1 Directly transmitted viral and bacterial infections of man

Roy M. Anderson

1.1 INTRODUCTION

The origins of contemporary epidemiological theory can be traced back to the early part of the 20th century. Just prior to this period the mechanisms by which infectious disease agents spread within populations had been revealed by microbiological research, notably that of Pasteur and Koch, and this, together with a statistical familiarity with epidemiological data (particularly the geometry of the epidemic curve (Farr, 1840; Brownlee, 1906)), laid the groundwork for future developments.

The first major theoretical contribution was that of Hamer (1906) who postulated that the course of an epidemic depends on the contact rate between susceptible and infectious individuals. This notion has become one of the most important concepts in mathematical epidemiology; it is the so-called 'mass action principle' in which the rate of disease spread is assumed to be proportional to the product of the density of susceptibles times the density of infectious individuals. This simple assumption is central to most deterministic and stochastic theories of disease dynamics.

Hammer formulated the mass action principle in a discrete time model but in 1908 Ronald Ross (celebrated as the discoverer of malarial transmission by mosquitoes) translated the problem into a continuous-time model in his pioneering work on the dynamics of malaria (Ross, 1911, 1915, 1916, 1917). The structured form of Ross's model meant that, for the first time, it was possible to use a clearly defined mathematical theory as a genuine research tool in epidemiology.

The ideas of Hamer and Ross were extended, and explored in more detail, by Soper (1929), who deduced the underlying mechanisms responsible for the often observed periodicity of epidemic outbreaks of disease, and by Kermack and McKendrick (1927) who established the celebrated threshold theorem. This theorem, according to which the introduction of infectious individuals into a community of susceptibles will not give rise to an epidemic outbreak unless the density of susceptibles is above a certain critical value, is, in conjunction with the mass action principle, a cornerstone of modern theoretical epidemiology.

Since this early beginning the growth in the literature concerned with

mathematical epidemiology has been very rapid indeed. Recent reviews of this literature have been published by Bailey (1975) and Becker (1979). From an early stage it became apparent that the elements of chance and variation were important determinants of disease spread and this led to the development of stochastic theories. Much of the literature over the past three decades has been concerned with probabilistic models (see for example Bartlett, 1955, 1960; Bailey, 1975).

In recent work there has been an emphasis on the application of control theory to epidemic models (Wickwire, 1977), the spatial spread of diseases (Mollison, 1977) and the extension of the threshold theorem to encompass more complex deterministic and stochastic models (Whittle, 1955; Becker, 1977a). Despite the current sophistication of the mathematical literature dealing with epidemic phenomena, the insights gained from theoretical work have in general had little impact on public health policy. This may be due, in part, to the abstractly mathematical nature of much of this research. Becker (1979), for example, notes that of 75 papers on epidemiological models published since 1974, only five contained any data. If theoretical work is to make a contribution to the solution of practical problems in disease control, there clearly is a need for more data-oriented studies.

Some progress in this direction has been made, particularly with respect to the estimation from epidemiological data of parameters such as the incubation period and the duration of infectiousness (see Bailey, 1975, for a review of this work). More recently, attention has begun to be focused on the estimation of the rate of disease reproduction within human communities (the basic reproductive rate of infection, R) and on the use of this measure to determine immunization levels for disease control or eradication (MacDonald, 1957; Dietz, 1974, 1976; Yorke, Hethcote and Nold, 1978; Yorke *et al.*, 1979; Anderson and May, 1982a, b).

This chapter attempts to summarize the main trends in the more data-oriented sections of the literature. The emphasis is placed on recurrent epidemic behaviour and on the manner in which simple models can provide broad insights into the factors controlling the persistence and stability of directly transmitted viral and bacterial infections within large human communities (large-scale epidemic phenomena, Anderson and May, 1979a; May and Anderson, 1979). Of all disease agents, these directly transmitted infections have attracted the greatest attention from theoreticians. This is due, in part, to the relative simplicity of their life cycles but is also a consequence of the availability of long-term records of the population behaviour of such infections in Britain and North America.

1.2 HISTORICAL PERSPECTIVE

The observed improvement in human mortality rates within Europe and North America over the past three centuries, with life expectancy increasing

from around 25–30 years in 1700 to around 70–75 years in 1970, comes mainly from a decline in deaths from directly transmitted infectious diseases. A combination of nature and nurture is implicated: higher standards of hygiene and nutrition have combined with probable changes in the genetic structure of human and parasite populations to decrease the pathogenicity of many common childhood infections (McNeill, 1976; McKeown, 1979). An illustration of this trend is displayed in Fig. 1.1 where the number of deaths in England and Wales attributed to measles is recorded for the period 1897 to 1939.

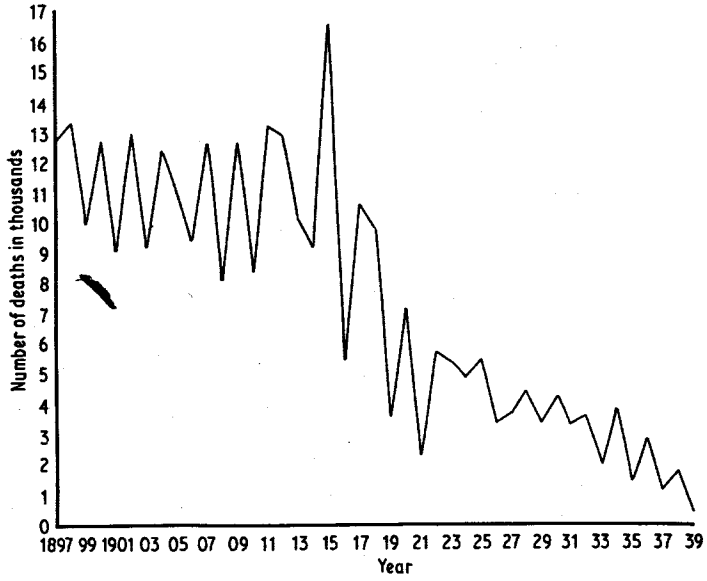

Figure 1.1 Yearly number of deaths attributed to measles in England and Wales over the period 1897 to 1939. (based on information from the Registrar General's statistical review of England and Wales).

In parallel with this decline in mortality, the frequency and magnitude of disease epidemics increased during the 18th and 19th centuries as a result of changing social patterns and the growth of large centres of population in increasingly industrialized societies. The reversal of this trend during the present century is largely due to the development and widespread use of vaccines to immunize susceptible populations against a variety of directly transmitted viral and bacterial diseases. The world-wide eradication of smallpox and the decline in the incidence of diptheria and paralytic

poliomyelitis in Europe are testimony to the effectiveness of this method of disease control.

In many regions of the world, however, many viral and bacterial infections remain endemic, despite the widespread use of vaccines. On the continents of Africa and Asia certain common childhood infections such as measles and whooping cough still remain a significant threat to life. The effective control of these infections will, in part, be dependent on an improved understanding of the population biology of the disease agents. Mathematical studies, combined with detailed statistical analysis of epidemiological data, can play an important role in the design of optimal control policies for these infections.

1.3 POPULATION DYNAMICS

In contrast to the extensive and sophisticated literature concerned with the analysis of simple epidemic phenomena, relatively less attention has been devoted to recurrent epidemic behaviour. Major advances have been made, however, by Bartlett (1956) and more recently by Dietz (1974, 1976) and Yorke and co-workers (Yorke and London, 1973; London and Yorke, 1973; Yorke *et al.*, 1979). In this section, first the main themes of this work are summarized, then the theoretical framework is extended to incorporate certain refinements, and finally existing data are analysed in the light of theoretical predictions.

The theoretical framework most commonly used to mimic the dynamics of viral and bacterial infections is one based on the division of the human population into categories containing susceptibles, infecteds who are not yet infectious (latent), infectious individuals, and those who are recovered and immune. The number (or density) of individuals in each of these categories will be denoted by the variables X, H, Y and Z respectively where the total population size N is $N = X + H + Y + Z$. Models based on this type of framework are compartmental in structure and do not explicitly describe changes in parasite population size. They simply mirror the dynamics of the number of infected people without reference to the abundance of organisms within each individual. Broadly speaking, they seek to answer such questions as: can the infection be stably maintained within the population? Is the disease endemic or epidemic in character? Does the infection exhibit recurrent epidemic behaviour? How do the proportions of susceptibles, infecteds and immunes change through time after the infection is introduced into a susceptible population? What is the critical density of susceptibles necessary to maintain the infection?

It is conventional to assume that the size (or density) of the human population, N, remains roughly constant, or at least changes on a time scale that is long compared to all other time scales of interest in an epidemiological context. This assumption is reasonable for most populations in western societies. The assumption corresponds to the net input of susceptibles into the

population (by births) being roughly equal to the net mortality μN (where μ is the per capita death rate and $1/\mu$ denotes life expectancy).

In accord with the 'mass action' transmission principle of Hamer (1906), we initially assume that the net rate at which infections are acquired is proportional to the number of encounters between susceptible and infectious individuals, βxy, where β is a transmission coefficient. Individuals pass from the latent state to the infectious stage at a per capita rate σ (such that the average latent period is $1/\sigma$), and recover to join the immune class at a per capita rate γ (where $1/\gamma$ represents the average infectious period).

Acquired immunity is taken to be lifelong, as it appears to be for most common childhood viral and bacterial diseases. This assumption, however, is easily modified in the model defined below. The assumption that all the rate parameters, β, σ, γ and μ, are simple constants is clearly artificial, but the resulting model provides a convenient point of departure for subsequent elaborations.

Under the above assumptions, a set of four first-order differential equations describe the dynamics of the infection within the human population (Dietz, 1974, 1976; Anderson and May, 1982a) and may be expressed as follows:

$$dX/dt = \mu N - \mu X - \beta XY \tag{1.1}$$

$$dH/dt = \beta XY - (\mu + \sigma)H \tag{1.2}$$

$$dY/dt = \sigma H - (\mu + \gamma)Y \tag{1.3}$$

$$dZ/dt = \gamma Y - \mu Z \tag{1.4}$$

Adding all four equations gives $dN/dt = 0$, corresponding to the original assumption that human population size, N, is constant.

The model defined by Equations (1.1) to (1.4) has two broad patterns of behaviour. The disease will be maintained within the population provided the 'basic reproductive rate', R, of the infection is greater than, or equal to, unity. The quantity R is of central importance to the dynamics of infectious disease agents and represents the expected number of secondary cases produced by an infectious individual in a defined population of X susceptibles (Dietz, 1974; Anderson and May, 1980). For the system defined above,

$$R = \frac{\sigma \beta X}{(\sigma + \mu)(\gamma + \mu)} \tag{1.5}$$

In more biological terms, secondary infections are produced at a rate βX throughout the expected lifespan, $1/(\gamma + \mu)$ of an infectious individual. Of these a fraction $\sigma/(\sigma + \mu)$ will survive the latent period to become the second generation of infectious individuals. Note the similarity of R to Fisher's net reproductive rate, R_0, a quantity widely used in the disciplines of ecology, population genetics and demography (Fisher, 1930). Also note that, although called a rate, R is in reality a dimensionless quantity. If $R < 1$, the disease cannot establish within the host population.

The criterion $R > 1$ for the establishment of the disease can equivalently be expressed as the requirement that the population of susceptibles exceed a critical 'threshold density', $X > N_T$, with the definition,

$$N_T = (\gamma + \mu)(\sigma + \mu)/\beta\sigma \tag{1.6}$$

This is the celebrated threshold theorem of Kermack and McKendrick (1927). More generally, therefore we may express R as

$$R = X/N_T \tag{1.7}$$

namely, the ratio of the number of susceptibles in the population divided by the threshold density necessary for disease persistence. In the absence of a continual inflow of susceptibles (when $\mu N = 0$), the criterion $R > 1$ represents the condition which must be satisfied for an epidemic to occur. Under such circumstances the disease will eventually die out once the supply of susceptibles drops below N_T.

For most of the common childhood viral and bacterial diseases, such as measles, whooping cough and chicken pox, the durations of the latent and infectious periods, $1/\sigma$ and $1/\gamma$, are of the order of a few days to a few weeks, while $1/\mu$ is of the order of 70 years or more. Under these circumstances ($\sigma \gg \mu$ and $\gamma \gg \mu$), Equations (1.5) and (1.6) may be accurately approximated as $R = \beta X/\gamma$ and $N_T = \gamma/\beta$.

Of the parameters determining the value of R, some are specific to the disease agents; examples are the parameters σ and γ, and that component of β which reflects the transmissibility of the disease. Other components of R, such as the density of susceptibles, X, and that component of β which reflects the average frequency of contact between individuals, vary greatly from one locality to the next depending on the prevailing environmental and social conditions. Even the value of $1/\gamma$ may be influenced by such conditions, since the isolation of infected children can substantially reduce the effective infectious period. The density of susceptibles depends mainly on the net birth rate in the community, which itself depends on the total population density, N. This observation underlies the observed correlation between endemic maintenance of disease without periodic fade out and community size. For measles in Britain and North America, the critical community size appears to be around 200 000–300 000 people (Table 1.1a) (Bartlett, 1957, 1960; Yorke *et al.*, 1979), although from an analysis of epidemiological data for relatively isolated island communities Black (1966) suggests a figure in the region of 500 000 (Table 1.1b). On a more local scale, in low-density rural communities (where $X < N_T$) epidemics will be unable to develop and the disease will not persist in the absence of a continual inflow of infecteds.

The model defined by Equations (1.1) to (1.4) predicts that provided $R > 1$, the system will exhibit damped oscillations to a stable state (Dietz, 1974). At equilibrium the proportions of susceptibles, infected but not infectious,

infectious and immune individuals, namely, x^*, h^*, y^* and z^* respectively are given by:

$$x^* = 1/R \tag{1.8}$$

$$h^* = (1 - 1/R)(\gamma + \mu)/(\sigma C) \tag{1.9}$$

$$y^* = (1 - 1/R)/C \tag{1.10}$$

$$z^* = \gamma(1 - 1/R)/(\mu C) \tag{1.11}$$

where R is as defined in Equation (1.5) and

$$C = 1 + \gamma/\mu + (\gamma + \mu)/\sigma$$

The damped oscillatory behaviour of the model is clearly at odds with observed patterns of disease behaviour such as the regular 2-year cycle in measles epidemics in Britain and North America. This issue will be discussed more fully in a later section of this chapter.

Table 1.1a Reported cases of measles in cities of North America, 1921–1940 (adapted from Bartlett (1960) and Yorke et al. (1979)).

City	*Population size (units of 100 000)*	*Years with a month in which no cases were reported*
New York	75	0
Chicago	34	0
Philadelphia	19	0
Detroit	16	0
Los Angeles	15	0
Montreal	10	0
Cleveland	9	1
Baltimore	9	0
Boston	8	0
Toronto	7	0
Washington	7	0
Pittsburgh	7	0
Milwaukee	6	0
Buffalo	6	0
Minneapolis	5	0
Vancouver	3	20
Rochester	3	3
Dallas	3	18
Akron	2	8
Winnipeg	2	7

Table 1.1b The persistence of measles within 19 island communities (adapted from Black (1966)).

Island	*Population size (units of 100 000)*	*Percentage of months in which no cases were reported*
Hawaii	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

1.4 PARAMETER ESTIMATION

The values of certain parameters of the model, such as the expectation of life, $1/\mu$, and the annual input of susceptibles, μN (the annual birth rate), can be obtained from published demographic statistics. Other parameters such as the average latent and infectious periods are more difficult to estimate and clearly vary for each disease agent.

To estimate these rate parameters it is necessary to turn to data reflecting transmission within family groups. The classic data on measles, collected by Hope Simpson in the Cirencester area of England during the years 1946–1952 (see Bailey, 1975), record the distribution of the observed time interval between two cases of measles in 219 families with two children under the age of 15 as listed in Table 1.2. The bulk of these observations represent case-to-case transmission 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. Early work on the estimation of latent periods was based on chain binomial models where it is assumed there is a constant incubation period which is terminated by a very short interval of high infectiousness since,

Table 1.2 Observed time interval distribution between two cases of measles in families of two (adapted from Bailey (1973); Hope Simpson's data from Cirencester, England 1946–1952)).

Time interval between the two cases (days)	*Total number of families observed*	*Presumed double primaries*	*Presumed case-to-case transmission*
0	5	5	.
1	13	13	.
2	5	5	.
3	4	4	.
4	3	2	1
5	2		2
6	4		4
7	11		11
8	5		5
9	25		25
10	37		37
11	38		38
12	26		26
13	12		12
14	15		15
15	6		6
16	3		3
17	1		1
18	3		3
19	.		.
20	.		.
21	1		1

once symptoms appear, the case is promptly removed from circulation. Susceptibles in contact with the case at the time of infectiouness have a certain probability of themselves becoming infected and this leads to successive crops of cases in a group of susceptibles (i.e. a family or school), the crops being separated in time by the incubation period (the time to appearance of symptoms). The variable number of cases actually observed at any given stage can be shown to have a binomial distribution (Wilson and Burke, 1942; Greenwood, 1931). Recently, more sophisticated models have been developed which assume that the latent period is variable, with say a normal distribution, and that it is followed by an extended but constant period of infectiousness (see Bailey, 1975, chapters 14 and 15; Becker, 1976, 1977b; and Gough, 1977). A rough guide to the latent, incubation and infectious periods of certain common viral and bacterial infections is recorded in Table 1.3. Some of these estimates are based on the detailed statistical analyses of household data while others are more speculative.

Table 1.3 Epidemiological parameters (information compiled from Fenner and White (1970); Christie (1974) and Benenson (1975)).

Infectious disease	*Latent period, $1/\sigma$ (days)*	*Infectious period, $1/\gamma$ (days)*	*Incubation period (time to appearance of symptoms; days)*
Measles	6–9	6–7	11–14
Chicken pox	8–12	10–11	13–17
Rubella	7–14	11–12	16–20
Infectious hepatitis	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 direct estimate of the basic reproductive rate, R, from Equation (1.5) (or equivalently, of N_T from Equation (1.6)) is usually impossible, because of the difficulties inherent in obtaining any direct estimate of the transmission parameter, β. Dietz (1974, 1976) has, however, shown that R can be estimated from the relation

$$R = 1 + L/A \tag{1.12}$$

Here L is the human life expectancy ($L = 1/\mu$), and A is the average age at which individuals acquire the infection. Dietz's derivation assumes all the rate parameters ($\sigma, \gamma, \mu, \beta$) are constants, independent of the age of the host. A more general expression for R has been derived by Anderson and May (1982a) and is discussed below, but Equation (1.12) remains a useful approximation even when the rate processes are age-dependent.

The average age at infection, A, can be estimated from data recording the proportion in each age class who have experienced the infection. The most accurate data of this form are provided by serological surveys but other, less accurate methods, are often employed, such as questionnaire surveys, or estimates can be obtained from age-classified notification records (Muench, 1959; Griffiths, 1974; Dietz, 1974). If the transmission parameter, β, of Equations (1.1) to (1.4) is age-independent, then $A = 1/\lambda$ where λ is the 'force of infection' of simple catalytic models (Muench, 1959). Under equilibrium conditions for example, in the context of Equations (1.1) to (1.4), $\lambda = \beta y^* N$. Typical data recording the proportion in each age class who have experienced the infection are displayed in Fig. 1.2. A more extensive collection of estimates of A for various diseases in various localities and times are presented in Table 1.4. Knowledge of A and L leads, via Equation (1.12), to an estimate of R as also catalogued in Table 1.4.

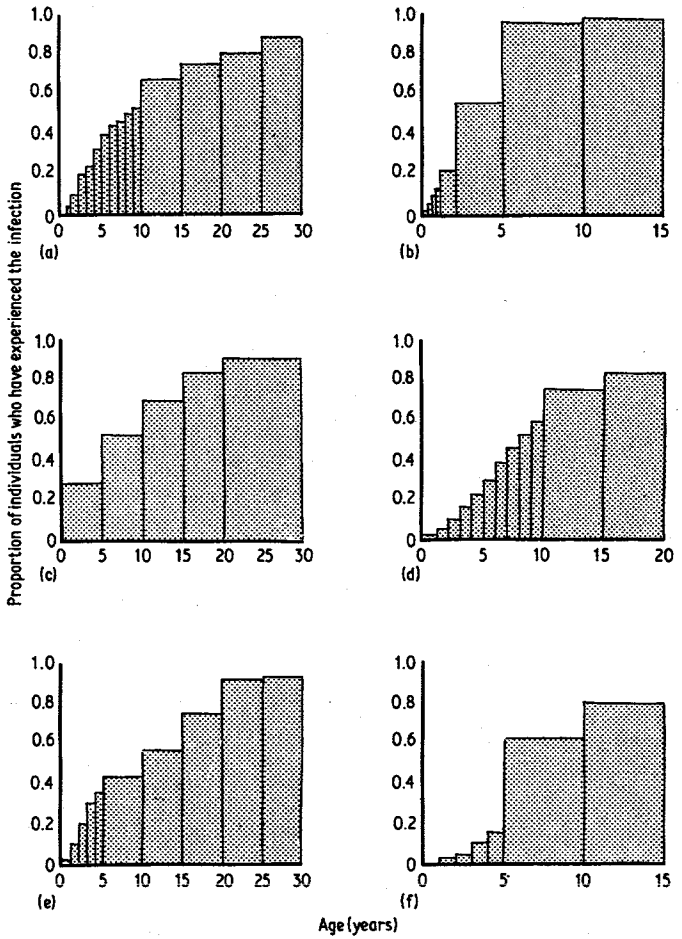

Figure 1.2 Examples of age-prevalence curves, based on either serological surveys or case notifications. (a) Poliomyelitis in the USA 1955 (serology) (b) Pertussis in England and Wales 1970 (case notifications) (c) Rubella in West Germany 1977 (serology) (d) Diphtheria in New York State 1915–1924 (case notifications) (e) Epstein–Barr virus in England and Wales 1970 (serology) (f) Mumps in Baltimore USA 1943 (case notifications).

Most commonly the transmission parameter, β, and hence the force of infection λ, are age-dependent; this complication will be treated in a later section (Anderson and May, 1982a; Griffiths, 1974).

Table 1.4 The average age, A, at which various infections are acquired (see Anderson and May (1982a) for data sources).

Infectious disease	*Average age at infection, A (years)*	*Geographical location*	*Type of community (r = rural, u = conurbation)*	*Time period*	*Assumed life expectancy (years)*	*R*
Measles	11.7	Kansas, USA	r	1918–1921	60	5.4
	10.5	Cattaraugus, New York, USA	r	1920–1930	60	6.0
	10.0	Maryland, USA	r	1907–1917	60	6.3
	7.7	Kansas, USA	u	1918–1921	60	8.3
	7.3	Massachusetts, USA	r and u	1918–1921	60	8.8
	7.3	Connecticut, USA	r and u	1921–1922	60	8.8
	6.9	Maryland, USA	u	1907–1917	60	9.4
	6.7	New Jersey, USA	r and u	1918–1921	60	9.7
	6.7	Massachusetts, USA	r and u	1932–1937	60	9.7
	6.1	England and Wales	r	1956–1968	70	12.5
	5.9	London, Ontario, Canada	u	1912–1913	60	11.1
	5.8	Providence, RI, USA	u	1919–1935	60	11.3
	5.6	Willesden, England	u	1912–1913	60	11.7
	5.5	Cirencester, England	u	1947–1950	70	14.0
	5.4	Baltimore, Maryland, USA	u	1916–1927	60	12.2
	5.3	Various localities in North America	r and u	1912–1928	60	12.5
	5.1	Hagerstown, Maryland, USA	u	1921–1923	60	13.0
	4.8	England and Wales	u	1956–1969	70	16.3
	4.4–5.6	England and Wales	r and u	1944–1979	70	13.7–18.0
	4.2	Gary, Indiana, USA	u	1920–1922	60	16.2
	3.5	Zambia, Rhodesia and S. Africa	r and u	1960–1968	40	11.4
	2.9	Ghana	r and u	1960–1968	40	13.8
	2.5	Eastern Nigeria	r and u	1960–1968	40	16.0
Whooping cough	9.4	Cattaraugus, New York, USA	r	1920–1930	60	6.4
	6.5	Various localities in North America	r and u	1912–1928	60	9.2
	6.5	Maryland, USA	r	1908–1917	60	9.2
	5.9	Hagerstown, Maryland, USA	u	1921–1923	60	10.2
	5.7	London, Ontario, Canada	u	1912–1913	60	10.6
	5.6	Connecticut, USA	r and u	1921–1922	60	10.7
	5.4	New Jersey, USA	r and u	1918–1921	60	11.1
	5.3	Massachusetts, USA	r and u	1918–1921	60	11.3
	4.9	Maryland, USA	u	1908–1917	60	12.2
	4.3	Baltimore, Maryland, USA	u	1943	70	16.23
	4.1–4.9	England and Wales	r and u	1944–1978	70	14.3–17.1
Chicken Pox	8.6	Maryland, USA	u	1913–1917	60	7.0
	8.5	Various localities in North America	r and u	1912–1928	60	7.1
	7.6	New Jersey, USA	r and u	1917–1921	60	7.9
	7.1	Massachusetts, USA	r and u	1918–1921	60	8.5
	6.8	Baltimore, Maryland, USA	u	1943	70	10.2
	6.7	Maryland, USA	u	1913–1917	60	9.0
Diphtheria	19.1	Pennsylvania, USA	r	1910–1916	60	3.1
	14.2	New York, USA	r	1918–1919	60	4.2
	12.8	Kansas, USA	r	1918–1921	60	4.7
	12.6	Maryland, USA	r	1908–1917	60	4.8
	11.6	Kansas, USA	u	1918–1921	60	5.2

Table 1.4 (*continued*)

Infectious disease	*Average age at infection, A (years)*	*Geographical location*	*Type of community (r = rural, u = conurbation)*	*Time period*	*Assumed life expectancy (years)*	*R*
	11.2	New York, USA	u	1918–1919	60	5.4
	11.0	Virginia and New York, USA	r and u	1934–1947	70	6.4
	10.4	Various localities in North America	r and u	1912–1928	60	5.8
Scarlet fever	14.9	Various localities in North America	r and u	1912–1928	60	4.0
	12.3	New York, USA	r	1918–1919	60	4.9
	10.8	Kansas, USA	r	1918–1921	60	5.5
	10.1	Maryland, USA	r	1908–1917	60	5.9
	10.0	Kansas, USA	u	1918–1921	60	6.0
	9.8	Pennsylvania, USA	r	1910–1916	60	6.1
	9.0	Pennsylvania, USA	u	1910–1916	60	6.7
	8.0	Maryland, USA	u	1908–1917	60	7.5
Mumps	13.9	Various localities in North America	r and u	1912–1916	60	4.3
	9.9	Baltimore, Maryland, USA	u	1943	70	7.1
Rubella	11.6	England and Wales	r and u	1979	70	6.0
	10.5	West Germany	r and u	1972	70	6.7
Poliomyelitis	17.9	USA	r and u	1955	70	5.9
	11.2	Netherlands	r and u	1960	70	6.2

1.5 THE INTER-EPIDEMIC PERIOD

Long-term records reveal that many common childhood diseases exhibit marked variations in incidence from year to year. These fluctuations are often of a regular nature, tending to arise as a broad consequence of the depletion and renewal of the supply of susceptibles. The 2–3-year cycles of measles, a typical example of which is shown in Fig. 1.3, is particularly remarkable. In general, the interval between major epidemics is termed the inter-epidemic period and values of this parameter, for a variety of childhood diseases, are recorded in Table 1.5.

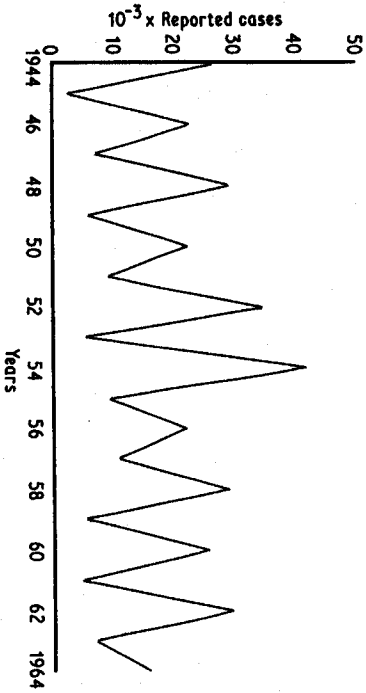

Figure 1.3 Measles in New York City from 1944 to 1964 (data from Yorke and London, 1973).

The deterministic model of Equations (1.1)–(1.4) exhibits damped oscillations, where, for diseases which are of short duration relative to the host lifespan ($\sigma \gg \mu, \gamma \gg \mu$), the period, T, of the oscillations is approximately (Anderson and May, 1982a):

$$T = 2\pi[LD/(R-1)]^{1/2} = 2\pi(AD)^{1/2} \tag{1.13}$$

Here D is the sum of the lengths of the latent and infectious periods (i.e. $D = 1/\sigma + 1/\gamma$), and R, L and A are as defined previously. The tendency of these oscillations to damp out is clearly at odds with the patterns of persistent oscillations shown in Fig. 1.3 and documented in the studies listed in Table 1.5. The damping rate, however, is small and hence the stability of the system is 'weak' (Grossman, 1980). It is therefore to be expected that relatively small destabilizing effects, due to either external excitation or structural changes in the model, could neutralize the damping and bring about undamped oscillatory behaviour.

Stochastic effects and seasonality in transmission can perpetuate the

Table 1.5 The average inter-epidemic period, T, for various diseases (see Anderson and May (1982a) for data sources).

Infectious disease	*Inter-epidemic period (years)*		*Geographical location*	*Time period*
	Average	*Range*		
Measles	2.2	2–4	England and Wales	1855–1968
	2.2	2–3	New York City, USA	1928–1964
	2.3	2–4	Glasgow, Scotland	1929–1968
	2.3	2–3	Providence, RI, USA	1900–1923
	2.6	2–4	Baltimore, Maryland, USA	1928–1964
	2.6	2–5	Hungary	1952–1972
	2.6	2–4	England and Wales	1968–1979
	3.0	2–4	Bulgaria	1952–1972
	3.8	2–6	Providence, RI, USA	1858–1899
Whooping cough	2.5	2–4	Glasgow, Scotland	1928–1955
	2.8	2–5	England and Wales	1855–1955
	3.2	2–4	Baltimore, Maryland, USA	1928–1954
	3.5	2–5	England and Wales	1956–1979
	3.5	3–4	Glasgow, Scotland	1956–1979
	4.0	3–4	Finland	1940–1970
	4.1	2–5	Bulgaria	1921–1972
	4.2	4–5	Hungary	1952–1972
Poliomyelitis	2.8	2–4	Bulgaria	1926–1956
	4.0	3–5	England and Wales	1950–1965
	4.2	2–5	Finland	1940–1972
	4.6	4–5	Netherlands	1950–1965
Chicken pox	2.5	2–4	New York City, USA	1928–1972
	2.8	2–4	Baltimore, Maryland, USA	1929–1973
	3.0	2–4	Glasgow, Scotland	1929–1972
Rubella	3.3	2–5	Glasgow, Scotland	1929–1964
	3.4	2–7	Baltimore, Maryland, USA	1928–1974
	3.7	2–6	New York City, USA	1933–1972
Mumps	3.0	2–4	Baltimore, Maryland, USA	1928–1973
	3.0	2–6	New York City, USA	1928–1967
Diphtheria	5.1	4–6	England and Wales	1897–1979
Scarlet fever	4.4	3–6	England and Wales	1897–1978

oscillations of the system indefinitely. Bartlett (1956), for example, demonstrated that a full stochastic model of an epidemic process, with renewal of susceptibles plus an immigration rate of new infectives, generates an undamped succession of outbreaks. These outbreaks will not follow a strict cycle

but will tend to be rather irregular in their occurrence.

More recently, a series of studies have shown that the inclusion of seasonal periodicity in the transmission parameter β of Equations (1.1) to (1.4) can 'pump' the otherwise-damped deterministic oscillations, locking the system into sustained cycles whose periods are an integral number of years (London and Yorke, 1973; Yorke and London, 1973; Dietz, 1976; Grossman, Gumowski and Dietz, 1977; Yorke *et al.*, 1979; Grossman, 1980). This mechanism has been succinctly described by Yorke *et al.* (1979) in the following manner. 'We may think of the level of susceptibles as similar to a pendulum swinging back and forth past equilibrium. Seasonal variation gives the pendulum a shove every year and these regular shoves are required to keep the pendulum in motion'.

The precise conditions under which this mechanism generates subharmonic resonance (imposed over the regular seasonal cycles in prevalence) depend, in a complicated manner, on the amplitude of the seasonal cycles and on the magnitude of the basic reproductive rate R (Dietz, 1976; Grossman, 1980). Broadly speaking, to induce 2- or more-year cycles both the amplitude of the seasonal transmission and the magnitude of R must be relatively large. These conditions are satisfied, for example, by measles which has very marked seasonality in transmission and hence exhibits the most regular and clearly defined inter-epidemic period (Fig. 1.3). Seasonality in transmission is a common feature of many common childhood infections, some examples of which are shown in Fig. 1.4, but the mechanisms which generate such patterns are poorly understood at present. The main causes are probably climatic effects such as temperature and humidity influencing the survival and dispersal of transmission stages, and seasonal changes in social behaviour such as children returning to school after vacation periods.

Irrespective of the mechanisms which perpetuate the oscillations, the resulting inter-epidemic period is approximately determined by the period T defined in Equation (1.13) which depends on the biological parameters $1/\sigma$, $1/\gamma$, R and L. These relations between D $(1/\sigma + 1/\gamma)$, R (or A) and T are illustrated in Fig. 1.5.

Anderson and May (1982a) have pointed out that there is striking agreement between the simple theoretical insights embodied in Fig. 1.5 and observed values of the epidemiological parameters T, D and A or R (catalogued in Tables 1.5, 1.4 and 1.3). For measles in Britain, the average value of D is around 10 to 16 days (Table 1.3) noting that D is the sum of the latent period plus the effective infectious period, allowing for the fact that children are usually withdrawn from circulation once symptoms appear. The average value of R is about 14 to 18 (Table 1.4), and life expectancy is approximately 70 years. With these figures, Equation (1.13) predicts a value of T between 2 and 3 years, in agreement with the observed inter-epidemic period (Table 1.5). In the case of whooping cough in Britain and North America, the value of D is around 20 to 30 days (Table 1.3) and R is about 14 to 17 (Table 1.4), leading to the prediction

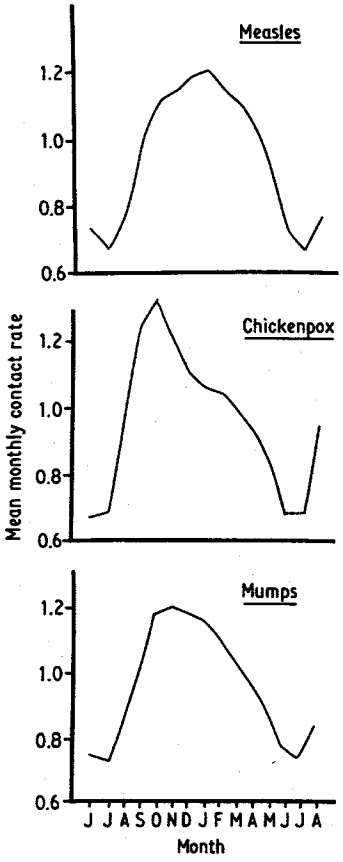

Figure 1.4 Examples of seasonal trends in the mean monthly contact rates (finite transmission rates for measles, chicken pox and mumps in New York City, USA (modified from London and Yorke, 1973)).

that the inter-epidemic period T is around 3 to 4 years. This again is in general agreement with the observations (Table 1.5).

It is important to note that the foregoing analysis pertains to macro-epidemiological patterns in large communities, and substantial variations from the predictions are to be expected in small subpopulations. In addition, within those developing countries where birth rates are high and life expectancy short, Equation (1.13) suggests a pronounced reduction in the inter-epidemic period compared with the corresponding period in a developed country (small L values and high R values in developing countries).

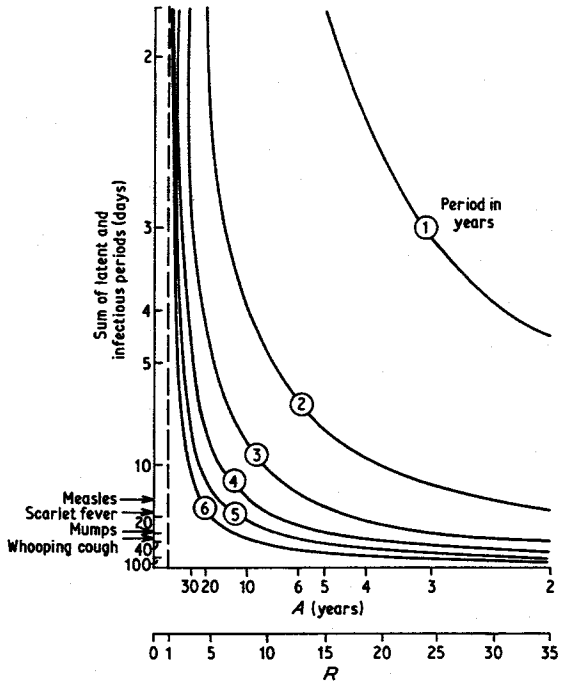

Figure 1.5 This figure shows the inter-epidemic period, T, as a function of the sum of the latent and infectious periods, D, where $D = 1/\sigma + 1/\gamma$, and the basic reproductive rate of the disease, R, for $L = 70$ years. (the approximate relationship defined in Equation (1.13)). Equivalently, as shown in Equation (1.12) T may be expressed as a function of D and A, the average age at which infection is acquired. The contour lines are for specified values of T, labelled according to the period in years. The values of D for some particular diseases are indicated on the y-axis (see Table 1.3) (after Anderson and May, 1982a).

1.6 CONTROL BY IMMUNIZATION

The concept of the basic reproductive rate, R, is central to an understanding both of the population behaviour of infectious diseases and of the impact of control policies (Anderson, 1982b). To eradicate an infection, it is necessary to reduce R below unity. This may be achieved by immunizing a proportion, p, of the population soon after birth, provided

$$p > [1 - 1/R] \tag{1.14}$$

This expression follows from the observation that, in such a population, the number of susceptibles is at most $N' = N(1-p)$, such that the effective

reproductive rate R', is $R' = R(1-p)$ with R as defined in Equation (1.7) (Dietz, 1974; Smith, 1970). The condition $R' < 1$ then gives Equation (1.14). The relation between p and R is illustrated in Fig. 1.6. If R is large, the proportion that must be immunized approaches unity. If other things are equal (such as the efficiency of a vaccine, or the ease of its application within the population) diseases with high R values will be much more difficult to control than those with low values. The effective equivalence between the basic reproductive rate R and the average age at infection, A, defined in Equation (1.12), is further illustrated in Fig. 1.6 where Equation (1.12) is used to re-scale the x-axis to give a relation between p and A for a specified value of L. In other words diseases which have a low average age at infection will be more difficult to control than those with a high average age at infection.

Immunization programmes clearly act to reduce the value of R and hence theory predicts they will tend to increase, both the inter-epidemic period, T, and the average age at infection, A (see Equations (1.12) and (1.13)).

For example, in England and Wales, as a consequence of widescale immunization, the average inter-epidemic period for measles has increased

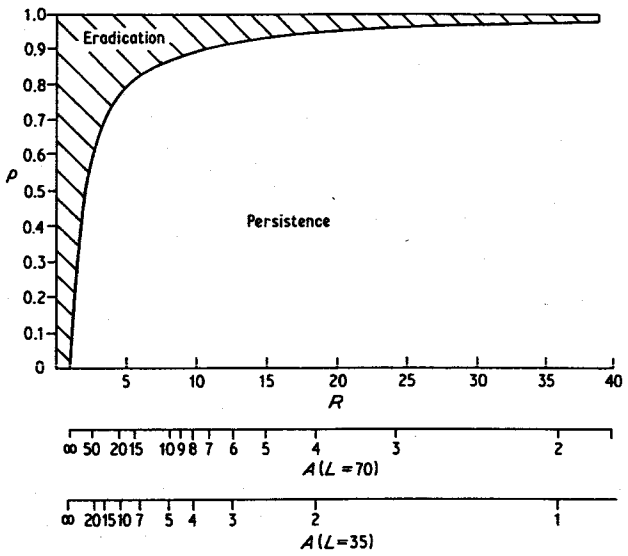

Figure 1.6 This figure illustrates the relation, Equation (1.14), between the proportion, p, of a community that must be immunized (at or near birth) to eradicate an infection, and the basic reproductive rate R. Alternatively p may be expressed as a function of A (using Equation (1.12)); this relationship is shown for $L = 70$ and $L = 35$ years. The infection is eradicated for values of p in the shaded area and it persists otherwise (Anderson and May, 1982a).

from 2.2 to 2.6 years, and for whooping cough it has increased from 2.8 to 3.5 years (Anderson and May, 1982a). The frequency distributions of inter-epidemic periods for both diseases, before and after the introduction of immunization, are displayed in Fig. 1.7. The epidemiology of rubella in the United States is a good example of the impact of immunization on the average age at which individuals acquire infection. As documented in Table 1.6, the introduction of

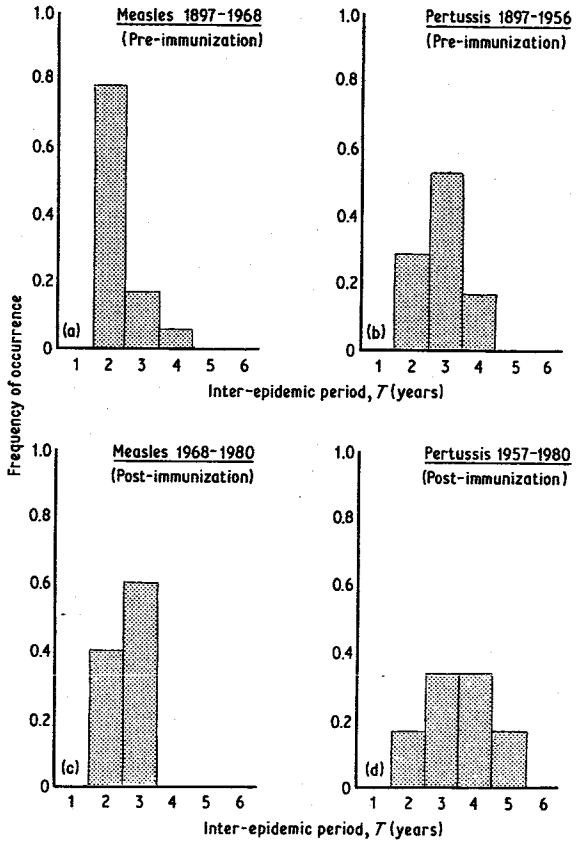

Figure 1.7 This figure shows the frequency distributions of the inter-epidemic period, T, for measles and whooping cough in England and Wales. (a) Measles prior to the introduction of immunization; 1897–1968. (b) Pertussis prior to the introduction of immunization; 1897–1956. (c) Measles after the introduction of immunization; 1968–1980. (d) Pertussis after the advent of immunization, 1957–1980. The mean periods for graphs (a), (b), (c) and (d) are 2.2, 2.8, 2.6 and 3.5 years respectively. Data from the Registrar General's statistical review of England and Wales.

Table 1.6 Cumulative proportion of reported cases of rubella in Illinois, Massachusetts and New York City, USA, by age group for three different periods. A nationwide immunization programme was introduced in the United States in 1969 (data from Hayden, Modlin and Witte, 1977).

	1966–1968	*1969–1971*	*1972–1974*
Age (years)			
0–4	0.216	0.215	0.132
5–9	0.601	0.566	0.345
10–14	0.771	0.726	0.555
15–19	0.898	0.897	0.870
20+	1.000	1.000	1.000
Total number of cases	17960	10709	6512
Average age A (years)	9.6	9.9	13.3
R	8.3	7.1	5.3

a nationwide immunization programme in 1969 increased A from 9.6 years prior to vaccination to 13.3 years by the end of 1974 in Illinois, Massachusetts and New York City (Hayden, Modlin and Witte, 1977). This attribute of immunization programmes has generated wide comment, since the risks associated with infection by certain common viral diseases such as measles and rubella increase with age. In the case of rubella, the infection is of major significance to pregnant women since the virus has a capacity to produce congenital abnormalities (Greenberg *et al.*, 1957), while, for measles, the likelihood of neurological complications (encephalitis) is known to increase with age (Miller, 1964). It is important to note, however, that although immunization will tend to increase the value of A, the *total number* of cases occurring in older individuals may decline as immunization coverage increases. This point is well illustrated by rubella data recorded in Table 1.6.

1.7 AGE-DEPENDENT EPIDEMIOLOGICAL PARAMETERS

The model defined by Equations (1.1) to (1.4) is helpful in illuminating certain basic principles but it suffers from one major short-coming, namely all the rate parameters are assumed to be constant and independent of host age. This is a reasonable assumption with respect to the latent and infectious periods, $1/\sigma$ and $1/\gamma$. Other parameters, however, such as the transmission rate, β, and the mortality rate, μ, are functions of host age. Furthermore, in the context of control, immunization programmes are often highly age-specific. To illustrate this point, Fig. 1.8 records a series of examples of transmission, mortality and vaccination rates at various ages for measles and pertussis in England and Wales.

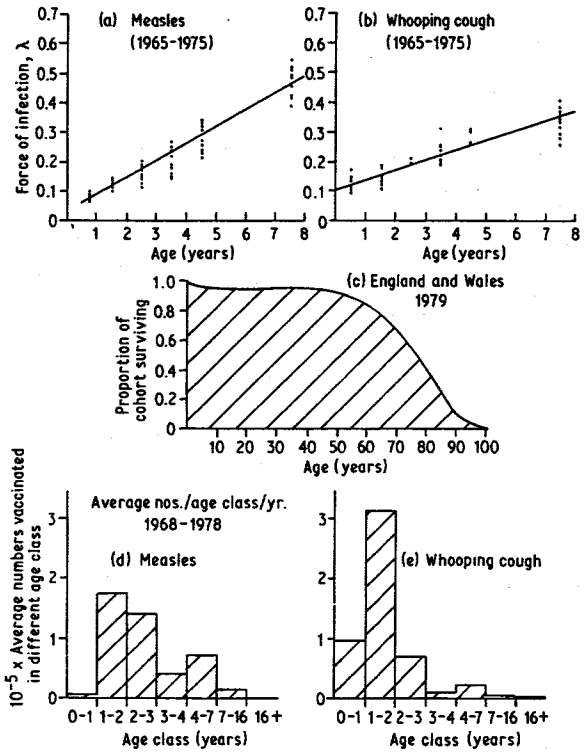

Figure 1.8(a) The 'force', or instantaneous rate, of infection, λ, is shown as a function of age, a, for measles in England and Wales between 1965 and 1975. This rate λ is estimated from case notifications presented in the Registrar General's Statistical Reviews, by methods described by Griffiths (1974); λ is defined per annum per susceptible. The dots represent yearly estimates of the age-dependent rates, while the solid line is the best-fit linear model. Up to the age of 10 years, the relation between λ and age, a, is well described by the linear expression $\lambda(a) = \alpha + \delta a$, where α and δ are constants: $\alpha = 0.030$, $\delta = 0.057$ ($r^2 = 0.99$). (b) As for Fig. 1.8(a), except that the data are for whooping cough. The linear relation between λ and a has coefficients $\alpha = 0.109$, $\delta = 0.033$ ($r^2 = 0.95$). (c) This figure shows the age-dependent survival curve for the population of England and Wales in 1977 (data from the Registrar General's Statistical Review, 1977). The age-specific mortality rate, $\mu(a)$, is the logarithmic derivative of this curve with respect to age, a. (d) The average number of individuals, in various age classes, who were vaccinated against measles in England and Wales between 1968 and 1978 (data supplied by the Department of Health and Social Security, UK). (e) As for Fig. 1.8(d), except the data are for whooping cough from 1965 to 1978.

In developed countries the mortality rate, μ, changes little in the early years of life (the ones most relevant to the dynamics of many common viral and bacterial infections) and hence age-dependency in the transmission parameter is the factor of greatest significance. The estimation of age-dependent infection is usually based on simple catalytic models in which the 'force of infection' is denoted by $\lambda(a)$ at age a. In the terminology of Equations (1.1) to (1.4), the parameter $\lambda(a)$ represents the per capita rate at which susceptibles acquire infection when the disease is at its endemic equilibrium (with the total number of infectious individuals having a constant value Y^*). The parameter $\lambda(a)$ can be written as $\lambda(a) = \beta(a)Y^*$, with $\beta(a)$ denoting the age-specific transmission parameter.

In an analysis of measles-incidence data in England and Wales, Griffiths (1974) suggests that the function $\lambda(a)$ is approximately linear over the age range 0–10 years of age (and this includes over 95% of all reported cases) where

$$\begin{aligned} \lambda(a) &= \alpha + \delta a, \quad t > \tau, \\ \lambda(a) &= 0, \quad t \leqslant \tau \end{aligned} \tag{1.15}$$

The parameters α and δ are constants while τ denotes the time period during which maternal antibodies provide protection from infection in newborn infants. For measles the value of τ is roughly 6 months. If $\bar{x}(a)$ is the proportion of the population that remains uninfected by age a, then the mean age at attack, A, is given by

$$A = \int_0^\infty \bar{x}(a)\,da \tag{1.16}$$

The proportion $\bar{x}(a)$ is defined as

$$\bar{x}(a) = \exp\left[-\int_0^\infty \lambda(a)\,da\right] \tag{1.17}$$

Griffiths discusses the estimation of the parameters of the function $\lambda(a)$ from case notification records for a series of age classes and describes a maximum-likelihood estimation procedure. Ideally seriological data would be the best information on which to base such estimation procedures, but in the case of measles, for example, case notifications provide a reasonable data base since over 95% of all cases occur in the young age classes of the population.

Catalytic models take no account of the recurrent epidemic cycles of infections but since the epidemic process is being averaged over a number of years, including several cycles, the fluctuations in incidence can be effectively ignored. Estimates of the age-dependent force of infection $\lambda(a)$ for measles in England and Wales are displayed in Fig. 1.8(a) and the fit of a simple catalytic model of the form defined in Equation (1.17), to case notification records is shown in Fig. 1.9

Dietz (1974, 1976) and more recently Anderson and May (1982a) have

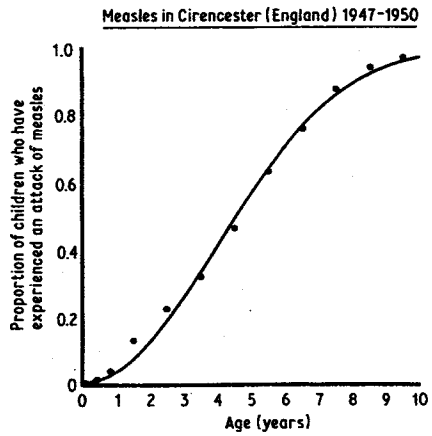

Figure 1.9 The fit of the 'catalytic' model defined in Equation (1.17) in the main text to the proportion of children who experienced an attack of measles at various ages in the Cirencester region of England during 1947–1950. This proportion, $y(a)$, is given by $1 - \bar{x}(a)$ where $\bar{x}(a)$ is as defined in Equation (1.17). The dots are observed points while the solid line is the best-fit 'catalytic' model (data from Griffiths, 1974).

generalized the model described in Equations (1.1) to (1.4) to form a set of partial differential equations describing the change in, for example, the number of susceptibles as a function of time and age, $X(a, t)$. The analysis of such models may be simplified by following the dynamics of a cohort of $\bar{N}$ newly born susceptibles, within a community where the population has a constant size and a stable age distribution, and where the disease is at its endemic equilibrium, Y^*. In order to increase the generality of the following model we assume that disease control is being attempted by an immunization schedule in which susceptibles are vaccinated at an age-dependent rate $c(a)$ and that vaccinated individuals join the immune class and remain protected for life.

The numbers of susceptibles, latent, infectious and immune individuals as functions of age a (denoted by $X(a)$, $H(a)$, $Y(a)$ and $Z(a)$, respectively) in the cohort of size $\bar{N}$ can now be represented by the following differential equations:

$$dX/da = -[\lambda(a) + \mu(a) + c(a)]X(a) \tag{1.18}$$

$$dH/da = \lambda(a)X(a) - [\sigma + \mu(a)]H(a) \tag{1.19}$$

$$dY/da = \sigma H(a) - [\gamma + \mu(a)]Y(a) \tag{1.20}$$

$$dZ/da = \gamma Y(a) + c(a)X(a) - \mu(a)Z(a) \tag{1.21}$$

Individuals leave the susceptible class as a result of natural mortality, vaccination (passing directly to the immune class), and infection passing into

the latent class and thence to the infectious class, and finally to the immune state unless mortality intervenes along the way. The latent and infectious periods, $1/\sigma$ and $1/\gamma$, are assumed to be age-independent. In developed countries the mortality rate, $\mu(a)$, will typically have the type of age-dependence shown in Fig. 1.8(c). The age-specific immunization rates, $c(a)$, can be estimated from the data for a particular programme, as illustrated in Figs 1.8(d) and (e) for measles and whooping cough, respectively, in England and Wales. The parameter $\lambda(a)$, as described earlier, represents the age-dependent 'force of infection', with $\lambda(a) = \beta(a)Y^*$. Equivalently, the total number of infectious individuals can be expressed as $Y^* = y^*N$ where y^* is the equilibrium proportion of the population N that are infectious, whence $\lambda(a) = \beta(a)y^*N$. The force of infection will usually depend linearly on the prevalence of infection, y^*, but it will not always depend linearly on the total population size N. As discussed in a later section of this chapter, for most sexually transmitted diseases, it is likely that λ depends only on y and not on N.

For most childhood viral and bacterial infections λ does depend on N, although, as noted by Anderson and May (1982a), the dependence is often less strong than the conventionally assumed linear dependence of Equations 1.1–1.4. Some evidence presented by these authors is documented in Tables 1.7 and 1.8 which show the mean age at first infection, A, as a function of community size (Table 1.7) and of the degree to which the population is an urban rather than a rural one (Table 1.8) for several diseases. The data are for unvaccinated populations and A is thus inversely proportional to λ. As expected, A tends to increase with decreasing N or decreasing urbanization but the effects are weaker than linear. These complications, however, can be avoided by using Equations (1.18) to (1.21) and simply determining $\lambda(a)$ from empirical data.

From the definition and discussion given earlier (see Equation (1.7)) the basic reproductive rate R for a disease in a given population is, in general, equal to the number of susceptibles there would be in the absence of the disease,

Table 1.7 Average age at first infection, A, as a function of community size, for various childhood diseases (data for New York state in the years 1918–19, see Anderson and May, 1982a).

Community size	*Mean age A (years)*			
	Measles	*Whooping cough*	*Scarlet fever*	*Diphtheria*
200 000–50 000	9.0	6.3	10.5	10.6
50 000–10 000	9.0	5.7	10.2	11.5
10 000–2 500	10.7	6.9	11.2	12.5
under 2 500	12.9	8.2	12.3	14.2

Table 1.8 Average age at first infection, A, in relation to the degree of urbanization for various childhood diseases (data for different states in the USA in the years 1910–22, from reference (data for different states in the USA in the years 1910–22, see Anderson and May, 1982a):

State	*Percentage of population living in urban, rather than rural, communities*	*Mean age A (years)*				
		Measles	*Whooping cough*	*Chicken pox*	*Scarlet fever*	*Diphtheria*
Mass.	94.8	7.3	5.4	–	9.5	9.7
N.J.	78.4	6.7	5.4	7.1	9.7	9.0
Conn.	67.8	7.3	5.6	7.6	10.4	10.5
Penn.	64.3	–	–	–	9.2	9.6
Maryland	60.0	8.4	5.7	7.6	8.9	10.4
N.Y.	57.6	–	–	–	11.1	11.6
Kansas	24.9	10.8	–	–	10.7	12.7

divided by the number of susceptibles when the disease is established at an endemic equilibrium. From Equations (1.18)–(1.21) we therefore obtain (Anderson and May, 1982a)

$$R = \frac{\int_0^\infty \exp\{-\int_0^a [\mu(v)+c(v)]\,dv\}\,da}{\int_0^\infty \exp\{-\int_0^a [\lambda(v)+\mu(v)+c(v)]\,dv\}\,da} \tag{1.22}$$

In the simplest limiting case when all the rate parameters (λ, c, μ) are constants, and in the absence of vaccination ($c = 0$) Equation (1.22) reduces to $R = 1 + (\lambda/\mu)$ where the average age at first infection $A = 1/\lambda$. This gives Equation (1.12) (where $L = 1/\mu$) as discussed earlier. If a proportion p of the population is immunized at the constant rate c, the effective reproductive rate R' under the pressure of vaccination is

$$R' = R[1 - cp/(c+\mu)] \tag{1.23}$$

Here R is the basic reproductive rate before the implementation of immunization.

To eradicate the infection we require $R' < 1$ and from Equations (1.12) and (1.23) the fraction of the population to be protected must satisfy

$$p > \frac{1+V/L}{1+A/L} \tag{1.24}$$

where V is the average age at which individuals are immunized ($V = 1/c$) and A remains the average age at first infection prior to control. Since p cannot exceed unity, it is clear that $V < A$ is a necessary condition for eradication to be achieved.

Provided R is estimated from the age-dependent rates $\lambda(a), \mu(a)$ and $c(a)$ (Equation 1.22) then Equation (1.24) remains a very good approximation for determining the proportion p where A, V and L are the reciprocals of the appropriately average values of the age-dependent rates. The conclusion that eradication is impossible if $V > A$ is of practical importance. In the case of rubella (German measles) in Britain, for example, evidence suggests the value of A is roughly 10–12 years (Knox, 1980). The adopted control policy is to vaccinate girls, and only girls, between 11 and 15 years of age, combined with selective post-partum vaccination in women found not to have antibodies during antenatal care. This clearly protects the individuals most at risk but Equation (1.24) suggests it will have little impact on the overall incidence of rubella in Britain. This prediction is in agreement with available evidence and with experience in the United States where a greater reduction in rubella prevalence has been achieved by vaccinating boys and girls at a pre-school age (see Table 1.6). A detailed appraisal of which of the two policies is the 'better' depends on other factors such as cost–benefit considerations.

The main conclusion to be drawn from the age-structured model defined by Equations (1.19)–(1.21) is that the optimum vaccination policy (optimum in the sense of either eradication or the degree of reduction in disease prevalence) will maintain the value of V as low as possible. It is important to note, however, that the duration of protection provided by maternal antibodies (τ, see Equation (1.15)) must be taken into account in the design of immunization programmes for children.

1.8 MEASLES AND WHOOPING COUGH IN ENGLAND AND WALES

A detailed examination of the impact of immunization programmes on the epidemiology of a disease agent is, of course, dependent on the availability of long-term records of both incidence and numbers vaccinated. Such data are available for both measles and pertussis in England and Wales (Anderson and May, 1982a; Griffiths, 1973).

1.8.1 Measles

The epidemiological trends for measles in England and Wales since 1940 are recorded in Fig. 1.10. Measles is a highly infectious disease with the average age of acquisition, A, being between 4 and 6 years in developed countries. In certain regions of Africa and Asia with high birth rates the value of A is much lower (Table 1.4). Significant differences also exist in the value of A between rural and urban communities (see Tables 1.7 and 1.8) with the average age being higher in smaller and less densely populated areas.

In England and Wales, A decreased from 5.5 years to 4.4 years between 1944 and the introduction of widescale immunization in 1968 (Fig. 1.11a), a trend

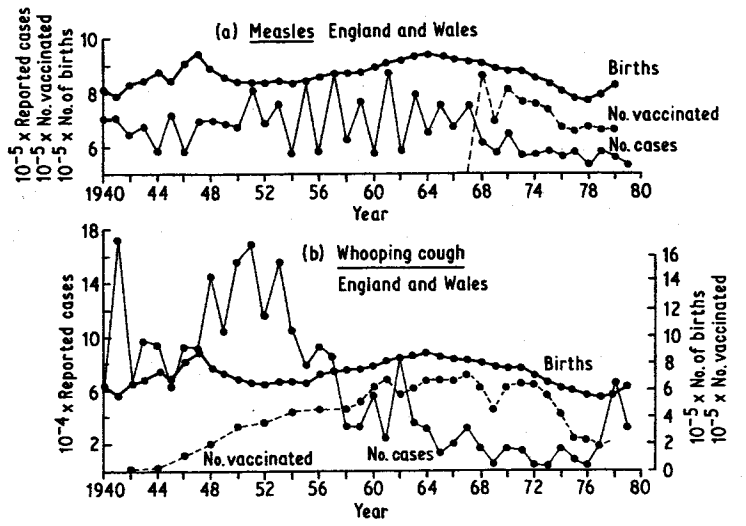

Figure 1.10 Reported cases of (a) measles and (b) whooping cough in England and Wales are shown, from 1940 to 1979. The figures show the total number of births (thick line), the total number of reported cases (thin line) and the total number of people vaccinated (dashed line) each year (Anderson and May, 1982a).

thought to be due to greater intermixing within the population and increased population density. Since the introduction of immunization, the trend has been reversed and both the inter-epidemic period, T (Fig. 1.7), and the average age at infection, A (Fig. 1.11a), have increased. Throughout the span of 1944 to 1979, 90–98% of reported cases have been in children less than 10 years old (Fig. 1.11b). By following specific cohorts through time to monitor the decline in the proportion that are susceptible, these data have been used to estimate that the degree of under-reporting of cases, on a national scale, lies between 40 and 45% (Anderson and May, 1982a). Similar estimates have been made in North America (Yorke *et al.*, 1979).

Vaccination coverage in England and Wales has been relatively low. As recorded in Fig. 1.12(b), of each yearly cohort, 15–35% have been vaccinated by an average age at vaccination, V, of between 2.0 and 2.6 years. In total, roughly 46–57% of each cohort has been vaccinated since 1968. This level of immunization has had relatively little impact on the reproductive rate R. For example, Fig. 1.13(a) records the decline in the susceptible population over time for the 1956 cohort (pre-vaccination) and for the 1970 cohort (post-vaccination). Anderson and May (1982a) estimated the value of R (using Equation (1.22)) to be in the range 14 to 18 pre-vaccination and in the range 12 to 13 post-vaccination. Thus the vaccination of a total of the order of 50% of

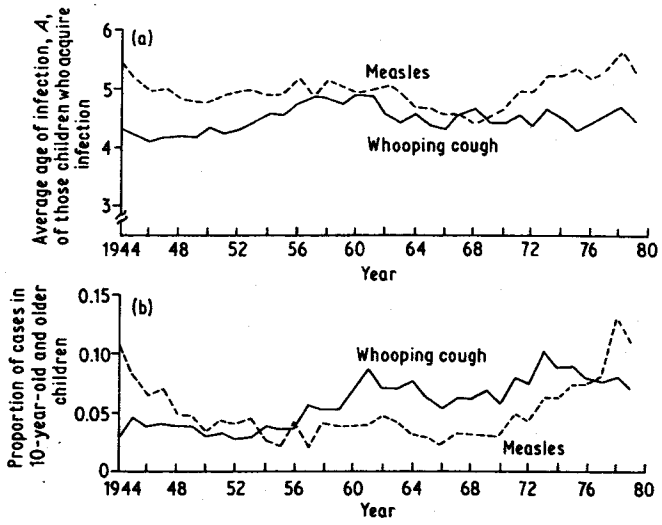

Figure 1.11 (a) This figure shows the average age, A, at which children experienced an attack of measles (dashed line) or pertussis (solid line) in England and Wales over the years 1944 to 1979. The values of A were estimated by numerical integration of the expression

$$A = \int_{-\tau}^{t} \exp\left[-\int_{\tau}^{a} \lambda(v)\mathrm{d}v\right]\mathrm{d}a$$

where the age-dependent infection rate, $\lambda(v)$, for each specific year is as defined in the text. The limits of the integral are τ, defined to be the interval of time after birth during which maternal antibodies protect a newborn child against infection (assumed to be 0.5 years for measles, and negligible for pertussis), and t, an arbitrarily determined upper limit (which is set at 10 years since 90–98 % of cases occur in children less than this age). In calculating A, the death rate is taken to be negligible during the first 10 years of life (Anderson and May, 1982a). (b) This figure shows the proportion of reported cases of measles (dashed line) and whooping cough (solid line) that were from individuals past the age of 10 years, in England and Wales from 1944 to 1979.

whom roughly 30 % are vaccinated by the average age of vaccination at roughly 2–2.2 years reduced the value of R by about 20 %.

On the basis of Equation (1.24), with average A and V values of 4.6 and 2.2 years respectively, it appears as though approximately 96 % of each cohort would have to be immunized for measles to be eradicated in Britain. Figure 1.14 depicts the relations between p, V and A as predicted by Equation (1.24). Even if the average age at vaccination was reduced to close to, or less than, 1 year, to eradicate the disease would require an immunization coverage of close to 94 %.

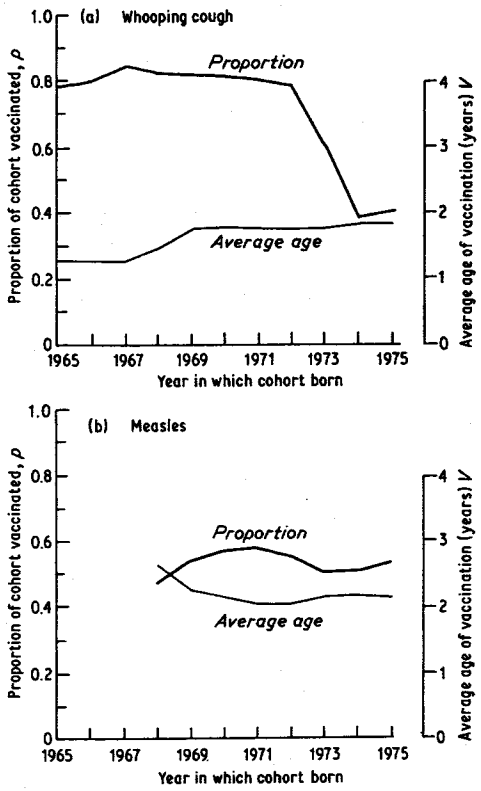

Figure 1.12(a) The thick line (corresponding to the vertical axis to the left) shows the proportion of the cohorts born in England and Wales during the years 1965 to 1975 who were vaccinated against whooping cough. The thin line (corresponding to the vertical axis to the right) represents the average age, V, at which vaccination was received by the individuals in these cohorts who were vaccinated (data from the Department of Health and Social Security, UK). (b) As for (a), except that the data are for vaccination against measles in cohorts born in the years 1968 to 1975.

This figure is an average value for England and Wales and higher levels of coverage would be required in densely populated cities and lower levels in rural communities. Stochastic effects, accentuated by seasonality in disease transmission, would probably result in the fade out of infection at marginally lower levels of protection than those predicted by deterministic models. The magnitude of the predicted level of immunization required to eradicate measles is a direct consequence of the high infectiousness of this disease (high R values and low A values).

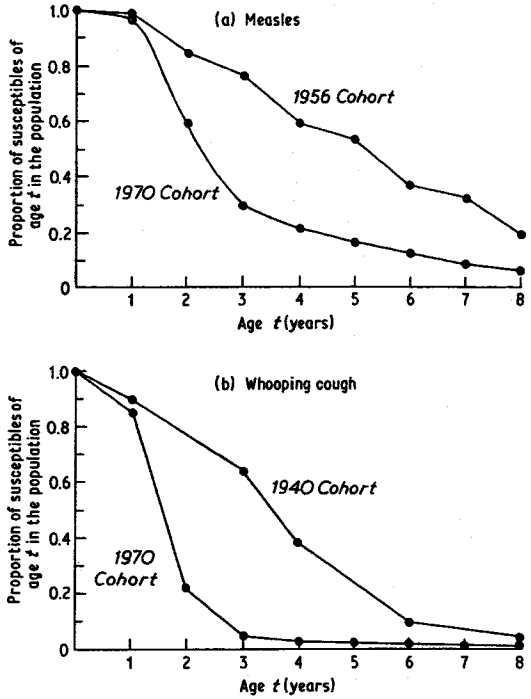

Figure 1.13(a) The estimated decline in the proportions of the 1956 and 1970 cohorts that were susceptible to measles is shown as a function of age, in England and Wales. The cohorts were chosen to represent epidemiological patterns before (1956) and after (1970) the introduction of immunization programmes for measles (see Anderson and May (1982a) for further details). The basic reproductive rate, R, for measles within these cohorts was estimated by numerical integration of Equation (1.22): for the 1956 cohort, $R = 16.0$; and for the 1970 cohort, $R = 12.8$. Of the 1970 cohort, a total of 57% were vaccinated at an average age, V, of 2.3 years. (b) As for (a) except this figure shows the decline in the proportions of the 1940 and 1970 cohorts susceptible to whooping cough as a function of age, in England and Wales. The cohorts were again chosen to show patterns before (1940) and after (1970) the introduction of immunization programmes. Values for R were estimated as $R = 16.3$ and 6.3 for the 1940 and 1970 cohorts respectively (see Anderson and May, 1982a). Of the 1970 cohort a total of 81% were vaccinated at an average age, V, of 1.7 years.

1.8.2 Whooping cough

Anderson and May (1982a) have also analysed, along the lines outlined above, the available data for pertussis in England and Wales between 1940 and 1979 (Fig. 1.10). Their analysis, based on the decline in the proportion of

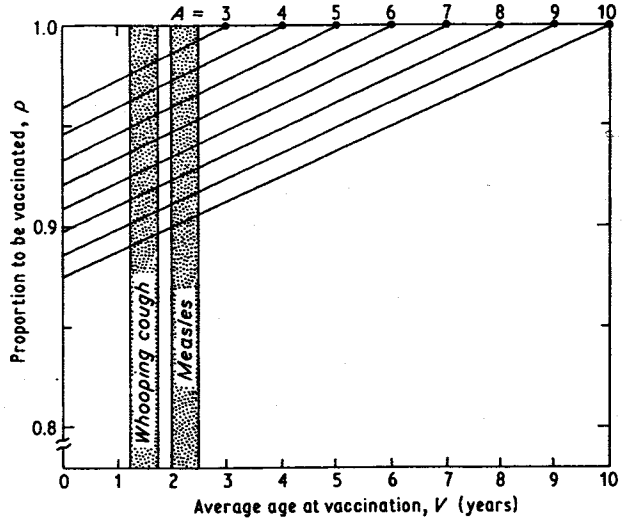

Figure 1.14 This figure shows the approximate relationship, Equation (1.24), between the proportion, p, of each cohort that must be vaccinated, and the average age at vaccination, V, for eradication of an infection. The solid lines represent the boundaries between eradication (above the line) and persistence (below the line) for various values of the average age at first infection, A, in the population before the introduction of immunization. The average life expectancy is taken to be 70 years. The shaded vertical bands depict the ranges of values of V for immunization against whooping cough and against measles, in England and Wales, during the periods 1965–1975 and 1968–1975 respectively (Anderson and May, 1982a).

susceptibles pre-vaccination and post-vaccination, gave values of R of roughly 16 and 6 respectively (Fig. 1.13(b)). In the 1970 cohort, for example, a vaccination coverage of 81% in total and 42% by the average age of vaccination at 1.7 years produced a 61% decline in the value of R and a very low overall incidence of the disease during the early 1970s. Since this time, however, public concern about the dangers of vaccination for whooping cough has led to a substantial decline in the level of vaccination and concomitantly the incidence of the disease has increased (Fig. 1.10). Anderson and May (1982a) suggest, on the basis of Equation (1.24), that with an average age of vaccination around 1.7 years (see Fig. 1.12(a)), a coverage of about 96% would be necessary to eradicate whooping cough in Britain.

1.9 SEXUALLY TRANSMITTED INFECTIONS

The widespread occurrence and increasing incidence of venereal disease, or sexually transmitted infections, is today a major public health problem in

developed and developing countries alike. In the United States in 1973, for example, approximately 767 000 cases of gonococcal infections and 91 000 cases of syphilis were reported. Relevant surveys suggest a high degree of under-reporting of such infections and it has been estimated that the overall incidence of gonorrhoea in the USA is in excess of 2.5 million out of a total population of 250 million. The total world incidence of sexually transmitted diseases, especially syphilis, gonorrhoea and non-gonococcal urethritis is probably surpassed only by such diseases as malaria and roundworm infections (Bailey, 1979). A variety of socioeconomic factors are thought to be responsible for the recent rise in the global prevalence of venereal disease. These include increasing urbanization, migrant labour, expanding tourist travel and new attitudes towards sexual behaviour in modern societies.

In recent years a variety of models have been developed to mimic the dynamics of sexually transmitted infections and some of these have been designed to explore the efficiencies of various control options (Cooke and Yorke, 1973; Hethcote, 1974, 1976; Lajmanovich and Yorke, 1976; Yorke, Hethcote and Nold, 1978). In this section one of the sexually transmitted infections is considered, namely the bacterium *Neisseria gonorrhaeae* (causal agent of gonorrhoea).

Gonorrhoea has distinctive epidemiological characteristics when compared with the viral and bacterial infections discussed in the preceding sections of this chapter. First, the disease occurs only within the sexually active portion of a community. Second, the duration of infectiousness is long and differs substantially between male and female individuals. Third, and finally, acquired immunity to reinfection is virtually non-existent and hence recovered individuals pass directly back to the susceptible pool.

Models of the dynamics of gonorrhoea conventionally consider a freely mixing community which consists of sexually active males and females of densities N_1 and N_2 respectively. In the following model we assume that in a short interval of time, X_2 susceptible females have sexual contact with S_2X_2 males of whom $S_2X_2Y_1/N_1$ are infected. The parameter S_2 represents the average number of partners that a female has sexual contact with in a defined unit of time. We further assume that the probability of infection passing from an infected male to a susceptible female during a single 'partner contact' is q_2. The net rate at which new female infections arise is therefore $S_2q_2X_2Y_1/N_1$. For convenience we define $\hat{\beta}_1 = S_2q_2/N_1$. In a similar manner the net rate at which male infections arise is $S_1q_1X_1Y_2/N_2$ and we define $\hat{\beta}_2 = S_1q_1/N_2$. The parameter S_1 denotes the average number of partners that a male has sexual contact with per unit of time and q_1 is the probability of infection passing from an infected female to a susceptible male during a single 'partner contact'. The average duration of male and female infectiousness is denoted by $1/\gamma_1$ and $1/\gamma_2$ respectively. If the populations of sexually active males and females are

constant, then the above assumption gives rise to the following differential Equations:

$$dX_1/dt = -\hat{\beta}_2X_1Y_2 + \gamma_1Y_1 \quad (1.25)$$

$$dY_1/dt = \hat{\beta}_2X_1Y_2 - \gamma_1Y_1 \quad (1.26)$$

$$dX_2/dt = -\hat{\beta}_1X_2Y_1 + \gamma_2Y_2 \quad (1.27)$$

$$dY_2/dt = \hat{\beta}_1X_2Y_1 - \gamma_2Y_2 \quad (1.28)$$

This model predicts that the infection will be maintained within the population provided the basic reproductive rate, R, is greater than, or equal to, unity. For the system defined in Equations (1.25)–(1.28)

$$R = (S_1S_2q_1q_2)/(\gamma_1\gamma_2) \quad (1.29)$$

Note that in contrast to the basic reproductive rate of the viral and bacterial infections discussed in earlier sections of this chapter (see Equation (1.5)), the expression defined in Equation (1.29) is independent of host population size. In other words there is no critical threshold density for disease persistence. The maintenance of gonorrhoea is thus simply dependent on the prevailing degree of sexual promiscuity within the community (i.e. the magnitudes of the parameters S_1 and S_2).

The estimation of the parameter R from empirical data is fraught with problems since it is dependent on the availability of accurate information on sexual habits. These clearly vary greatly both within different sections of a given community and between communities. With respect to the infectious periods, $1/\gamma_1$ and $1/\gamma_2$, the most widely quoted values are 10 days for men and 100 days for women (Constable, 1975; Reynolds and Chan, 1975). These values, however, are fairly crude and based on limited evidence (Yorke *et al.*, 1978). Some attempts have been made to estimate average values for S_1 and S_2 on the basis of interviews, but such information is known to be highly unreliable. In a study by Darrow (1975), for example, of a large city in North America, the average patient with gonococcal infection reported 1.46 partners during the preceding 30 days. This figure is too low since at endemic equilibrium ($R = 1$) theory suggests that on average each infective person must have two effective contacts during the course of an infection, namely, a contact with an infector (the source of the infection) and a contact with a person to whom the infection is transmitted. An infective person, may of course have additional sex partners to whom the infection is not transmitted (the probabilities q_1 and q_2).

The discrepancy between theory and data may be due to several factors not least of which is the honesty of the individuals interviewed. Other factors, however, may also play an important role. The model defined in Equations (1.25) to (1.28) is based on the assumption that the sexually active proportion of the community mixes in a homogeneous manner. This is clearly far from the

truth, some individuals have many more sex partners than others. In fact, many epidemiologists believe that in most communities the frequency distribution of sexual partners per unit of time is highly skewed (the variance being much greater than the mean) with the majority of individuals having one or no contacts (in a unit time interval) and a few individuals having very many contacts. The sexually active individuals in the tail of this distribution in effect form a 'core' population which is, by itself, almost entirely responsible for the maintenance of gonorrhoea in the community as a whole. In the 'non-core' segment of the population the value of R is probably much less than unity so that the activities of the core maintain the infection by continually reintroducing it into the remainder of the population.

The core is thought to consist primarily of women, whose infections would often be asymptomatic for long periods. It will also contain some asymptomatic men and people who continue sexual intercourse in spite of symptoms. Control programmes (based on screening and drug treatment) aimed at this core segment of the population are potentially highly effective (Yorke *et al.*, 1978).

One final point to note concerning the epidemiology of sexually transmitted infections relates to the concept of an endemic disease equilibrium. As illustrated in Fig. 1.15 the incidence of infections such as gonorrhoea has increased substantially in many regions of the world over the past two decades. This implies that the parameters which control the dynamics of such infections are not in reality constants (as defined in Equations (1.25)–(1.28)) but are in fact changing with time. We therefore have a situation in which the equilibrium predicted by simple deterministic models is moving in time as the parameters which determine this state change. For sexually transmitted infections, it is highly probable that changes in the contact rates S_1 and S_2 are responsible for the observed epidemiological trends.

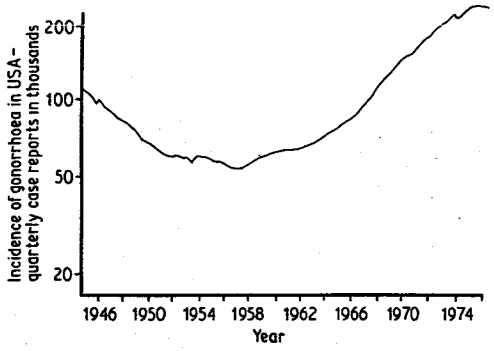

Figure 1.15 This figure shows the quarter-yearly case reports of gonorrhoea in the United States between the years 1944 and 1976 (data from Yorke *et al.*, 1978).

1.10 CONCLUSIONS

The development of safe, effective and cheap vaccines which provide lasting (ideally lifelong) protection is clearly a first step for the successful control of many common viral and bacterial infections. Once these aims have been achieved, however, as they have for diseases such as measles, polio and rubella, important epidemiological questions remain to be resolved. Careful thought, for example, must be given to such issues as the level of artificially induced herd immunity required to eradicate an infection, the degree of reduction in case notifications to be expected from a given immunization policy and the influence of vaccination upon the average age at first infection and upon the time period between major epidemics. In certain cases, the failure to tackle these issues has led to some degree of confusion concerning the type of immunization policy to be adopted (Sutherland and Fayers, 1971; Roden and Heath, 1977; Galbraith, Forbes and Mayon–White, 1980).

Any attempt to answer these questions involves a knowledge both of the typical course of infection within an individual (e.g. the latent period and the duration of infectiousness) and of the overall population biology of the disease agent and its host. Theoretical studies of the population dynamics of infectious diseases can play a central role in the acquisition of both types of information. In the case of parameters which determine the course of an infection within an individual, mathematical models provide a framework for parameter estimation from empirical data. With respect to population behaviour, theory can help to create a basic understanding of the factors controlling dynamical behaviour. In this latter context, the concept of a basic reproductive rate, and its definition in terms of a few simple biological properties of the disease agent and its host population, are of central importance. Theory suggests that the design and implementation of immunization programmes should ideally be based on quantitative assessments of their overall impact on the quantity R, the aim being to reduce the basic reproductive rate to a level, close to, or below, unity.

Assessments of the extent to which specific infections may be controlled by immunization requires serological surveys to measure the average age at which infection is acquired and thence to estimate typical values of R (at local, regional and national scales). For those infections for which vaccines are, or soon will be, available (such as cytomegalovirus and hepatitis B virus), it would appear prudent to acquire this information before immunization schedules are introduced. Once the level of protection necessary for eradication, or a defined reduction in incidence, has been determined, other issues which influence public health policy, such as social and cost–benefit considerations, may be examined in much more rigorous terms. Further research of a statistical and mathematical nature could be of great value to such deliberations particularly if it is based on the analysis of data. There is a need, for example, for more sophisticated statistical methods to aid in the estimation of R from serological data or case notification records, and for a better understanding of the significance of non-homogeneous mixing to disease dynamics.