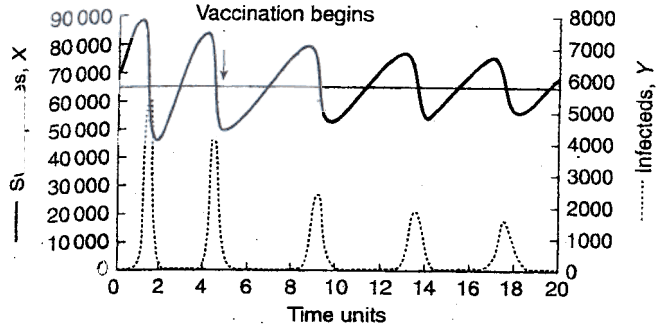


Fig. 17 Predicted impact of vaccination on the interepidemic period. Vaccination of 50 per cent of all births was introduced at time 5 into the model given in Fig. 10 (with the same initial settings).

which was central to the decline in the uptake of pertussis vaccine in the United Kingdom during the mid to late 1970s—can be overcome by legislation to enforce vaccination (as in the United States), but its final resolution is only achieved by global eradication of the disease agent so that routine vaccination can cease.

Other problems concern doubts over the role played by exposure to natural infection in boosting vaccine-induced immunity and, in some cases, worries over the duration of protection provided by vaccination. If enough is understood about these problems mathematical models could be used to decide whether or not to revaccinate a proportion of the immunized population and, if so, what is the best age to revaccinate. Similarly, recent evidence for measles suggests that passive immunity in infants of mothers whose own protection was vaccine derived wanes more rapidly than in infants whose mothers were naturally infected (Markowitz *et al.* 1996). The consequences of this are that, on the one hand, infants become susceptible to infection at an earlier age than was previously the case, but, on the other hand, it may allow for the lowering of the age of vaccine delivery. The merits of this latter issue could well be addressed using mathematical models.

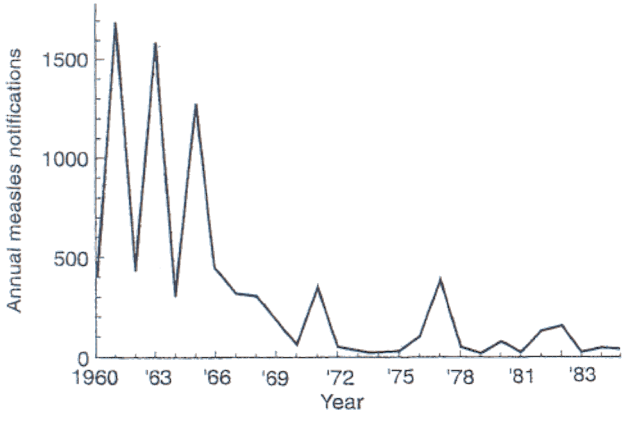


Fig. 18 Annual measles notifications for the city of Oxford, England, for the period 1960 to 1985. The introduction of measles vaccination in 1966 has resulted in a significant increase in the period between epidemics. (Source: Office of Population Census and Surveys, London.)

Variation in vaccine uptake

Ideally, vaccination coverage should be high and constant both through time and in different regions of a country. In practice, however, this is rarely the case. With respect to time, once incidence is reduced to a low level, problems can arise in stimulating public health workers to maintain coverage at high levels. More importantly, after introduction, most immunization programmes show a slow increase in rates of coverage. This obviously results in a delay in experiencing the full benefits and must be recognized in assessing the impact of a given policy. It takes many decades before the full benefits of a cohort immunization programme are manifest. Model simulations of the impact of such programmes on the incidence of infection and disease clearly illustrate this point (Anderson and May 1983, 1985*a,c*). Of greater concern, however, is the variation in vaccine uptake in different regions of a country. Levels of vaccine coverage for sentinel antigens (measles, diphtheria 3, and pertussis 3) in the United Kingdom, for example, varied widely between different regions in the late 1980s (Fig. 19), a problem which has been greatly diminished as a result of improved vaccine programme co-ordination. To block transmission countrywide effectively it is necessary to ensure that the targets laid out in Table 3 are attained in each area. Otherwise, pockets of infection in regions of low uptake will continue to trigger small epidemics in other areas. The upsurge of mumps in certain states in the United States in the late 1980s (Wharton *et al.* 1988) is an example of the potential hazards of spatial variation in vaccine uptake.

Non-uniformity in human population density

Non-uniformity in the spatial distribution of humans, with some people living in dense aggregates and others living in isolated or small groups, can lead to heterogeneity in transmission rates. Models suggest that this can result in the transmission potential of an infection (R_0) being greater on average than suggested by estimation procedures which assume spatial homogeneity (Anderson and May 1984; May and Anderson 1984). Under these circumstances, theory suggests that the optimal solution appears to involve 'targeting' vaccination coverage in relation to group size with dense groups receiving the highest levels of coverage. The optimal programme is defined as that minimizing the total, communitywide number of immunizations needed for elimination or for a defined level of control. This strategy reduces the overall proportion that must be vaccinated to block transmission, compared with that estimated on the assumption of spatial homogeneity. This conclusion has practical significance for the control of infections such as measles and pertussis in some developing countries, where rural–urban differences in population density tend to be much more marked than in developed countries (Anderson and May 1991). It is probable that in many regions of Africa and Asia, diseases such as measles cannot persist endemically in rural areas without frequent movement of people between low-density (rural) and high-density (urban) populations. Under these circumstances, transmission might be blocked in both regions by high levels of mass immunization in the urban centres alone.

Age-dependent factors

Analyses of case-notification records and serological profiles suggest that, for many common infections (measles, rubella, and pertussis), the per capita rate of infection ($\lambda(a)$) depends on the ages of

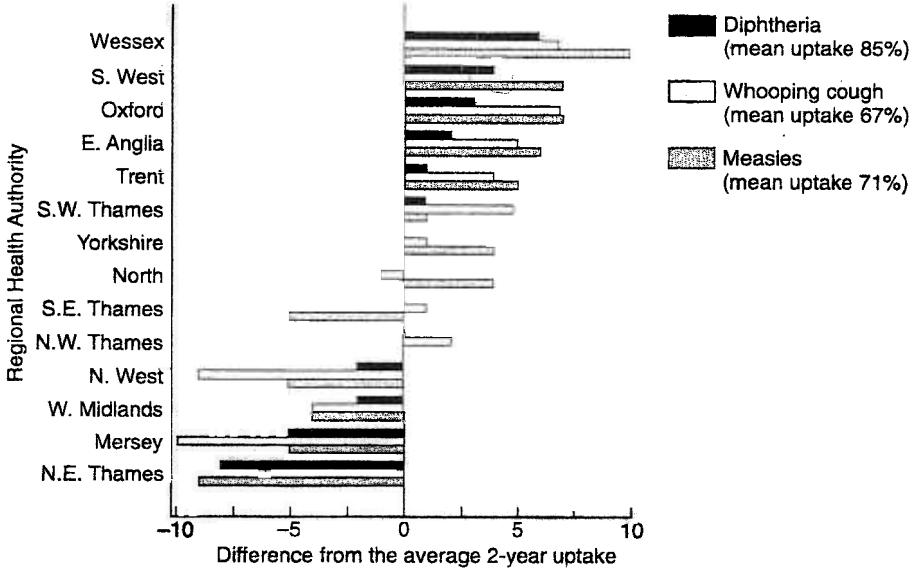


Fig. 19 Regional variation in immunization uptake for sentinel agents in England, 1986. (Source: Nokes and Anderson 1988.)

susceptible individuals, changing from a low level in the 0- to 5-year age classes, via a high level in the 5- to 15-year age classes, back to a lower level in the adult age classes (Fig. 5). This is of interest both because it reflects behavioural attributes of human communities and because of its impact on the predicted level of vaccination required to eliminate transmission. The high levels of the force of infection in the 5- to 15-year-old classes are thought to arise as a consequence of frequent and intimate contacts within school environments (Anderson and May 1985*c*; Nokes *et al.* 1986; Anderson *et al.* 1987). Theoretical studies which take account of age dependence in the force of infection predict somewhat lower rates of vaccination than those arrived at under the simple mass-action assumption (Table 3). However, it should be emphasized that the values listed in Table 3 provide a good first approximation of the targets to be obtained in a vaccination programme. The reason why the observed age-related changes in the force of infection influence the predicted level of coverage relates to the tendency for mass vaccination to shift the age distribution of susceptibility (Figs 15 and 16). Susceptibles who avoid infection and vaccination may move from an age class with a high force of infection into an older class with a lower rate.

Does mass vaccination always reduce disease incidence?

The risk of complications arising from infection is often dependent upon the age at which exposure occurs. The newborn are particularly vulnerable due to their immunological immaturity and are therefore more likely to suffer morbidity and even mortality (Fig. 20). Protection by maternally derived antibody moderates the risk during this time of great vulnerability but, in developing countries, factors such as malnutrition and high incidences of secondary 'opportunist' infections can result in high mortality rates as a result of infant and childhood viral and bacterial infection. In general where the risk of serious disease is higher in the young than old people, mass vaccination will always act to reduce the incidence of disease.

In developed countries case fatalities are much less common and the greatest problem is morbidity and the risk of serious disease. Of particular concern are infections where the risk of severe complications increases with age (Fig. 21). Whether this trend is important depends on the quantitative details of such factors as how risk changes with age, the average age at which the vaccine is administered, the average age at infection, and how the rate (or force) of infection changes with age prior to the introduction of immunization (Knox 1980; Anderson and May 1983, 1985*a*; Anderson *et al.* 1987; Nokes and Anderson 1991) (Fig. 21).

Rubella and mumps are clear examples because of the risk of congenital rubella syndrome in infants born to mothers who

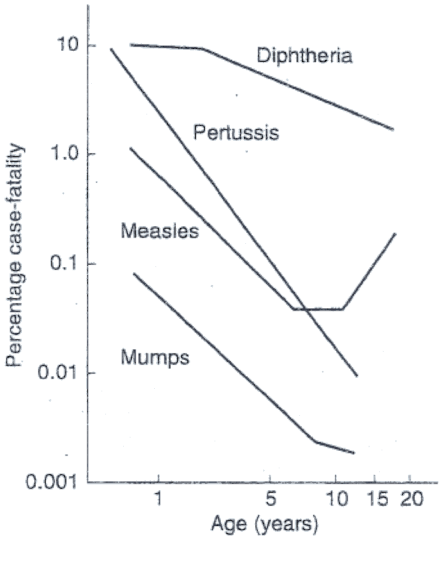


Fig. 20 Age-dependent mortality associated with infection from a variety of childhood viruses and bacteria. (Source: MIMS 1987.)

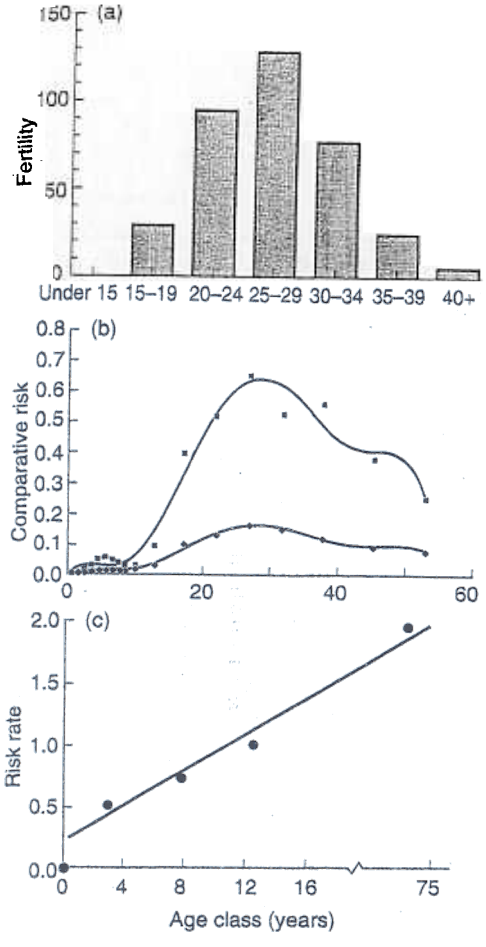


Fig. 21 Age-dependent risk of complications from infection: (a) the likelihood of fetal transmission of rubella virus with concomitant risk of congenital rubella syndrome is directly related to age-specific fertility of women (data for England and Wales 1985 from OPCS Monitor FMI 86/2) (b) changes in the risk of complications from mumps infection in the United Kingdom relative to age and sex. In addition to meningitis and encephalitis, males (■) may suffer orchitis (data from Anderson et *al.* 1987). Note here that the term comparative risk, refers to the risk compared with other age classes. (c) Measles encephalitis per 100 000 cases in the United States. (Source: Anderson and May 1983.)

contracted rubella in their first trimester of pregnancy and the occurrence of orchitis and the associated risk of sterility in post-pubertal males plus infection of the central nervous system following mumps infection. The crux of the problem relates to how mass vaccination changes the age profile of the incidence of infection. Any level of coverage will reduce the incidence of infection but by increasing the average age at which those still susceptible acquire infection certain levels of coverage may increase the incidence of disease. The important question is whether the increase in the proportion of cases in older people will result in an increase in the absolute numbers of cases of serious disease.

This problem has resulted in the adoption of different vaccination programmes against rubella (to control congenital rubella syndrome) in different countries (Table 5). Until the introduction of MMR vaccine in the United Kingdom in 1988, girls only were vaccinated at an average age of around 12 years, so as to allow rubella virus to circulate in males and young females and create naturally acquired immunity in the early years. By contrast, it has always been the case in the United States for both boys and girls to be vaccinated at around 2 years of age, with the aim of blocking rubella virus transmission. Mathematical models predict that the United States policy is best if very high levels of vaccination (80 to 85 per cent of each yearly cohort) can be achieved at a young age, while the United Kingdom policy is better if this cannot be guaranteed (Fig. 22). A mixed policy is predicted to be of additional benefit over the selective policy alone if moderate to high levels of vaccine uptake among boys and girls can be achieved at a young age (60 per cent) (Anderson and Grenfell 1986).

The process of using mathematical models to evaluate the impact of a particular mass vaccination policy in a community is detailed in Box 4, in this case for mumps. At the time of the introduction of MMR infant vaccination in the United Kingdom in November 1988 such studies as these suggested that provided moderate to high levels of coverage (60–65 per cent) could be achieved then the change in policy was unlikely to increase the incidence of serious disease (Anderson *et al.* 1987). Following the implementation of the MMR vaccine, coverage rose from the level of uptake for measles vaccine at the time of around 70 per cent by age 2 years (the level of update for measles vaccine at the time) to 90 per cent within the space of 2 years. Thoughts have now turned to the required strategy for elimination of these three infections and use is being made of mathematical models to explore the possible options, such as a two-dose schedule (Babad *et al.* 1995). For rubella, and specifically for the issue of when to remove the selective arm of the vaccination strategy, we now have the example from the Scandinavian countries to guide our policy. Data from Finland, for example, clearly show the need to continue schoolgirl vaccination until the cohorts with high-level immunity through infant vaccination span the entire high-fertility age groups. Note that this concurs with predictions made prior to the observations becoming available (Nokes and Anderson 1987).

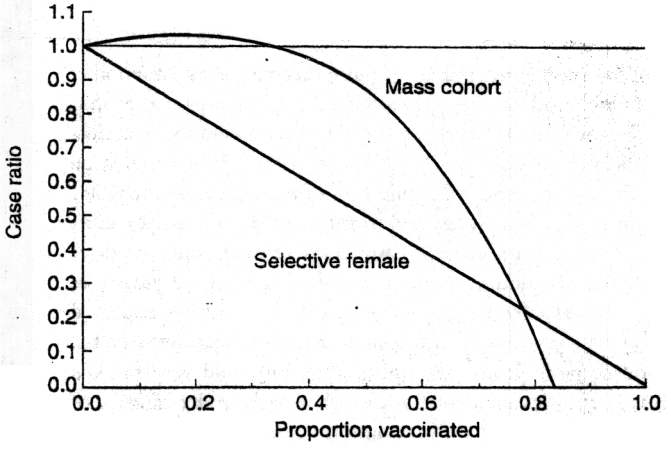


Fig. 22 Effectiveness of different rubella immunization programmes. Changes in the predicted case ratio (i.e. the average number of rubella infections in pregnant women after the introduction of immunization divided by average prevaccination number) under increasing levels of coverage for two types of policy, namely, selective immunization of girls of average age 12 years or mass vaccination of children (aged 2 years). Low to medium levels of uptake favour adoption of a selective immunization programme compared with mass vaccination which has the undesirable effect of increasing the average age at infection.

Table 5 Strategies of rubella immunization

	Selective	Mass cohort
Aim	Eliminate congenital rubella, not rubella infection	Eliminate rubella infection, and so congenital rubella
Age at vaccination	Prepubertal girls (10–15 years)	Boys and girls of 1–2 years
Philosophy	(i) Build upon levels of herd immunity attained through childhood (ii) Reduce the proportion of susceptible women of childbearing age (ii) Allow continued circulation of virus in male and young female segments of the population	(i) Reduce circulation of wild virus in community, especially children (ii) Lower the probability of susceptible women catching infection via the action of herd immunity
Overall incidence of infection	Very little impact at any level of coverage	(i) Reduction in cases in a non-linear manner as vaccine level increases (see Fig. 13) (ii) Increase in average age at infection
Other concerns	(i) Cannot eradicate congenital rubella unless 100 per cent of women 'at risk' are immune (via infection in childhhod or immunization) (ii) Herd immunity largely natural with continued re-exposure to infection and boosting of antibody response	(i) Proportion of remaining cases increases in older age classes, hence possible to increase congenital rubella at certain levels of immunization (ii) Herd immunity ultimately all vaccine induced. Less solid? No boosting of immunity by re-exposure to virus
Which policy?	Suitable for lower levels of vaccination coverage (see Fig. 22)	Suitable if high levels of uptake can be achieved (see Fig. 22)
Country (as example)	UK	USA

The strategy of pulse vaccination

The use of the alternative strategy of pulse vaccination as a method of control of childhood vaccine-preventable diseases has gained prominence in the early 1990s largely as a result of success in the Americas against polio and measles (De Quadros *et al.* 1991). Pulse vaccination may be defined as the repeated application of vaccine across a wide age range (Agur *et al.* 1993; Nokes and Swinton 1995) and usually takes the form of vaccination days or campaigns repeated once or twice yearly in which all children under a specified age (e.g. 15 years) are offered vaccine (usually irrespective of vaccination history). Repeated vaccination days or weeks in Central and South America have seen the elimination of polio from the region since 1991 and very marked reductions in measles incidence. Although a basic understanding of the rationale underlying pulse vaccination guided its use in the Latin American context, there is good reason to seek greater quantitative insight into the underlying mechanism of action prior to advocating more widespread use in other regions with different social patterns and health infrastructure.

Remember that it is the presence of a threshold density or proportion of susceptibles in a population which enables endemic persistence of acute vaccine-preventable infections (i.e. infections requiring close contact to effect transmission and which develop lasting immunity following recovery). Vaccination of a fraction of an endemic proportion susceptible lowers the effective reproductive rate below unity and incidence declines (this may be quite a considerable reduction if a pulse is administered across a wide age range). Lowering the number of infectious persons in the population results in a lowering of the force of infection acting upon susceptibles. In turn fewer infections leads to a build-up once more in susceptible numbers to the threshold level. The principle behind pulse vaccination rests upon these simple conditions. The aim of repeatedly pulsing is to maintain susceptible numbers below the threshold density or fraction and thereby maintain a continual decline in incidence (i.e. by maintaining $R < 1$). In practical quantitative terms we are interested in the timing of successive pulses to achieve this objective.

The interpulse interval depends upon three factors: what fraction of the population are susceptible at endemic equilibrium, how much of this susceptible population is immunized as a result of the campaign, and how rapidly are susceptibles replenished after a campaign. Translating these into epidemiological terms, we note that the proportion of the susceptible population vaccinated in a single pulse is $p'x^*$ where p' is the vaccination coverage and x^* is the endemic fraction susceptible (related to the basic reproductive rate in the form $x^* = 1/R_0$). In addition, if the total population is approximately constant in size, then, ignoring any further infection, the rate of replenishment of susceptibles by births is equal to the death rate, $\mu = 1/L$ (where L is life expectancy at birth). Therefore the minimum time taken after pulsing to recover the equilibrium fraction, i.e. the interpulse period T_v, is

$$T_v = p'x^*L \tag{39}$$

and, since $x^* = 1/R_0 = A/L$, we obtain

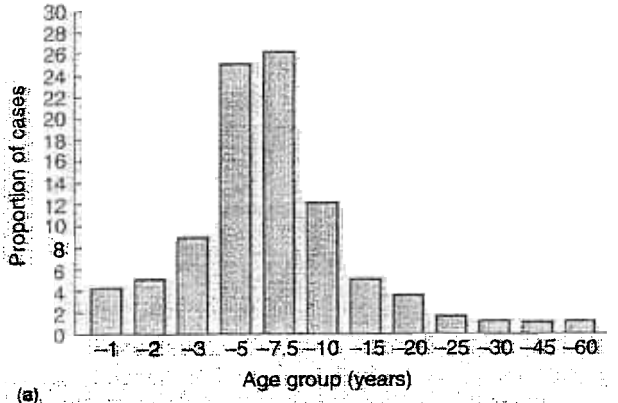
$$T_v = p'A \tag{40}$$

(Agur *et al.* 1993). This gives the common-sense result that if all susceptibles were to be immunized by a pulse of vaccine, i.e. $p' = 1.0$, then the time taken to recover the threshold fraction would be

Box 4 Epidemiology and control of mumps virus infection

Incidence of infection

Mumps is typical of the childhood viral infections with peak incidence in the young age classes and relatively few cases occurring in adulthood (figure (a)).

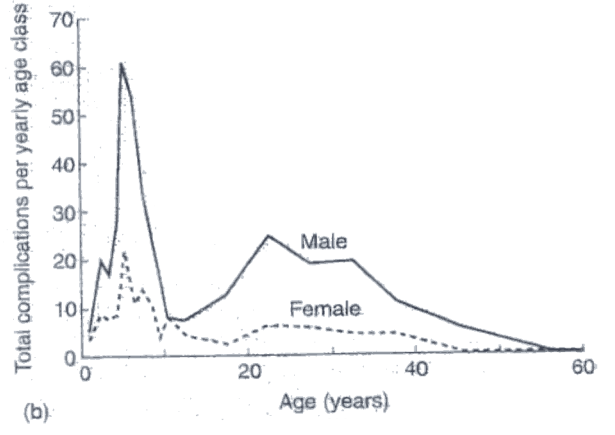


(a)

Information of this sort, obtained from age-specific case notification data or age-serological profiles, is used to derive age-dependent rates of transmission as shown in Fig. 5(b) in the main text.

Incidence of disease

Various types of complications are associated with mumps virus infection (figure (b)). In the prevaccination era mumps was the most common cause of viral meningitis in the United Kingdom, and is also a significant cause of encephalitis and, in postpubertal males, of orchitis.

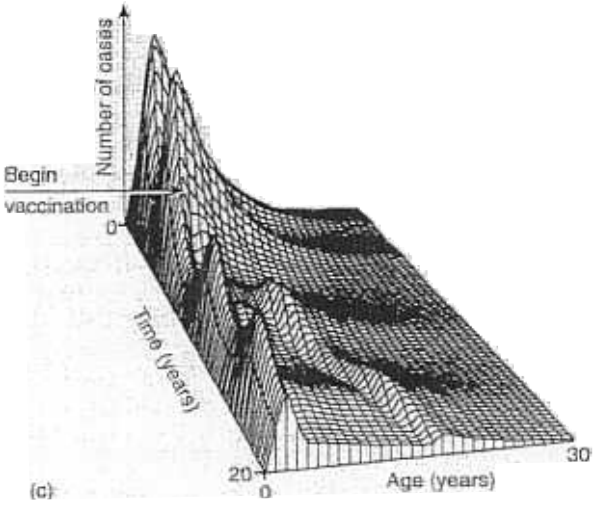


(b)

Scaling these age-complication data by the proportion of cases of infection in the corresponding age classes (shown above), it is possible to drive the relative risk of complications from infection, as shown in Fig. 21(b) in the main text. What becomes apparent from these data analyses is that, although fewest cases of infection occur in the older age classes, there remain substantial numbers of cases of complications, such that infection in older persons runs a considerably greater risk of resulting in complications when compared with infection in the young.

Mass vaccination and the incidence of infection

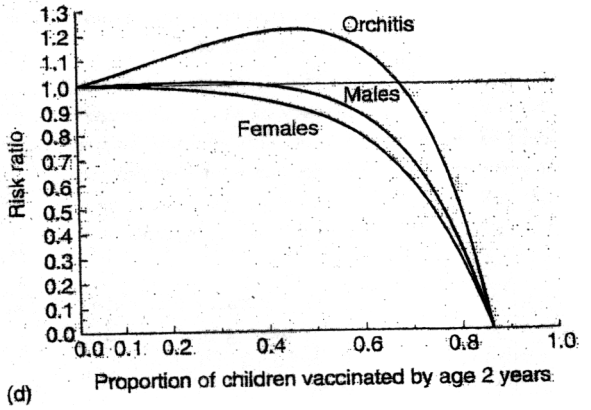
Figure (c) shows the predicted numbers of cases of mumps infection across a wide range of age classes, through time, before and after the introduction of a programme of mass cohort immunization (60 per cent of 2-year-olds). The force of infection is assumed to remain constant with age at 0.15 per year (corresponding to an average age at infection before immunization of 6.7 years). The epidemic peaks in the prevaccine period show the majority of cases occurring in the youngest age classes. Subsequent to the initiation of immunization two changes should be noted: (a) the obvious and expected decline in infection incidence (particularly in the young), and (b) an increase in the age at which the remaining cases occur, indicated by the wave of infections migrating, in time, into the older age classes. The implications of this shift in the age distribution of cases on the incidence of disease are addressed below.



(c)

Mass vaccination and the incidence of disease

The effect of a rise in the proportion of cases in the older age classes (predicted above) on the incidence of complications is dependent upon two things: (a) the level of cohort immunization that can be attained, and (b) the age-dependent nature of the risk of complications seen in Fig. 21. Simulations that help to unravel this problem are shown in figure (d) (adapted from Anderson *et al.* 1987), recording the change in the predicted risk ratio (i.e. the average number of complications after immunization has begun divided by the average number of complications occurring before) over various levels of childhood immunization (note that the risk ratio is unity for no benefit from immunization). Obviously there is little benefit to be obtained by vaccination at less than 60 per cent, and indeed vaccination at anything less than 70 per cent is potentially hazardous when considering orchitis alone. Such a phenomenon is a direct result of the combination of increased average age at infection and of the risk of complications with age.



(d)

equivalent to the average age at infection. The higher the average age at infection, that is the lower the transmission potential of the infection, the higher the endemic fraction susceptible and the consequent increase in the permitted interval between pulses. An observation from this analysis is that it is possible to eliminate an infection using a pulse vaccination proportion, p', which is less than the critical level of coverage predicted to be required for a continuous immunization process. The reason for this is that by vaccinating repeatedly across an age range, some individuals will receive multiple opportunities to receive the vaccine and there is therefore a build-up of vaccine-induced immunity with increasing age.

One complication that ought to be considered is the effect of combining a routine vaccination programme with a pulse regime (Nokes and Swinton 1995). Clearly, if a fraction of susceptibles are being vaccinated at or near to birth, then the rate of replenishment of susceptibles following a pulse will be lowered. It may be shown that, provided that the vast majority of individuals acquire immunity some time during their lives, the interpulse period in the presence of a routine vaccination programme in which a proportion p are vaccinated is

$$T_v = p'A/(1-p). \quad (41)$$

In other words the pulse interval is lengthened in direct relation to the new fraction of births which are susceptible, $(1-p)$.

Major simplifying assumptions underlie these simple relationships for the interpulse period. It is assumed that a proportion p' of susceptibles of any age are vaccinated and that on successive occasions each individual in a population has the same likelihood of being vaccinated. Such assumptions entail that the expressions given provide simple guidelines to aid understanding. Models of greater complexity are required which expand upon these ideas to give more practical guidelines (Nokes and Swinton 1995).

Monitoring the impact of control programmes

There is an ever-growing need to establish a co-ordinated surveillance programme to monitor the impact of control programmes against microparasitic infections (Nokes and Cutts 1993). The needs include the following.

1. To establish the impact of a specified control programme on a particular outcome variable, such as incidence of infection. This is of increasing importance as control programmes near their goals of elimination, where indicators of process, such as vaccination coverage, simply do not relate well enough to outcome.
2. To establish the accuracy of outcome indicators, such as notifications of infectious disease, the efficiency of which commonly fall off dramatically as incidence declines. This is crucial to the identification of outbreaks (and perhaps areas of low vaccine uptake) and to the validation of elimination targets, for example surveillance of acute flaccid paralysis as a marker for poliomyelitis.
3. To monitor the appearance of wild-type variants, either introduced from other countries, or which have gained selective advantage over persisting strains in the presence of high selective pressure of vaccination or chemotherapy.

Various modern tools are now at our disposal to assist in this process. For example, saliva antibody assays which are being used to confirm clinical diagnoses and may be useful in establishing longitudinal surveillance systems and molecular probes by which to identify the origins of strains in infection outbreaks and the arrival of variants able to circulate in the presence of high-level vaccine-associated immunity.

Mathematical models also have a role to play in this area of epidemiology. They facilitate the assessment of the impact of mass vaccination programmes through their predictive capability, where suitable outcome indicators may not be available (e.g. infections with poor differential diagnosis) or may only be measurable many years after a programme has begun (e.g. hepatitis B virus and the occurrence of hepatic disorders). In addition, models can be used to explore the potential (and the time course) for strain variants to establish themselves in highly vaccinated populations, where they would otherwise normally be out-competed by a dominant (higher R_0) strain, for which immunity through vaccination is more solid.

Changes in sexual behaviour and the transmission of sexually transmitted infections

The current pandemic of HIV and AIDS, and the absence of effective drugs and a vaccine to combat infection, has focused much attention in recent years on how to induce changes in sexual behaviour via education and media publicity campaigns to slow the spread of infection. The most important behaviour relevant to the rate of spread is the distribution of the rates of acquiring new sexual partners within a defined population (Fig. 8). A major characteristic of this behaviour is the heterogeneity between individuals within a given community. A central question in this problem is whether it is best to aim health educational programmes at the whole population, with the aim of reducing average rates of sexual-partner change or whether it is best to target education at high-risk groups such as those with very high rates of sexual-partner change (in either male homosexual or heterosexual communities). This is a complicated question and its resolution depends, in part, on a detailed quantitative knowledge of the pattern of sexual behaviour within a given population. However, simple mathematical models can help to provide some clues to the resolution of this issue. Of particular importance in understanding the dynamics of transmission of HIV is determining how sexual behaviour influences the magnitude of the basic reproductive rate, R_0. As discussed earlier, for a sexually transmitted disease such as HIV, the magnitude of R_0 is (in simple terms) defined by the probability of transmission per partner contact, B, multiplied by the effective rate of sexual-partner change, c, multiplied by the average duration of infectiousness, D, of an infected person. As noted earlier the variance in the rate of sexual-partner acquisition is typically much larger in value than the mean and, hence, those with high rates of partner change play a disproportionate role (relative to their proportional representation within a sexually active population) in the spread of infection (Fig. 14). This simple theoretical result suggests that greater benefit is to be gained (in terms of reducing R_0) by targeting education at those with higher than average rates of sexual-partner change. In practice the identification of such individuals is problematic in the absence of detailed survey data that relate this behaviour to other characteristics. The surveys of sexual behaviour that have been completed to date show a strong age dependency (with young adults having the highest rates of sexual-partner change) but little else of help in identifying correlates (ACSF 1992; Johnson *et al.* 1992). However, attendees at sexually transmitted disease clinics are an important target group, since sexually transmitted diseases other than HIV are more frequently present among those with high rates of partner change. Small

changes in behaviour in the highly sexually active are likely to have a major impact on the prevalence of sexually transmitted diseases in a community.

Conclusion

We have glossed over much detail and ignored many complications in model formulation and analysis in this chapter. The interested reader is therefore urged to consult the source references. Our aim has been to define, as simply as possible, the central concepts underpinning the study of the transmission dynamics of infectious diseases and the major conclusions that have emerged from the development and analysis of mathematical models of transmission and control.

The recent convergence of mathematical theory and observation in epidemiology has created a powerful set of tools for the study of the population biology of infectious disease agents. At present the potential value of these techniques is not widely appreciated by public health scientists and medical personnel. Many people have rightly criticized models that pursue the mathematics for its own sake, making only perfunctory attempts to relate the findings to epidemiological data. But there is a converse danger which is less widely understood. The complexities of the course of infection within an individual and its spread between people are such that years of clinical experience and the most refined intuition will not always yield reliable insights into the factors that control the transmission dynamics of a given infectious agent and how these are influenced by perturbations introduced by control measures. Moreover, insensitive use of a computer will not always help in understanding these problems, for if a computer is given inappropriate instruction it will usually give inappropriate answers. What is needed, in our view, is increased collaboration between epidemiologists and mathematicians, with the models being founded on data (and with their predictions being tested against available facts) and with verbal hypotheses being founded on clear mathematical statements of the assumptions. We hope that the contents of this chapter stimulate interest in this goal.

References

ACSF (1992). AIDS and sexual behaviour in France. *Nature*, **360**, 407–9.

Agur, Z., Cojocaru, L., Mazor, G., Anderson, R.M., and Danon, Y.L. (1993). Pulse mass measles vaccination across age cohorts. *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 11698–702.

Anderson, R.M. (1982). Directly transmitted viral and bacterial infections of man. In *Population dynamics of infectious diseases—theory and applications* (ed. R.M. Anderson), pp. 1–37. Chapman & Hall, London.

Anderson, R.M. (1988). Epidemiology of HIV infection: variable incubation plus infectious periods and heterogeneity in sexual activity. *Journal of the Royal Statistical Society, Series A*, **151**, 66–93.

Anderson, R.M. and Grenfell, B.T. (1986). Quantitative investigation of different vaccination policies for the control of congenital rubella syndrome (CRS) in the U.K. *Journal of Hygiene (Cambridge)*, **96**, 305–33.

Anderson, R.M. and May, R.M. (1982). Directly transmitted infectious diseases: control by vaccination. *Science*, **215**, 1053–60.

Anderson, R.M. and May, R.M. (1983). Vaccination against rubella and measles: quantitative investigations of different policies. *Journal of Hygiene (Cambridge)*, **90**, 259–325.

Anderson, R.M. and May, R.M. (1984). Spatial, temporal and genetic heteogeneity in host populations and the design of immunization programmes. *IMA Journal of Mathematics Applied in Medicine and Biology*, **1**, 233–66.

Anderson, R.M. and May, R.M. (1985*a*). Age-related changes in the rate of disease transmission: implications for the design of vaccination programmes. *Journal of Hygiene (Cambridge)*, **94**, 365–436.

Anderson, R.M. and May, R.M. (1985*b*). Herd immunity to helminth infection and implications for parasite control. *Nature*, **315**, 493–6.

Anderson, R.M. and May, R.M. (1985*c*). Vaccination and herd immunity to infectious diseases. *Nature*, **318**, 323–9.

Anderson, R.M. and May, R.M. (1986). The invasion, persistence and spread of infectious diseases within animal and plant communities. *Philosophical Transactions of the Royal Society of London*, **314**, 533–70.

Anderson, R.M. and May, R.M. (1988). Epidemiological parameters of HIV transmission. *Nature*, **333**, 514–22.

Anderson, R.M. and May, R.M. (1991). *Infectious diseases of humans: dynamics and control*. Oxford University Press.

Anderson, R.M., Grenfell, B.T., and May, R.M. (1984). Oscillatory fluctuations in the incidence of infectious disease and the impact of vaccination: time series analysis. *Journal of Hygiene (Cambridge)*, **93**, 587–608.

Anderson, R.M., Medley, G.F., May, R.M., and Johnson, A.M. (1986). A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS. *IMA Journal of Mathematics Applied in Medicine and Biology*, **3**, 229–63.

Anderson, R.M., Crombie, J.A., and Grenfell, B.T. (1987). The epidemiology of mumps in the UK: a preliminary study of virus transmission, herd immunity and the potential effect of immunization. *Epidemiology and Infection*, **99**, 65–84.

Anderson, R.M., May, R.M., and McLean, A.R. (1988). Possible demographic consequences of AIDS in developing countries. *Nature*, **332**, 228–34.

Babad, H.R., Nokes, D.J., Gay, N.J., Miller, E., Morgan-Capner, P., and Anderson, R.M. (1995). Predicting the impact of measles vaccination in England and Wales: model validation and analysis of policy options. *Epidemiology and Infection*, **114**, 319–44.

Bailey, N.T.J. (1973). Estimation of parameters from epidemic models. In *Mathematical theory of the dynamics of biological populations* (ed. M.S. Bartlett and R.W. Hiorns), p. 253. Academic Press, London.

Bailey, N.T.J. (1975). *The mathematical theory of infectious diseases and its implications*. Griffin, London.

Ball, F. (1983). The threshold behaviour of epidemic models. *Journal of Applied Probability*, **20**, 227–41.

Becker, N. (1979). The uses of epidemic models. *Biometrics*, **35**, 295–305.

Behets, F.M., Edidi, B., Quinn, T.C., *et al.* (1991). Detection of salivary HIV-1-specific IgG antibodies in high risk populations in Zaire. *Journal of Acquired Immune Deficiency Syndromes*, **4**, 183–7.

Bernoulli, D. (1760). Essai d'une nouvelle analyse de la mortalité causée pour la verole et des avantages de l'incubation pour la prevenir. *Memoires Mathematiques et Physiques de l'Academie Royale des Sciences (Paris)*, **1**, 1–45.

Black, F.L. (1966). Measles endemicity in insular populations: critical community size and its evolutionary implications. *Journal of Theoretical Biology*, **2**, 207–11.

Bolker, B.M. and Grenfell, B.T. (1993). Chaos and biological complexity in measles dynamics. *Philosophical Transactions of the Royal Society of London. B Biological Sciences*, **251**, 75–81.

Brown, D.W.G., Ramsay, M.E.B., Richards, A.F., and Miller, E. (1994). Salivary diagnosis of measles: a study of notified cases in the UK, 1991–3. *British Medical Journal*, **308**, 1015–17.

Brownlee, J. (1906). Statistical studies in immunity: the theory of an epidemic. *Proceedings of the Royal Society of Edinburgh*, **26**, 484–521.

Cliff, A., Haggett, P., and Smallman-Raynor, M. (1993). *Measles: an historical geography*. Blackwell Scientific, Oxford.

De Quadros, C.A., Andrus, J.K., Olive, J.-M., *et al.* (1991). Eradication of poliomyelitis: progress in the Americas. *Pediatric Infectious Disease Journal*, **10**, 222–9.

Dietz, K. (1987). Mathematical models for the control of malaria. In *Malaria* (ed. W.H. Wensdorfe and J.A. MacGregor), p. 1087. Churchill Livingstone, Edinburgh.

Farr, W. (1840). *Progress of epidemics. Second Report of the Registrar General of England and Wales*, pp. 91–8. HMSO, London.

Fine, P.E.M. (1993). Herd immunity: history, theory, practice. *Epidemiologic Reviews*, **15**, 265–302.

Garnett, G.P. and Anderson, R.M. (1993). Contact tracing and the estimation of sexual mixing patterns: the epidemiology of gonococcal infections. *Sexually Transmitted Diseases*, **20**, 181–91.

Garnett, G.P., Swinton, J., Brunham, R.C., and Anderson, R.M. (1992). Gonococcal infection, infertility and population growth: II. The influence of heterogeneity in sexual behaviour. *IMA Journal of Mathematics Applied in Medicine and Biology*, **9**, 127–44.

Grenfell, B.T. and Anderson, R.M. (1985). The estimation of age-related rates of infection from case notifications and serological data. *Journal of Hygiene*, **95**, 419–36.

Gupta, S. and Anderson, R.M. (1992). Sex, AIDS and mathematics. *New Scientist*, 12 September, pp. 34–8.

Hamer, W.H. (1906). Epidemic disease in England. *Lancet*, **i**, 733–9.

Hethcote, H.W. and Yorke, J.A. (1984). Gonorrhoea: transmission dynamics and control. *Lecture Notes in Biomathematics*, **56**, 1–105.

Holmstrom, P., Syrjanen, S., Laine, P., Valle, S.-L., and Suni, J. (1990). HIV antibodies in whole saliva detected by ELISA and Western blot assays. *Journal of Medical Virology*, **30**, 245–8.

Hope Simpson, R.E. (1952). Infectiousness of communicable diseases in the household. *Lancet*, **i**, 1145–55.

Johnson, A.M., Wadsworth, J., Wellings, K., Bradshaw, S., and Field, J. (1992). Sexual lifestyles and HIV risk. *Nature*, **360**, 410–12.

Johnson, A.M., Wadsworth, J., Wellings, K., Field, J., and Bradshaw, S. (1994). *Sexual attitudes and lifestyles*. Blackwell Scientific, Oxford.

Kermack, W.O. and McKendrick, A.G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London, Series A*, **115**, 700–21.

Knox, E.G. (1980). Strategy for rubella vaccination. *International Journal of Epidemiology*, **9**, 13–23.

McLean, A.R. and Anderson, R.M. (1988*a*). Measles in developing countries. Part II. The predicted impact of mass vaccination. *Epidemiology and Infection*, **100**, 419–42.

McLean, A.R. and Anderson, R.M. (1988*b*). Measles in developing countries. Part I. Epidemiological parameters and patterns. *Epidemiology and Infection*, **100**, 111–33.

McLean, A.R. and Blower, S. (1993). Impefect vaccines and herd immunity to HIV. *Proceedings of the Royal Society London B*, **253**, 9–13.

Markowitz, L.E., Albrecht, P., Rhodes, P., *et al.* (1996). Changing levels of measles antibody titers in women and children in the United States: impact on response to vaccination. *Pediatrics*, **97**, 53–8.

May, R.M. and Anderson, R.M. (1984). Spatial heterogeneity and the design of immunization programs. *Mathematical Biosciences*, **72**, 83–111.

May, R.M. and Anderson, R.M. (1987). The transmission dynamics of HIV infection. *Nature*, **326**, 137–42.

Mortimer, P.P. and Parry, J.V. (1991). Non-invasive virological diagnosis: are saliva and urine specimens adequate substitutes for blood? *Reviews in Medical Virology*, **1**, 73–8.

Nokes, D.J. and Anderson, R.M. (1987). Rubella vaccination policy: a note of caution. *Lancet*, **i**, 1441–2.

Nokes, D.J. and Anderson, R.M. (1988). The use of mathematical models in the epidemiological study of infectious diseases and in the design of mass immunization programmes. *Epidemiology and Infection*, **101**, 1–20.

Nokes, D.J. and Anderson, R.M. (1991). Vaccine safety versus vaccine efficacy in mass immunization programmes. *Lancet*, **338**, 1309–12.

Nokes, D.J. and Anderson, R.M. (1992). Mathematical models of infectious agent transmission and the impact of mass vaccination. *Reviews in Medical Microbiology*, **3**, 187–95.

Nokes, D.J. and Anderson, R.M. (1993). Application of mathematical models to the design of immunization strategies. *Reviews in Medical Microbiology*, **4**, 1–7.

Nokes, D.J. and Cutts, F.T. (1993). Immunizations in the developing world: strategic challenges. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **87**, 353–4, 398.

Nokes, D.J. and Swinton, J. (1995). The control of childhood infection by pulse vaccination: an epidemiological approach. *IMA Journal of Mathematics Applied in Medicine and Biology*, **12**, 29–53.

Nokes, D.J., Anderson, R.M. and Anderson, M.J. (1986). Rubella epidemiology in south-east England. *Journal of Hygiene (Cambridge)*, **96**, 291–304.

Parry, J.V., Perry, K.R., Panday, S., and Mortimer, P.P. (1989). Diagnosis of hepatitis A and B by testing saliva. *Journal of Medical Virology*, **28**, 255–60.

Perry, K.R., Brown, D.W.G., Parry, J.V., Panday, S., Pipkin, C., and Richards, A. (1993). Detection of measles, mumps and rubella antibodies in saliva using antibody capture radioimmunoassay. *Journal of Medical Virology*, **40**, 235–40.

Rogers, D.J. (1988). A general model for the African trypanosomiases. *Parasitology*, **97**, 193.

Ross, R. (1911). *The prevention of malaria* (2nd edn). Murray, London.

Scott, M.E. and Smith, G. (ed.) (1994). *Parasitic and infectious diseases epidemiology and ecology* (1st edn). Academic Press, London.

Sokal, R.R. and Rohlf, F.J. (1981). *Biometry* (2nd edn). W.H. Freeman, San Francisco, CA.

Soper, M.A. (1929). Interpretation of periodicity in disease prevalence. *Journal of the Royal Statistical Society A*, **92**, 34–61.

Ukkonen, P. and Von Bonsdorf, C.-H. (1988). Rubella immunity and morbidity: effects of vaccination in Finland. *Scandinavian Journal of Infectious Diseases*, **20**, 255–9.

Van Den Akker, R., Van Den Hoek, J.A.R., Van Den Akker, *et al.* (1992). Detection of HIV antibodies in saliva as a tool for epidemiological studies. *AIDS*, **6**, 953–7.

Wharton, M., Cochi, S.L., Hutcheson, R.H., Bistowish, J.M., and Shaffner, W. (1988). A large outbreak of mumps in the post vaccine era. *Journal of Infectious Diseases*, **158**, 1253–60.

Wickwire, K. (1977). Mathematical models for the control of pests and infectious diseases: a survey. *Theoretical Population Biology*, **11**, 182–238.

Yorke, J.A., Nathanson, N., Pianigiani, G., and Martin, J. (1979). Seasonality and the requirements for perpetuation and eradication of viruses in populations. *American Journal of Epidemiology*, **109**, 103–23.