J. Hyg., Camb. (1985), **95**, 419–436
Printed in Great Britain

The estimation of age-related rates of infection from case notifications and serological data

By B. T. GRENFELL and R. M. ANDERSON

Department of Pure and Applied Biology, Imperial College, London University, London SW7 2BB

(*Received* 11 *March* 1985; *accepted* 7 *May* 1985)

SUMMARY

The paper describes a maximum-likelihood method for the estimation of age-related changes in the *per capita* rate of infection, from case notification records or serological data. The methods are applied to records of measles incidence in the UK and USA, for which the estimated rates of infection tend to rise to a maximum value at around 10 years of age and then to decline in the older age-classes. Longer-term and seasonal trends are analysed by reference to changes in the estimated average age at infection; a statistic derived from a knowledge of the age-specific rates of infection. Future data needs in the epidemiological study of directly transmitted viral and bacterial diseases are discussed with reference to the detection and interpretation of age-dependent rates of disease transmission.

INTRODUCTION

In the study of the transmission dynamics and epidemiology of common viral and bacterial infections of man, case reports stratified according to age are an important source of information (for a recent review see Anderson & May, 1983). The dimensions of age and time are equivalent and hence age-related changes can reflect temporal changes in the rate or force of disease transmission within a community. Quantitative studies of age-incidence distributions have a long history in the epidemiological literature, where early work was based on case reports or patient interviews, while recent research is more commonly founded on serological survey (Wilson, 1904; Collins, 1924, 1929; Sydenstricker, 1926; Fales, 1928; Hedrich, 1930; Griffiths, 1974; Anderson, 1982; Anderson & May, 1982, 1983, 1985). Past and present surveys are normally horizontal, as opposed to longitudinal, in design.

One of the earliest quantitative investigations was that of Collins (1924, 1929) in North America. This study examined age-specific case reports for a variety of common childhood infections, such as measles, mumps, chicken pox, scarlet fever, whooping cough and diphtheria. Analyses were based on an age-specific '*incidence rate*', defined as the number of reported cases per unit of time in a given age class divided by the total number of individuals in that age class. Today this parameter is termed the age-specific '*attack rate*' (often defined per 1000 head of population) and is widely employed in epidemiological research. This statistic, however, has serious limitations since, by definition, it takes no account of the numbers in each

age class who are actually *susceptible* to infection. For example, low attack rates in old age classes may simply reflect low proportions susceptible to infection (most being immune as a consequence of infection at an earlier age), as opposed to a real decline in the rate at which susceptible individuals acquire infections. In other words, the interpretation of age-related changes in attack rates is confounded by age-related changes in the proportion immune (or susceptible) in each age-class.

A precise measure of the rate at which *susceptibles* acquire infection was first proposed by Muench (1959), who employed simple mathematical models to mirror age-related changes in the proportion of individuals who had experienced infection. These models were described as *catalytic* because of their structural similarities to equations commonly employed in the study of chemical reactions. Muench (1959) employed a parameter termed the '*force of infection*' (denoted by the symbol λ), which he defined as the instantaneous *per capita* rate at which susceptible individuals acquire infection. For example, if there are $x(a, t)$ susceptibles of age a at time t and $x(a+1, t+1)$ susceptibles of age $a+1$ at time $t+1$, then λ (per unit of time) is simply defined as $\lambda = -\ln[x(a+1, t+1)/x(a, t)]$. The magnitude of the force of infection is inversely related to the average age, A, at which an individual typically experiences infection within a community. If the value of λ is independent of age, A is simply $1/\lambda$ (Anderson & May, 1982, 1983).

The *force of infection* is of much greater value in epidemiological study than the widely used *attack rate*, as a direct consequence of its ability to distinguish underlying age-related changes in rates of infection independent of changes in the proportion of susceptibles in each age-class. The first extensive study of the use of the force of infection as a measure of age-related changes in disease incidence was that of Griffiths (1974), who employed a model in which λ increased linearly with age over the range 0·5–10 years of age. Griffiths (1974) developed a maximum-likelihood method for estimating the parameters of the model, and applied the technique in an analysis of age-specific data on measles incidence in England and Wales. Anderson & May (1982) have since employed Griffiths' technique in a more extensive study of measles and pertusssis in the United Kingdom. This analysis was subsequently extended to cover a variety of common childhood infections, which revealed a general pattern, in which infection rates (λ's) tended to increase during the early childhood years and then decline in older age classes (Anderson, Grenfell & May, 1984; Anderson & May, 1983, 1985).

Observed patterns of age-related change in λ are of considerable applied significance, although their interpretation is a matter of some uncertainty at present (Schenzle, 1985; Anderson & May, 1983, 1985). The force of infection acting on a given age group of susceptibles is determined by their effective contact rate (in the sense of contacts which result in infection) with infectious individuals in all other age classes. The highest contact rates invariably occur among children of 5–10 years. Mass vaccination acts to increase the average age at which an individual typically acquires an infection over that pertaining prior to control. As such, control may shift the average infection age from an age group with a high intrinsic contact rate to one with a lower contact rate. The outcome of this change is to increase the predicted effectiveness of a given level of vaccination coverage over that derived from calculations based on the assumption that the force of infection is constant and independent of age (Schenzle, 1985; Anderson & May,

1985). The precise details of such calculations depend critically on the manner in which the force of infection changes with age (Anderson & May, 1985).

This present paper examines age-related changes in λ by reference to the epidemiology of measles. We extend Griffiths' (1974) model to encompass a general polynomial description of changes in λ with age and derive a stepwise maximum-likelihood method for parameter estimation from data sets consisting of case notifications or serological information. We examine changes in the patterns of age-dependent forces of infection, both with respect to annual and longer-term temporal trends, by analyses of measles notifications for England and Wales over the past 36 years.

METHODS

(1) *Polynomial 'catalytic' infection model*

The simple catalytic model of Muench (1959) assumes that the force of infection, λ (defined per susceptible person per unit of time), is constant and independent of age. Furthermore, infection is taken to induce life-long immunity and not to influence the mortality rate of infected individuals. Within a population of constant size and stable age distribution, these assumptions lead to a simple differential equation model for changes in the proportion of susceptibles in age-class a, $x(a)$, with respect to age:

$$dx(a)/da = -\lambda x(a). \tag{1}$$

Given the assumption that all individuals are susceptible at birth ($a = 0$, $x(0) = 1$), equation (1) has the solution

$$x(a) = \exp(-\lambda a). \tag{2}$$

The proportion immune in the population at age a, $y(a)$, is simply $y(a) = 1 - x(a)$. More generally, equation (1) can be expressed in terms of the cumulative distribution function of age (a) at infection, $F(a)$ (the proportion of a cohort, all of whom were susceptible at birth, who have experienced the infection (i.e. who are immune) by age a), where

$$dF(a)/da = \lambda[1 - F(a)]. \tag{3}$$

This model has two major shortcomings: it assumes that λ is independent of age and time and that all members of a given cohort are susceptibles at birth. The parameter λ can be expressed as a function of age, $\lambda(a)$, such that the solution of equation (3) is

$$F(a) = 1 - \exp\left[-\int_0^a \lambda(\alpha)\, d\alpha\right]. \tag{4}$$

To account for maternally derived antibodies in children born from mothers who have experienced the infection, $\lambda(a)$ can be set to zero below a lower age threshold L (the value of L is roughly 0·25–0·5 years for infections such as measles and rubella; Anderson & May, 1983). The effects of vaccination on our estimation of the force of infection are discussed in Section (4) below.

https://doi.org/10.1017/S0022172400062859 Published online by Cambridge University Press

Griffiths assumed that the function $\lambda(a)$ was linear in form. We generalize this model to a polynomial function (of degree k), where

$$\lambda(a) = \sum_{i=0}^{k} b_i a^i \quad (L < a \leqslant U), \tag{5}$$
$$\lambda(a) = 0 \quad (a \leqslant L).$$

The upper age limit U denotes human life expectancy or the oldest age class for which data is available to estimate the parameters (the b_i's) of equation (5). Combining equations (4) and (5) gives

$$F(a) = 1 - \exp\left[-\sum_{i=0}^{k} [b_i(a^{i+1} - L^{i+1})/(i+1)]\right] \quad (L < a \leqslant U), \tag{6}$$
$$F(a) = 0 \quad (a \leqslant L).$$

(2) *The mean age at infection, A*

The average age, A, at which individuals experience infection is defined as

$$A = \int_0^\infty [1 - F(a)]\,da. \tag{7}$$

For a constant, age-independent force of infection, A is simply the reciprocal of λ. More generally, combining equations (6) and (7),

$$A = L + \exp\left[-\int_L^\infty [b_i(a^{i+1} - L^{i+1})/i+1)]\right]. \tag{8}$$

Given estimates of L and the b_i's this equation can be evaluated by standard numerical methods. Since the force of infection polynomial (equation (5)) cannot be evaluated above age U (the oldest age-class for which data is available), the upper limit of the integral in equation (8) must be set to U. This results in an underestimate of A if U is significantly less than the life expectancy of individuals within the population. For infections such as measles, however, the magnitude of this error is small, since most individuals experience infection by the age of 10–15 years.

(3) *Parameter estimation*

The parameters of equation (6) can be estimated from observed data recording the change in the function $F(a)$ with age (either cumulative case reports or serological data). We follow Griffiths (1974) by assuming a binomial distribution for $F(a)$ and estimating the parameters (the b_i's) by maximum likelihood. We define n samples at ages a_j ($j = 1, \ldots, n$), drawn from a population in which the infection is endemic. Each sample consists of N_j individuals, of which R_j have experienced the infection (Griffiths, 1974). The binomial likelihood of these observations is given by

$$L = \prod_{j=1}^{n} f(a_j)^{R_j} [1 - F(a_j)]^{N_j - R_j}, \tag{9}$$

https://doi.org/10.1017/S0022172400062859 Published online by Cambridge University Press

where $F(a_j)$, the cumulative proportion infected at age a_j, is given by equation (6). Maximum-likelihood estimates of the parameters (L and b_j, $i = 0, \ldots, k$) are then obtained by minimizing the quantity, $-\log(L)$, where

$$-\log(L) = -\left[\sum_{j=1}^{n} [\log[F(a_j)]R_j + \log[1-F(a_j)](N_j - R_j)]\right] \tag{10}$$

by a standard numerical algorithm.

To assess the relative goodness of fit of different force of infection polynomials (equation (5)) we employ the statistic, D, where

$$D = -2\left[\sum_{j=1}^{n} \log[F(a_j)/E_j]R_j + \log[1-F(a_j)]/(1-E_j)](N_j - R_j)\right]. \tag{11}$$

Here $E_j = R_k/N_j$ and $F(a_j)$ is the value of $F(a_j)$ at the maximum-likelihood solution. Equation (11) is the definition of deviance given by Nelder & Wedderburn (1972). In particular, we compare different force of infection polynomial fits by the quantity

$$M = D_k/\nu_k, \tag{12}$$

where D_k is the deviance associated with a polynomial of degree k, and $\nu_k = n-(k+2)$ is the degrees of freedom. The polynomial which minimizes equation (12) is defined as the 'best fit' to the data allowing for the degrees of freedom available in any given data set.

(4) *Case notification and serological data*

The observed proportional histories of measles infection by age ($F(a)$) in any given community can be derived from two different sources. Firstly, the cumulative proportion infected by age can be estimated from age-stratified case-notification records (see Anderson & May, 1983; Griffiths, 1974). In principle, this type of data should be weighted according to the number of people in each age-class within the total population. In most developed countries, however, the age distribution is relatively uniform over the age range 0–40 years (this is not the case in developing countries with high child-mortality rates). The major limitation of case reports concerns age-related biases in case reporting. It is widely believed, for example, that the probability of a case being reported in very young children is somewhat higher than that for the adult age-class. Since we use proportional case reports, the problem of under-notification (Fine & Clarkson, 1982*b*) will not affect our analysis as long as notifications can be assumed to be representative of the true age distribution of cases.

The second source of data arises from age-stratified serological surveys, which provide information on the proportion of immunes. In the absence of vaccination, such data, in principle, correspond directly to the cumulative proportion infected ($F(a)$) by age (a), although the distinction between seropositive and seronegative results is somewhat arbitrary. Care must also be exercised in the interpretation of antibodies detected in very young children since they may either be maternally derived or a result of experience of infection (for $a < 0{\cdot}5$ years). Serology, if based on random samples drawn from the population (as opposed to samples arising from hospital admissions or blood sera submitted for testing on the grounds of suspected

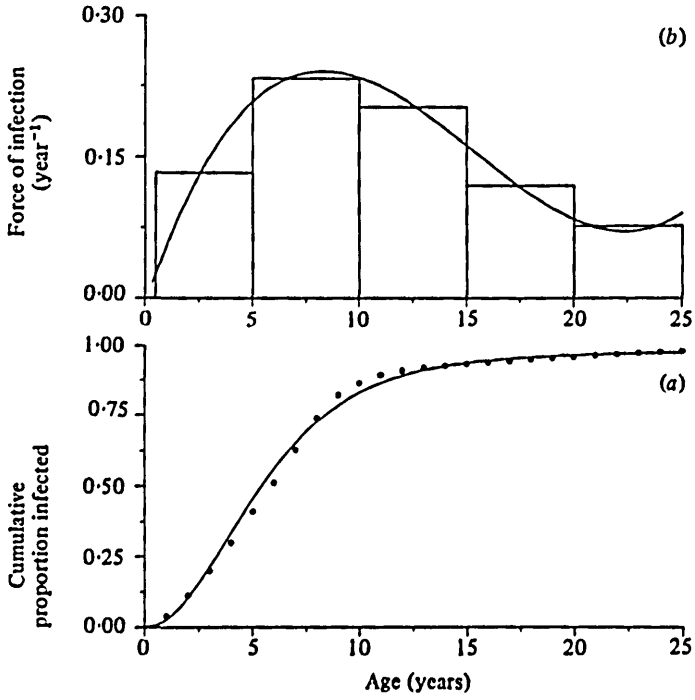

Fig. 1. Analysis of measles notifications for Baltimore, USA, 1906–15. (*a*) Observed and expected cumulative proportions infected by age ($F(a)$) in equation (5). (*b*) The fitted force of infection polynomial ($\lambda(a)$ in equation (4)). Here and in Figs. 2–7 the histogram represents average force of infection estimates (in the age ranges 0·5–5 years, 5–10 years, etc.) derived from the fitted polynomial, which is documented in Table 1.

infection) provide a better source of information than case notifications (given the biases and inaccuracies of case notification systems). Serological profiles, however, are still subject to bias. For example, the ability of current serodiagnostic techniques to detect low antibody titres in the older age-groups of a population (where the interval between infection and test is many decades) is uncertain at present. In practice, the choice of data is primarily governed by availability; information on case reports data is much more abundant than detailed age-stratified serology.

The effects of mass vaccination on our estimation of the force of infection depend upon the form of data used. Estimates of λ based on case reports implicitly exclude vaccinated individuals from the cohort considered. As long as vaccination takes place in early childhood (such is the case for measles) estimates of the subsequent force of infection are unaffected. Age-stratified serological data (which are currently rare for the vaccination era, and are not considered here) include vaccinated individuals in the immune class.

(5) *Data sources*

Our analyses are based on measles case-reports and serology from the United Kingdom and the United States. The case notifications are accumulated over several epidemic cycles (2-year periods) to average out short-term temporal variations in age-specific forces of infection (Fine & Clarkson, 1982*a*). They derive from the following sources.

https://doi.org/10.1017/S0022172400062859 Published online by Cambridge University Press

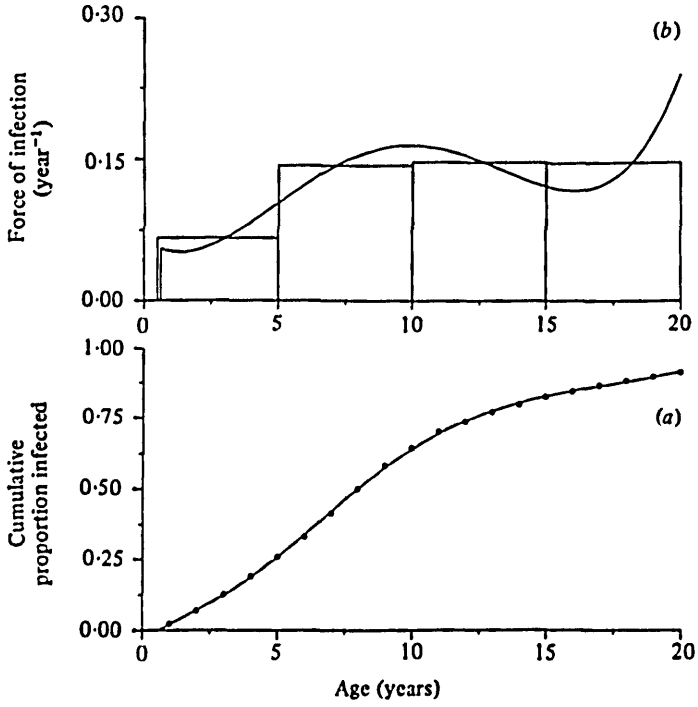

Fig. 2. Analysis of measles notifications for Maryland, USA (excluding Baltimore), 1908–17. (*a*) and (*b*) are as for Fig. 1.

(i) *Prior to mass vaccination.* (*a*) Measles notification for Maryland, USA, in the period 1908–17 (Fales, 1928), in Baltimore, USA, in the period 1906–15 (Hedrich, 1930). (*b*) Notifications for the city of Aberdeen, UK, in the period 1883–1902 (Wilson, 1904). (*c*) Age-stratified notifications for England and Wales in the period 1948–68 (data from the Annual Reviews of the Registrar General of England and Wales).

(ii) *Since the onset of mass vaccination.* (*d*) Age-stratified notifications for England and Wales over the period 1968–83 (data from the Annual Reviews of the Registrar General of England and Wales and the Statistics of Infectious Diseases published by the Office of Population Censuses and Surveys, (OPCS)).

Serological data is also analysed for small (fewer than three children) and large (three or more children) families, as recorded by Black (1959). The study of Black (1959) documented, by age, the proportion of people positive for measles antibodies, using a test for neutralizing antibodies on samples collected in New Haven, USA, during 1957, i.e. before mass vaccination.

Analyses of long- and short-term trends in measles infection rates are based on quarterly measles notifications for England and Wales in the period 1948–83 (Quarterly Reports of the Registrar General and the Quarterly Infectious Disease Monitors published by OPCS).

RESULTS

Pre-vaccination data

Figs. 1–4 display the results of applying the maximum-likelihood estimate procedure for the age-dependent force of infection to notifications from Baltimore

https://doi.org/10.1017/S0022172400062859 Published online by Cambridge University Press

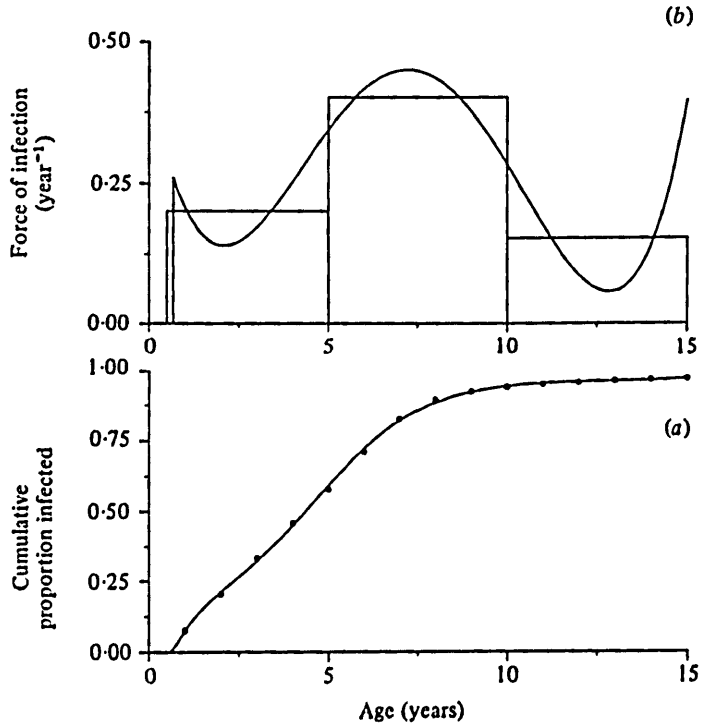

Fig. 3. Analysis of measles notifications for Aberdeen, 1886–1902. (*a*) and (*b*) are as for Fig. 1.

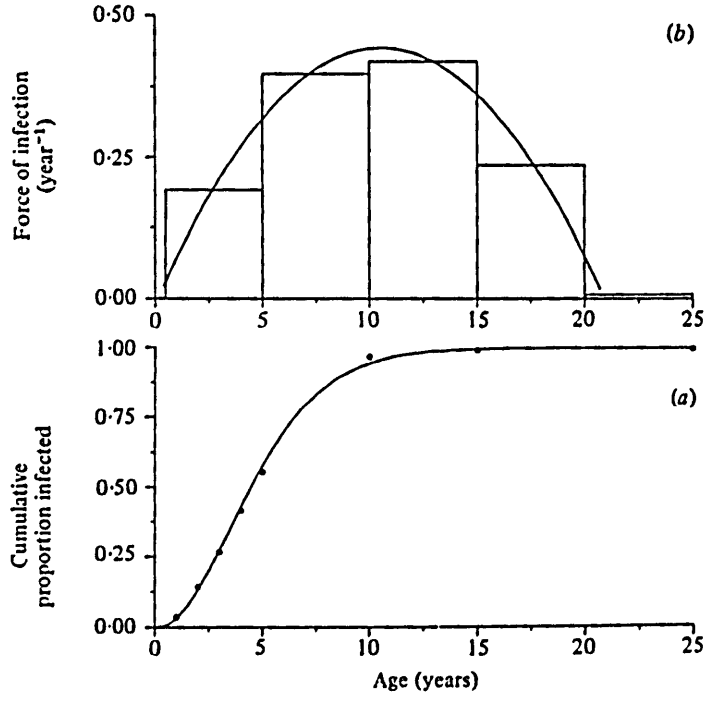

Fig. 4. Analysis of measles notifications for England and Wales, 1948–68. (*a*) and (*b*) are as for Fig. 1.

https://doi.org/10.1017/S0022172400062859 Published online by Cambridge University Press

Table 1. Details of the 'best fit' polynomial relationships between force of infection and age for the measles notification and serological data documented fully in the main text

Polynominal parameters (see equation (4)), with standard errors in parentheses, and an estimate of the mean age at infection (see Appendix) are given for each fit. The figures n and m refer respectively to the total number of individuals comprising each data-set analysed, and the size of the minimum age class used (which is always at the upper age limit, U). n and m are not available for the serological data sets.)

Data source	Upper age limit (U)	Poly-nomial degree	L	b_0	b_1	b_2	b_3	b_4	Mean age at attack A (years)
Pre-vaccination notifications									
Baltimore	25	3	0·0881	−0.00594	0·0679	−0·00561	−0·000122	—	6·72
1906–15			(0·046)	(0·0012)	(0·0009)	(0·00012)	(0·0000042)	—	
$n = 20998$, $m = 77$									
Rural Maryland	20	4	0·625	0·0663	−0·0228	0·0102	−0·000951	0·0000261	9·27
1908–17			(0·028)	(0·0055)	(0·0041)	(0·00089)	(0·000073)	(0·000002)	
$n = 29118$, $m = 456$									
Aberdeen	15	4	0·666	0·429	−0·325	0·113	−0·0124	0·00042	4·76
1883–1902			(0·012)	(0·017)	(0·015)	(0·0043)	(0·00046)	(0·000016)	
$n = 40374$, $m = 150$									
England and Wales	25	2	0·122	−0·0105	0·0864	−0·0411	—	—	4·96
1948–68			(0·0038)	(0·000073)	(0·000046)	(0·000003)	—	—	
$n = 8{\cdot}56$ million,									
$m = 41321$									
Pre-vaccination serology									
New Haven, small families	15	2	4·738	−1·475	0·411	−0·021	—	—	8·01
			(2·15)	(2·36)	(0·66)	(0·042)	—	—	
New Haven, large families	15	2	1·564	−0·261	0·186	−0·0125	—	—	5·51
			(13·7)	(0·6)	(0·29)	(0·025)	—	—	
Post-vaccination notifications									
England and Wales	25	2	0	0·013	0·0688	−0·00323	—	—	5·23
1969–82			—*	(0·00014)	(0·00009)	(0·000005)	—	—	
$n = 1{\cdot}83$ million,									
$m = 12613$									

* Because L is on its lower bound ($L = 0$) a quadratic surface cannot be fitted, so that S.E. (L) cannot be estimated.

(1906–15), rural Maryland (1908–17), Aberdeen (1883–1902) and England and Wales (1948–68), respectively. The results for the 1957 New Haven serological study are given in Fig. 5 (small family serology) and Fig. 6 (large family serology). Each figure records the observed and expected proportions who have experienced the disease by age a, $F(a)$, and the best fit polynomial force of infection, $\lambda(a)$. On the graphs portraying $\lambda(a)$, mean forces of infection over specified age intervals (0·5–5 years and thereafter in 5-year blocks) are recorded as histograms (the means were calculated from the polynomial fits). The parameter estimates for each fit are given, along with their standard errors and an estimate of the mean age at infection, A, in Table 1. The close correspondence between observed and expected age profiles ($F(a)$) in general illustrates the goodness of fit of the polynomial models, although in some cases the number of parameters (Table 1) is quite close to the number of age-classes used in their estimation (Figs. 1–7). A statistical test of agreement is inappropriate given the large numbers of notification records underlying each observed value of $F(a)$ (Griffiths, 1974).

The analyses of the notification data sets indicate a general and consistent trend for $\lambda(a)$ to increase from zero (between birth and approximately 1 year of age) up to a maximum value at around 10 years of age. This rise in the force of infection is followed by a decline in the older age-classes. The consistency of these patterns is best illustrated by the histograms in Figs. 1–4, which record mean values of λ over specified age classes. The estimates of the average age at infection, A, provide a simple one-parameter description of the overall pattern of infection within each sample population (Table 1). For the prevaccination data the value of A varies in the range 4·8–9·3 years. Much of the variation in A can be accounted for by differences in population densities and birth rates in the study areas.

The density (related to frequency of contact) and overall population size (determining the net birth rate of the community) are important determinants of the rate of disease transmission (the magnitudes of the $\lambda(a)$'s) and the average age of infection (A). In general, large urban communities experience higher rates of infection (and hence smaller A values) than low density rural communities (Godfrey, 1928; Griffiths, 1974; Anderson, 1982). This effect is illustrated by our analyses of the Baltimore city data ($A = 6·7$ years) and the whole state of Maryland, which includes large rural areas ($A = 9·3$ years). The corresponding force of infection polynomials (Figs. 1*b*, 2*b*) are strikingly different. Within the rural community, λ rises comparatively slowly during childhood and then remains high in the adolescent age-classes. By contrast, in the urban area λ rises steeply up to age 10 years, after which it falls away in the adolescent classes.

The results for Aberdeen show a lower average age at infection ($A = 4·8$ years; Fig. 3, Table 1). This is primarily a consequence of high numbers of cases in the first two age-classes (very young children). Adjustments of the data to reflect cases per 1000 head of population within an age class do not alter this pattern. The low value of A is probably a reflexion of high population density and the high birth-rate prevailing in the city during the period 1883–1902.

Analyses of the prevaccination data from England and Wales (1948–68) reveal a mean age at infection of 5·0 years; a value which is in broad agreement with the cruder estimates published by Griffiths (1974) and Anderson & May (1982). A quadratic fit for the force of infection captures the general trend in the notification

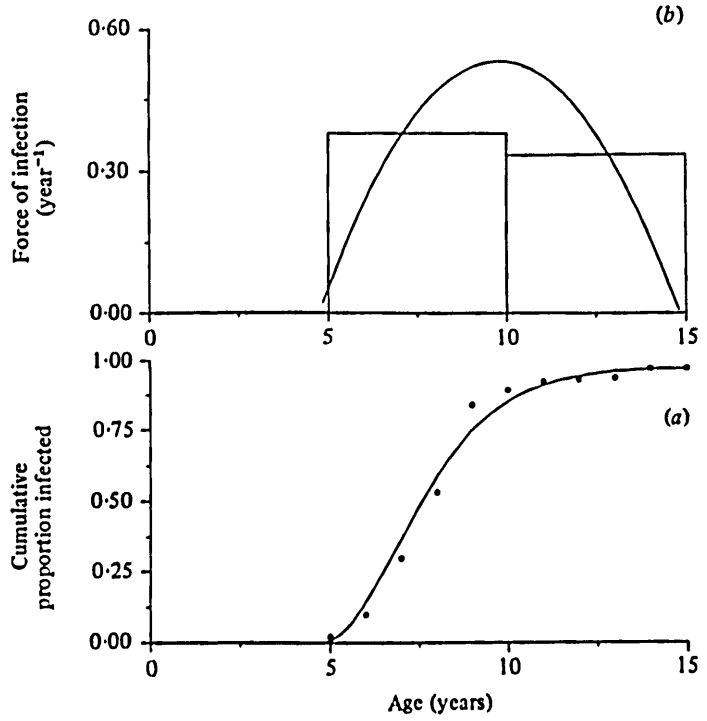

Fig. 5. Analysis of measles serological for data from small families (fewer than three children) for New Haven, USA, 1957. (*a*) and (*b*) are as for Fig. 1.

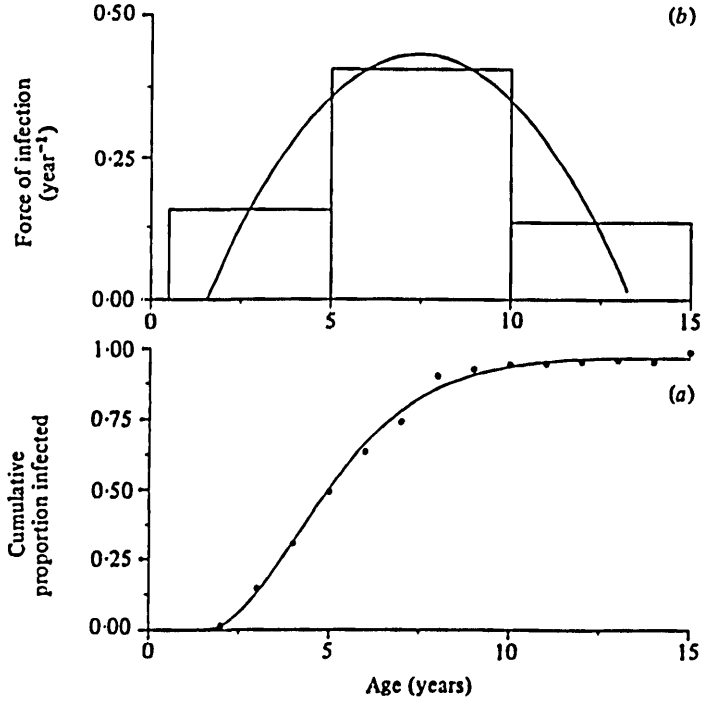

Fig. 6. Analysis of measles serological for data from large families (more than three children) for New Haven, USA, 1957. (*a*) and (*b*) are as for Fig. 1.

https://doi.org/10.1017/S0022172400062859 Published online by Cambridge University Press

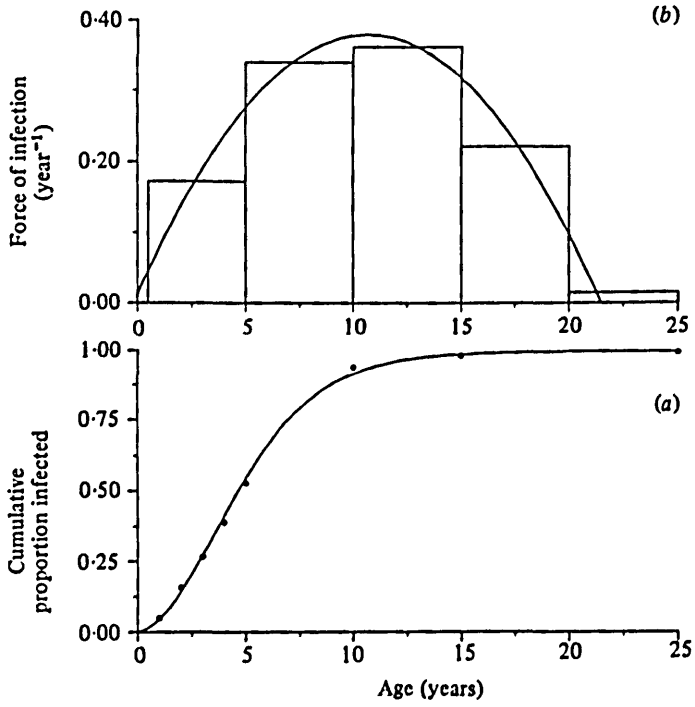

Fig. 7. Analysis of measles notifications for England and Wales, 1969–82. (*a*) and (*b*) are as for Fig. 1.

data although the estimates of λ in the older age-classes are approximate given the crude blocking of the notification above the age of 5 years (5 or more year blocks). The quadratic fit shows a rise in λ up to around age 10 years followed by a decline in the older age-classes (Fig. 4). The Fine & Clarkson (1982*b*) analyses of a more finely age-stratified data set from London revealed a decline in the force of infection after about 7–8 years of age. The crude age-blocking of notification for England and Wales precludes us from exploring the finer details of age-related changes in rates of infection amongst teenage and adult age classes. It is probable that adults in the range 20–35 years of age have slightly higher rates of infection than older adult groups as a consequence of frequent contact with young children within a family setting.

The analysis of Black's serological data from New Haven again shows a pattern of increase in the value of λ up to around 10 years of age, and a decline thereafter in the older age-classes (Black, 1959; Figs. 5 and 6). The importance of family size is well illustrated in Figs. 5 and 6. Pre-school children from small families, in Black's study, appear to have had little contact with measles before attending school at around 5 years of age. By contrast, within the large family group, the proportion seropositive rises steadily over the age range 1–5 years. The difference is well reflected by the respective average ages at infection, which are $A = 8{\cdot}0$ years for small families and $A = 5{\cdot}5$ years for large families. The pronounced difference should, however, be interpreted cautiously, since the small family samples were not chosen at random from the population (Black, 1959).

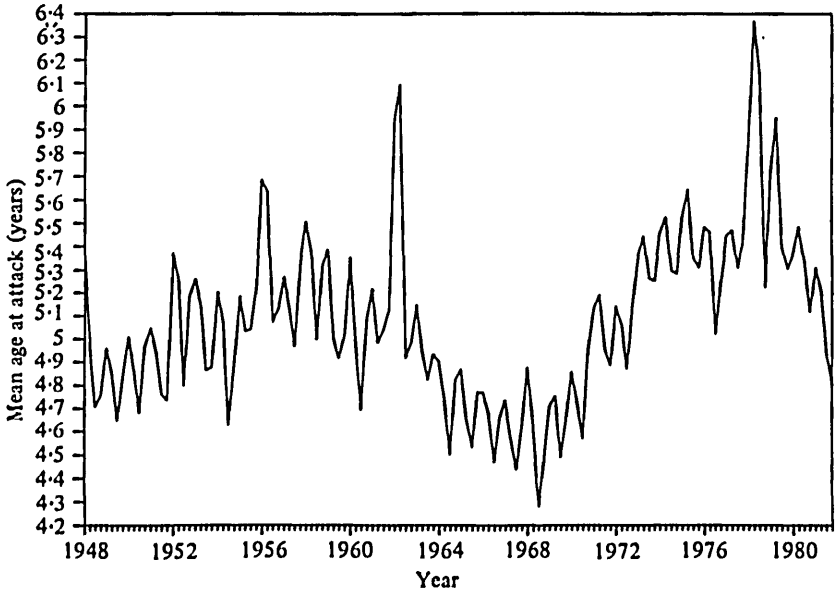

Fig. 8. Estimated mean age at attack by quarter for measles in England and Wales in the period 1948–82; based on quarterly measles notifications.

Post-vaccination data

The results of the analyses of the case notifications for measles in England and Wales over the period 1969–82 are summarized in Fig. 7 and Table 1. They reveal a similar pattern to that observed in the pre-vaccination era, with the peak force of infection occurring around 10 years of age. As predicted by theoretical studies, mass vaccination acts to raise the average age at infection over that pertaining prior to control (Anderson & May, 1982, 1983, 1985; Fine & Clarkson, 1982*b*; Griffiths, 1974); the average value of A is 5·0 years in the pre-vaccination era and 5·3 years over the post-vaccination period. We estimate that under a 50–60% vaccination coverage of each cohort of children over the period 1969–82, control has decreased the force of infection in the age-classes 6 months to 5 years from 0·19 to 0·17 year^{-1}. The data are too crudely blocked by age to permit the estimation of other than a very simple polynomial function (quadratic) for (a) (Fig. 7).

Long-term trends and short-term periodicities in the average age at infection (A)

Fig. 8 records changes in the estimated mean age at infection for measles in England and Wales, by quarter-year periods for the years 1948–82. The values of A were derived by fitting a quadratic force of infection polynomial to the measles notification by age for each quarter year. The estimates reveal interesting long- and short-term trends.

The estimated average age at infection has risen from around 4·8 years in 1948 to 5·1 years in 1960. Following this period, A falls to a value of around 4·6 years in 1968 and then rises to attain a plateau around 5·5 years in the 1980s. The comparatively low value of A in the late 1940s is probably a reflexion of the post-war 'baby boom', although a change of practice in notification recording may also be

https://doi.org/10.1017/S0022172400062859 Published online by Cambridge University Press

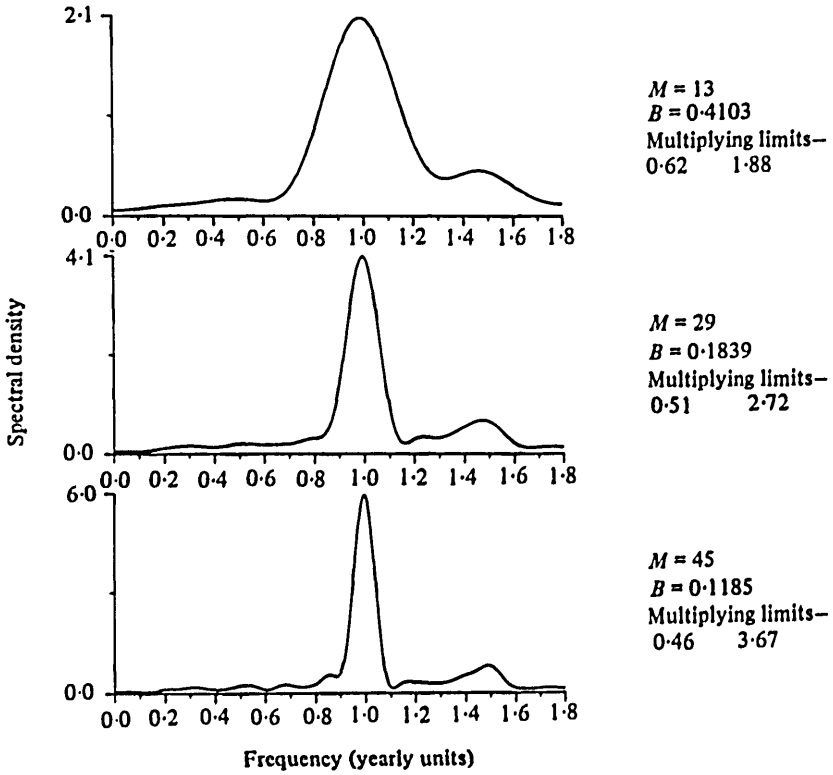

Fig. 9. Spectral analysis for the time series of estimated mean ages at attack of measles in England and Wales (1948–82), shown in Fig. 8. The data were logged and de-trended by differencing before spectral estimation. Full details of the spectral analysis are given by Anderson, Grenfell & May (1984). We smooth the spectrum with a Tukey smoothing window, and present the results at a range of window cut-off points (M) and therefore bandwidths (B cycles per year; B measures the degree of smoothing of the window). Multiplicative 95% confidence limits for the spectra are also given. The sharp peak at a frequency of 1 year indicates a pronounced annual oscillation in the series, which is significantly non-random ($P < 10^{-4}$, based on the equivalent auto-correlogram).

of importance. The reduction in A during the early 1960s is a well-documented phenomenon (Griffiths, 1974; Fine & Clarkson, 1982*b*; Anderson & May, 1982) and is thought to be due to a variety of factors, including an increased birth-rate in the late 1950s and early 1960s, changes in notification practice (Fine & Clarkson, 1982*b*) or changes in school attendance patterns, with greater numbers of children attending play-schools prior to entry to primary school. The increase in A from 1968 onwards reflects the impact of mass vaccination. It is interesting to note that A attains a plateau in the mid- to late-1970s and 1980s since vaccination coverage in England and Wales has remained relatively constant over this period at around 50–60% of each birth cohort (Anderson & May, 1982).

Fig. 8 also records a marked annual periodicity in the value of A. Time-series analysis techniques reveal a highly significant correlation at a period of 1 year. Fig. 9 records a spectral analysis of the data displayed in Fig. 8 (see Anderson, Grenfell & May (1984) for further details). The raw values of A were log-transformed

https://doi.org/10.1017/S0022172400062859 Published online by Cambridge University Press

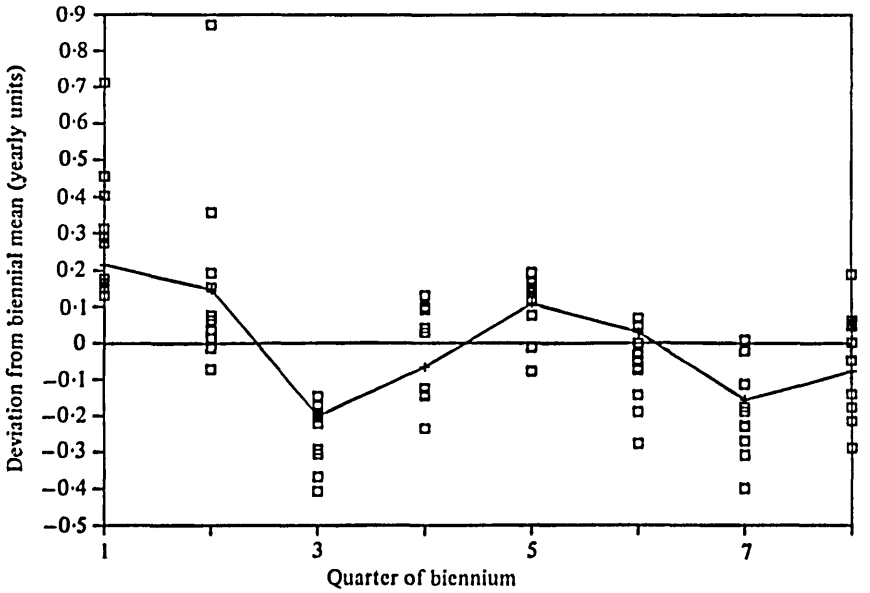

Fig. 10. Biennial variation in the estimated mean ages at attack for measles in England and Wales (1948–82), shown in Fig. 8. Long-term trends are removed by subtracting from each estimate the mean for the biennium in which it lies. +, Mean deviations.

and differenced in order to stabilize the variance and remove the long-term trends (Chatfield, 1975). Fig. 9 records the highly statistically significant peak (and therefore periodicity) at a frequency of one cycle per year with very little variation at other frequencies. Note that the 2-year epidemic cycles of measles do not have a marked impact on A since the statistic is calculated from information recording cases over age-classes in which individuals will have experienced a number of epidemic and non-epidemic years.

The annual pattern of changes in A is explored in more detail in Fig. 10, which records variation over the eight quarterly segments of the two-year measles cycles over the period 1948–82 (giving 17 biennial periods). Quarters 1–4 record the values of A in inter-epidemic years and quarters 5–8 the values of A in the epidemic years. Long-term changes in A were corrected for by subtracting from each quarterly estimate the mean value of A over the biennium in which the quarter-year falls. Two points emerge from this analysis. Firstly, A is on average lower in the epidemic years, and secondly, in both epidemic and inter-epidemic years the estimated average age at infection is lowest in the third quarter (the summer and early autumn months). This latter trend is at first sight puzzling since it is widely assumed that contact between the children (and hence the force of infection) is at its lowest during the summer months. Indeed, the number of reported cases is invariably at an annual minimum in these months (London & Yorke, 1973; Fine & Clarkson, 1982*a*; Anderson, Grenfell & May, 1984). Thus, paradoxically, our estimates of the average age at infection are lowest (implying high forces of infection according to the catalytic model) in the very months when case reports are at their lowest ebb, and where in reality transmission is minimal. The pattern is a simple consequence of the fact that, during the third quarter of each year, the

https://doi.org/10.1017/S0022172400062859 Published online by Cambridge University Press

proportion of measles notification in groups above 5 years of age is at its lowest level with respect to the other quarters of the year. This effect is probably a consequence of the summer vacation period, when children above 5 years of age mix less intensively with other children of their age than during their time in school. As such, of the few cases reported, most occur in very young children. This results in a low average age at infection despite the fact that the overall rate of transmission within the child and teenage community is at its minimum seasonal level.

This example reveals a problem in the use of horizontal (as opposed to longitudinal) age-stratified case-notification records to estimate age-specific forces of infection. From horizontal data, the magnitude of $\lambda(a)$ within an age-class is calculated on the assumption that all individuals eventually acquire infection, so that the total number of reported cases reflects the total number of susceptibles who will at some age acquire the disease. This assumption breaks down if the numbers of reported cases are unevenly distributed throughout a year (or a 2-year epidemic cycle) and if horizontal records are used from a small segment of that period (such as a quarter of a year). The bias is minimized if we employ age-stratified reports collected over 1- or 2-year (or longer) periods. It is completely eliminated, of course, if serological data (as opposed to case notification records) is employed to estimate $\lambda(a)$.

In spite of the problem discussed above, our analyses of changes in A within a year confirm the importance of seasonal fluctuations in the force of infection as a determinant of measles incidence. A better evaluation of the magnitudes of seasonal change in the age-specific forces of infection could be derived from age-stratified serological data collected at frequent intervals throughout a year. Alternatively, longitudinal cohort studies of quarterly case-notifications, corrected for age and seasonal biases in reporting efficiency (difficult to estimate in practice), could also provide valuable information as to the true extent of seasonal changes in the force of infection.

CONCLUSIONS

The most striking feature of our analyses is the consistency of the patterns of change with age in the force of infection for measles, irrespective of whether the data was collected in the UK or USA, whether it consists of case-notification records or serology, or whether it represents the pre- or post-vaccination eras. The rate of infection rises to a peak at around 10 years of age and declines thereafter. These patterns probably reflect age-related changes in the degree of mixing and contact within- and between-age classes. School attendance is undoubtedly an important component in the generation of such trends. Whether the observed decline in the older age-classes is real or an artifact arising from notification biases, the inability of serological tests to detect low antibody titres in adults, or genetic or spatial heterogeneities in exposure to infection, is uncertain at present. These and related issues, concerned with the implications of age-dependent rates of infection for the design of vaccination programmes, are discussed elsewhere (Schenzle, 1985; Anderson & May, 1985).

Our primary concern in this paper was to develop methods for the estimation of age-related trends and to examine their limitations. The methods are most

robust if applied to serological data, since the model is based on a description of age-related changes in the cumulative proportion in a population who have experienced infection. Ideally, the data should reflect longitudinal changes in the proportion seropositive, as specified cohorts age within the population. To our knowledge, no such serological data is available at present, covering an age-range of, say, 0–20 years. This is understandable given the long duration of study required to collect such information. More serious, however, is the scarcity of horizontal age-stratified serological profiles. In the UK, for example, no such survey has been published for the pre-vaccination era. The interpretation of post-vaccination surveys is complicated, but estimates of the age-specific forces of infection can be obtained if information on age-related vaccination rates is available. There is an urgent need to collect such serological data (by surveys with fine age-stratification), given the importance of age-related changes in the force of infection to the design of vaccination programmes and the interpretation of epidemiological trends in disease incidence under the impact of control (Anderson & May, 1983, 1985). In addition, useful insights into the mechanisms that generate low apparent rates of infection in adult age-classes could be acquired if serological surveys were designed to record information on antibody titres and certain genetic characteristics of the sample population (e.g. HLA typing).

Finally, although our analyses are based on data for measles infection, the methods outlined in this paper can be applied to the epidemiological study of a wide variety of viral and bacterial infections, provided they are directly transmitted by contact between individuals and they induce life-long immunity to reinfection.

We gratefully acknowledge financial support for this research from the Department of Health and Social Security.

REFERENCES

Anderson, R. M. (1982). Directly transmitted viral and bacterial infections of man. In *Population Dynamics of Infectious Disease Agents: Theory and Applications* (ed. R. M. Anderson), pp. 1–37. London: Chapman and Hall.

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

Anderson, R. M. & May, R. M. (1982). Directly transmitted infectious diseases: control by vaccination. *Science* **215**, 1053–1060.

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

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

Black, F. L. (1959). Measles antibodies in the population of New Haven, Connecticut. *Journal of Immunology* **83**, 74–83.

Chatfield, C. (1975). *The Analysis of Time Series: Theory and Practice.* London: Chapman and Hall.

Collins, S. D. (1924). Past incidence of certain communicable diseases common among children. *Public Health Reports* **36**, 1553–1567.

Collins, S. D. (1929). Age incidence of the common communicable diseases of children. *Public Health Reports* **44**, 763–820.

Fales, W. T. (1928). The age distribution of whooping cough, measles, chicken pox, scarlet fever and diphtheria in various areas in the United States. *American Journal of Hygiene* **8**, 759–799.

FINE, P. E. M. & CLARKSON, J. A. (1982*a*). Measles in England and Wales. I. An analysis of the factors underlying seasonal patterns. *International Journal of Epidemiology* **11**, 5–14.

FINE, P. E. M. & CLARKSON, J. A. (1982*b*). Measles in England and Wales. II. The impact of the measles vaccination programme on the distribution of immunity in the population. *International Journal of Epidemiology* **11**, 15–29.

GODFREY, E. S. (1928). The age distribution of communicable diseases according to size of community. *American Journal of Public Health* **18**, 616–631.

GRIFFITHS, D. A. (1974). A catalytic model of infection for measles. *Applied Statistics* **23**, 330–339.

HEDRICH, A. W. (1930). The corrected average attack rate from measles among city children. *American Journal of Hygiene* **11**, 576–600.

LONDON, W. P. & YORKE, J. A. (1973). Recurrent outbreaks of measles, chicken pox and mumps. I. Seasonal variation in contact rates. *American Journal of Epidemiology* **98**, 453–468.

MUENCH, H. (1959). *Catalytic Models in Epidemiology.* Harvard University Press.

NELDER, J. A. & WEDDERBURN, R. W. M. (1972). Generalised linear models. *Journal of the Royal Statistical Society* A **135**, 370–384.

SCHENZLE, D. (1985). Control of virus transmission in age-structured populations. *Lecture Notes in Biomathematics.* (In the Press.)

SYDENSTRICKER, E. (1928). The incidence of various diseases according to age. *Public Health Reports* **43**, 1124–1156.

WILSON, G. N. (1904). Measles: its prevalence and mortality in Aberdeen. *Report of the Medical Officer of Health, Aberdeen*, pp. 41–51.