PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Croonian Lecture, 1994: Populations, Infectious Disease and Immunity: A Very Nonlinear World

R. M. Anderson

Phil. Trans. R. Soc. Lond. B 1994 **346**, 457-505
doi: 10.1098/rstb.1994.0162

The Croonian Lecture, 1994. Populations, infectious disease and immunity: a very nonlinear world

R. M. ANDERSON

Department of Zoology, University of Oxford, Oxford OX1 3PE, U.K.

CONTENTS

	PAGE
1. Introduction	457
2. Childhood infections, herd immunity and mass vaccination	460
(*a*) High and low efficacy vaccines	461
(*b*) Mass immunization and herd immunity	463
3. Heterogeneity in exposure to infection	466
(*a*) Mixing patterns between sexual activity classes	467
(*b*) Sex acts and sex partners	468
4. The population dynamics of the immune system	470
(*a*) Regulation of cell proliferation	470
(*b*) Malaria and the human immune system	473
(*c*) Antigenic variation within the host	475
5. Pathogen genetic variation and between host transmission	479
(*a*) Superinfection and virulence	479
(*b*) The transmission dynamics of multiple strains of HIV	481
(*c*) Antigenic variation and the transmission dynamics of malaria	484
6. Infectious disease and human population growth	488
(*a*) Respiratory and gastro-intestinal infection that act to increase mortality	489
(*b*) Sexually transmitted infections that influence fertility	489
(*c*) The demographic impact of AIDS	491
7. Discussion	496
References	498

SUMMARY

The interaction between the variables that determine the typical course of infection in an individual patient and those that determine transmission in communities of people is often complex and very nonlinear in form. Mathematical models of infection and immunity are used to study the interaction in a wide variety of problems including the role of antigenic variation in pathogen persistence in the host, the design of vaccination policies for the control of childhood viral infections, the role of heterogeneity in sexual behaviour as a determinant of the epidemiology of sexually transmitted diseases and the demographic impact of infectious disease on human population growth. The themes of dynamical complexity in outcome deriving from simple biological assumptions, the evolution of the parasite under selection by the immune system, heterogeneities in the interacting systems, and the necessity of comparing prediction with observation reoccur in each problem. It is argued that much is to be gained from the use of mathematics in biology, concomitant with experiment and observation, in providing precision in interpretation and in facilitating the formulation and testing of hypotheses to explain observed pattern. Special emphasis is placed on the need for interdisciplinary research on the epidemiology of infectious diseases that combines molecular, immunological, field study and theoretical approaches.

1. INTRODUCTION

Throughout human history epidemics of infectious diseases have been a major cause of mortality and, concomitantly, a major force of natural selection in the evolution of our species (McNeill 1977; Hill *et al.* 1991). A dramatic example of the strength of this selective pressure is the 14th century epidemic of plague (the Black Death) which caused the death of

Phil. Trans. R. Soc. Lond. B (1994) **346**, 457–505
Printed in Great Britain

between one-third to one half of the population in Europe. Most of the major episodes documented by historians are associated either with rapid population growth in urban settings as a consequence of stable governance and growth in average wealth, or with the arrival of immigrants or invaders from other countries and continents who introduced new infections to immunologically naive communities. For example, it is estimated that 90% of the population of Native Americans died as a result of the European conquest of the New World. Many died in combat, but most succumbed as a result of the introduction of new infectious diseases such as tuberculosis, scarlet fever, smallpox, measles, syphilis and typhoid (Black 1993).

In more recent times the pattern of infection and disease in the world has changed. Three factors have been of particular importance. First, rapid population growth and increased urbanization (leading to high density populations) has facilitated the spread and persistence of many directly transmitted respiratory and gastro-intestinal infections, especially in poor communities without adequate resources to maintain good standards of hygiene and sanitation. Second, greatly improved transport systems, both within and between countries, have acted to dramatically change patterns of global mixing such that new strains of an infectious agent that emerge in one area are rapidly spread across countries and continents. This change has been most pronounced in very recent times due to growth in air transportation. Last, there has been considerable improvements in the diagnosis, treatment and prevention of many common infections over the past few decades due to the development and widespread use of drugs and vaccines (World Bank 1993).

In the industrialized countries, the latter factor induced a degree of complacency about infection and disease with research on, and teaching about, infectious diseases declining in the majority of medical schools. This trend was in part understandable. In urban centres in developed countries in the 1930s hospital wards were filled with victims of infectious disease such as pneumonia, tuberculosis, typhoid fever and encephalitis induced by a variety of viral infections. However, by the early 1980s, a visitor would have found the same wards dominated by patients with non-infectious problems such as heart disease, strokes and various types of cancer. This change was brought about by the development and use of effective antimicrobial agents, the community wide use of vaccines and the implementation of effective infectious disease surveillance and public health prevention programmes. This shift in the balance of morbidity and mortality caused by infectious and non-infectious disease generated the belief that the era of microbial diseases as a threat to human communities had passed. Even the continued toll of mortality and morbidity in the developing world, where infectious diseases have remained the leading cause of premature mortality throughout the 20th century, was viewed as a problem of resource allocation and distribution, and not one that science and technology could not solve given sufficient monies for research and drug plus vaccine distribution.

In the past decade, a series of events have acted to perturb our complacency about infection. The emergence of AIDS as a worldwide pandemic, and the difficulties encountered in the development of effective drugs and vaccines against a pathogen with enormous genetic variability both within and between infected individuals, have been particularly influential. However, many other factors have also been of significance. These include the continual emergence of new infectious agents across a very broad spectrum of pathogen types in a rapidly expanding global population with greatly increased patterns of mixing between societies (Lederberg *et al.* 1992; Morse 1993; Krause 1992). Aside from the immunodeficiency viruses, examples are the newly recognized hanti virus in the western United States in 1993 (CDC 1993), the hepatitis C and D viruses and Lyme disease (Barbour & Fish 1993). These 'emerging diseases' are either truly new or are rapidly increasing their incidence or expanding their geographic range (Morse & Schulenberg 1990; Morse 1993). Another factor of great importance has been the difficulties encountered in the past decade in keeping one step ahead of the rapid evolution of drug resistant bacterial strains in hospital settings by the development of yet more antimicrobial agents. The introduction of new drugs is invariably accompanied by the development of resistance. But a particularly worrying trend has been the emergence of multiple drug resistant strains of some bacteria and protozoa, combined with a decline in the rate at which new antimicrobial drugs are being developed. Without a continued investment of resources for basic research on microbes and drug development, these recent trends raise serious concern over the continued general effectiveness of antimicrobial agents, particularly in large hospital settings in urban areas. Surveillance, outbreak investigation and prospective studies testify to the rapidity with which these new strains spread from one location to another. An example is provided by the resurgence of tuberculosis in some large urban centres in certain developed countries due to multiple drug resistance (Bloom & Murray 1992). In part this is due to the persistence of infection (lengthened infectious period) in immunocompromised hosts such as AIDS patients, but it reminds us that even the most effective drugs will not necessarily eliminate infection without the help of a fully functioning immune system.

Heightened awareness of these problems is due, in part, to the development of new diagnostic and descriptive methods made possible by advances in the molecular, cellular and immunological research fields of biology. Such techniques include monoclonal antibodies, the polymerase chain reaction (PCR) and quantitative derivations of the technique, DNA (or RNA) signal amplification assays in general, DNA hetroduplex mobility assays, DNA fingerprinting, and various techniques of immune system cell activation and quantification. They are beginning to provide the necessary tools to chart with precision, and in quantitative terms, the course of infection in

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

individual patients, the genetic diversity of pathogen strains and the pattern of infection and disease in communities of people.

In combination with other techniques they provide very exciting opportunities for epidemiological research in the coming decades. In particular, they have already begun to shift the emphasis towards a more evolutionary approach in the study of the interaction between pathogen and host. Evolution is at the core of the relationship between the two. In the search for new drugs and vaccines, or in the study of immunological responses to infection, we often forget the enormous potential of viruses, bacteria and protozoa to evolve rapidly (either via mutation or recombinant events) on timescales very much shorter than the generation time of their human host, or indeed, in some cases, the time required to mount an effective immunological response. As the intensity of drug or vaccine application (or of immunological attack within the host) increases, the selective pressure for change is raised. Genetic variation via antigenic change is often a central pillar of the pathogen's strategy to persist in the host in the face of immunological attack. Molecular techniques that measure genetic diversity are beginning to reveal how common the strategy is, particularly amongst the persistent infections (i.e. Hepatitis B, HIV-1, *Mycobacterium tuberculosis*) but also in some more transient ones (i.e. *Niesseria gonorrhoea*). One spin off of the very rapidly expanding data base on the antigenetic diversity of important human pathogens (sequence data bases), has been the development of statistical techniques to look at relatedness between different isolates via the construction of phylogenetic trees (Nee *et al.* 1994). Precision in such techniques is improving and it is possible in some cases to infer who acquired infection from whom via sequence information from pathogen isolates from different patients. Retracing the history of events is not easy, and many assumptions underpin the construction of phylogenetic trees of relatedness between isolates. An exciting new trend in research, however, is the inclusion of an evolutionary framework within such methodologies, which tries to consider how the biology of the infectious agent and its interaction with its host will influence the intensity and direction of selection for genetic change (Harvey *et al.* 1994). These techniques also offer new insights into epidemiological processes, since the shape of a phylogenetic tree and rates of diversification in the genome of the infectious agent can reflect the temporal state of an epidemic or, indeed, whether an epidemic or endemic situation pertains (Harvey *et al.* 1994).

In parallel with the advances in molecular biology that have focused on genetic diversity in pathogen populations and the functional complexity of the human immune system, mathematical methods have increasingly been applied in the study of the persistence, spread and control of infectious agents (Anderson & May 1991). This rapidly expanding body of research uses mathematics to provide precision in the formulation and testing of hypothesis concerning the factors responsible for observed pattern. It tries to delineate the interplay between the factors that determine the typical course of infection in a patient and those that determine transmission between patients. At both levels of study (within and between hosts) patterns of infection, disease and immunity are typically complex and relationships between variables are invariably nonlinear in form. Even very simple model systems can exhibit cyclic oscillations or indeed chaos (Schaffer & Knot 1985; Anderson & May 1991; Bolker 1993; Bolker & Grenfell 1993). A particularly encouraging trend in this field has been the beginnings of convergence between theory and observation, where the former is prevented from soaring free from the constraints of data, while the interpretation of the latter is grounded on a firm theoretical (= mathematical) framework. Most recently, a further stimulus in melding the theoretical and experimental fields has been the growing recognition of the great dynamical complexity of the immune system and an enhanced appreciation of the need for careful study of the dynamics of evolutionary processes to facilitate interpretation of the rapidly growing volume of sequence information on pathogen diversity.

This present paper reviews recent trends in the field of infectious disease epidemiology, from the perspective of a population biologist. A particular emphasis is placed on the insights provided by mathematical studies of the interplay between the factors that determines the course of infection within the host and those that determine transmission between hosts. The review is selective, focusing on areas of current excitement and activity, and is organized as follows. The first section examines the transmission dynamics of some of the vaccine preventable childhood viral infections that induce life-long immunity and that do not appear to exhibit great genetic diversity. This is an area where theory has provided important insights into the interpretation of observed pattern and where mathematical models are used in practice as predictive tools to assess the merits and disadvantages of different immunization policies. The following section examines some of the complexities introduced by heterogeneity in exposure to infection arising from variable patterns of human behaviour. These are most apparent in the spread and persistence of sexually transmitted infections and the section focuses on how various forms of heterogeneity in sexual behaviour influence observed pattern.

The third section moves down a level of organization to consider the dynamic interaction between the human immune system and infectious agent populations within the host. Antigenical homogeneous and heterogeneous populations of the pathogen are examined. A step up from this, to examine the transmission of more than one strain of a particular pathogen, is considered in the fourth section, with an emphasis on the interaction between strain specific and cross-reactive immunological responses and the evolution of virulence. The sixth section addresses the interplay between epidemiology and demography and examines the impact of infectious agents on human

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

population growth. Particular reference is made to the current AIDS pandemic. The paper ends with a brief discussion of current trends and future directions in infectious disease epidemiological research.

2. CHILDHOOD INFECTIONS, HERD IMMUNITY AND MASS VACCINATION

One of the major success stories of modern preventative medicine is the development and wide scale use of vaccines to protect against a variety of directly transmitted viral and bacterial infections that were a major cause in children of morbidity in the developed world and of morbidity and mortality in the developing world. Examples include the measles, mumps, rubella, polio, diphtheria, pertussis and tuberculosis (TB), immunization campaigns (Nokes & Cutts 1993; Patriarca *et al.* 1993).

Theoreticians have also focused their attention on some of these infections because of the comparative simplicity of the processes that influence their transmission dynamics, the existence of long runs of data over many decades and also as a consequence of the remarkable periodicity of the incidences of some of the common viral infections of children such as measles, mumps and rubella (Brownlee 1906; Kermack & McKendrick 1927; Soper 1929; Farr 1840; Bartlett 1957; Bailey 1975; Anderson & May 1982*b*, 1991; Dietz & Schenzle 1985) (figure 1). For example, in many urban centres in western societies prior to the introduction of wide scale immunization, measles exhibited a major biennial cycle with a minor seasonal pattern.

A cornerstone of modern theoretical epidemiology is the threshold theorem according to which the introduction of a few infectious individuals into a community of susceptibles will not give rise to an epidemic outbreak unless the density or number of susceptibles is above a critical value (Kermack & McKendrick 1927). A corollary of this theory leads to a criteria for the critical level of mass immunization, p, required to block transmission within a defined community (Dietz 1976; Anderson & May 1982*b*, 1991). Central to this notion is the concept of a basic reproductive rate or number, R_0, which defines the average number of secondary cases of infection generated by one primary case in a susceptible population (Macdonald 1957; Dietz 1976; Anderson & May 1991). The magnitude of R_0 is determined by a number of parameters that describe the typical course of infection in an individual (the average incubation and infectious periods) and those that determine transmission within the population (the probability of transmission and the density of susceptibles). A growing literature is focused on the measurement of R_0 via cross-sectional serological studies of the fraction of individuals exposed to infection at different ages (Muench 1959; Griffiths 1974; Anderson & May 1982; Grenfell & Anderson 1985; Nokes *et al.* 1986). The development of new serological tests based on the detection of antibodies specific to viral antigens in saliva raises exciting opportunities to extend surveillance of prevailing levels of herd immunity (either due to vaccination or recovery from infection) in both developed and developing countries (Perry *et al.* 1993) (figure 2).

At present, this body of research is developing in two separate directions. On the one hand, work is

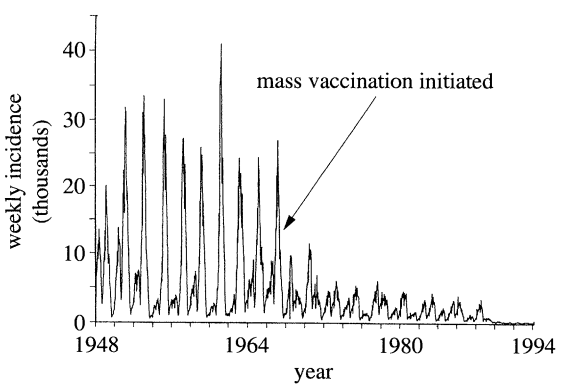

Figure 1. Longitudinal trends in the weekly incidence of measles in England and Wales, 1948 to 1994. (Data from the Office of Population Census and Survey, 1994.)

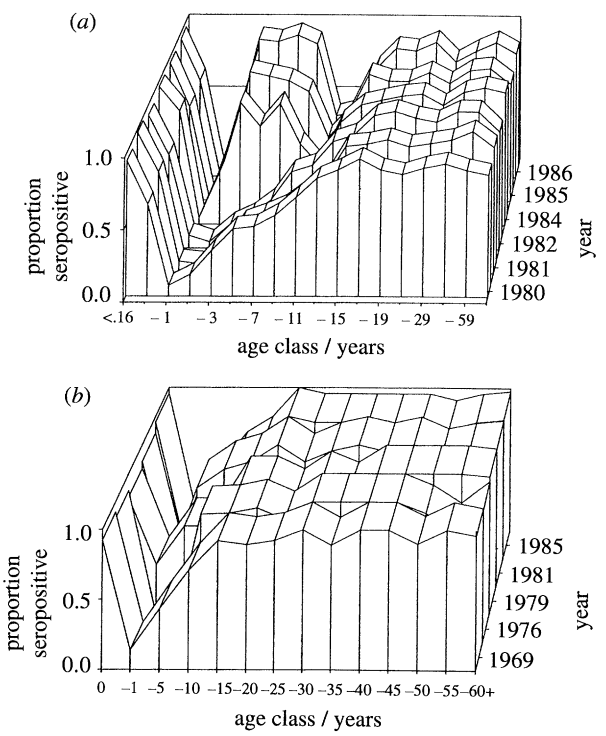

Figure 2. Impacts of two different immunization programmes on herd immunity to rubella virus (proportion of the population with antibodies to the virus, stratified by age and year of observation) (from Anderson & May 1990). (*a*) Finland in the period 1980 to October 1986 (Ukkonen & von Bondsdorff 1988). From 1980 to October 1982 vaccination was targeted at 13-year-old girls but in November 1982, MMR (measles, mumps and rubella) vaccine was administered to boys and girls of 14–18 months and 6 years of age. (*b*) Cross-sectional profiles for males and females combined in South Yorkshire 1969–1985 (Nokes & Anderson 1988). Rubella immunization of teenage girls was introduced in 1980; the profiles are similar in pre- and post-vaccination years.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

continuing on the dynamical properties of simple compartmental models that mirror the passage of individuals via a class of susceptibles, to a class of infecteds but non-infectious individuals, to a class of infectious people and finally to a class of immune individuals (the so-called SEIR models). Of particular interest is the oscillatory dynamics of even the simplest model (noted originally by Soper 1929), which predicts damped oscillation over long periods of time. Interestingly the predicted periods of these oscillations, with appropriate parameter assignments for different childhood infections, well mirror observed trends (e.g. the two year period of measles in unvaccinated urban centres in Europe and North America, Anderson & May 1982*b*).

The cycles are generated by the exhaustion by infection, and replenishment by new births, of the supply of susceptibles. They are most pronounced for viral and bacterial infection with short infectious and latent periods (a few days to a few weeks), with no carrier state and which induce lasting immunity to reinfection in those who recover. The interepidemic period may be significantly influenced by differing birth rates in different geographical locations (Anderson & May 1991). These damped cycles may be perpetuated by seasonal changes in the rate of transmission, $\beta(t)$, resulting from the aggregation and disaggregation of children for school terms and holiday periods. The seasonally forced SEIR model generates a rich array of dynamical behaviours including chaotic patterns (Smith 1982; Schwartz & Smith 1983; Aron & Schwartz 1984). Much recent discussion has focused on whether observed patterns, particularly the recorded temporal trends in the incidence of measles in developed countries, support this prediction (Schaffer & Kot 1985; Olsen & Schaffer 1990; Sugihara *et al.* 1990; Tidd *et al.* 1993). The case for chaotic behaviour in the incidence of measles over time rests on the evidence that historical epidemics show various 'fieldmarks' of chaos, such as positive Liapunov exponents, and that phase portraits reconstructed from observed time series (case records) bear a striking resemblance to the chaotic solution generated by the seasonally forced SEIR model (Bolker & Grenfell 1993; Tidd *et al.* 1993) (figure 3).

As further complexity is added to these models to mirror age structure, age dependency in the seasonal forcing term and patterns of 'who acquires infection from whom', the complex dynamical properties tend to be lost. In a recent paper Bolker (1993) documents a detailed numerical study of the sequential addition of biological complexity to the basic SEIR framework and an alternative framework called the RAS (realistic age-structured) model (Schenzle 1984). In the case of parameter settings appropriate to measles the family of models show a range of behaviours from large-amplitude chaos to regular biennial patterns. However, most importantly, added biological complexity (and heterogeneities) generates a buffering effect that prevents unrealistic incidences of infection in inter-epidemic periods which tend to reduce or eliminate the likelihood of chaotic behaviour (Bolker 1993; Bolker & Grenfell 1993). The lesson to be learned from this important body of theory is that biological complexity, and heterogeneities introduced by age or spatial factors, matter a great deal to the interpretation of observed pattern and to an understanding of the factors that enhance the long term persistence of infection in defined communities.

The second thrust of mathematical epidemiological research on these infections has been towards the application of theory to the applied problem of how best to design mass vaccination programmes to control infection and disease (Anderson & May 1991; Agur *et al.* 1993*b*; Babad *et al.* 1994; Nokes & Swinton 1994). There is much activity in this field and I wish to focus on two different problems of relevance in most developed countries. These are (*a*) the relative merits of high and low efficacy vaccines and (*b*) the merits of different policies for the use of combined MMR (measles, mumps & rubella) immunization to eliminate the transmission of these diseases. In both cases the basic mathematical framework employed is an SEIR type with age structure and seasonal transmission.

(*a*) *High and low efficacy vaccines*

Although the number of major pharmaceutical companies active in vaccine manufacture is dwindling rapidly (a cause for concern to departments of health), of those active the search for more efficacious variants of current vaccines continues. The reasons are twofold. First, as vaccination coverage continues to rise in many western countries (England and Wales provide a good example where greater than 90% coverage with MMR vaccine and with pertussis and diphtheria vaccines is achieved by age 24 months by more than 90% of district health authorities) vaccine failure becomes of greater significance in attempts to eradicate local transmission. Vaccine efficacy varies by vaccine type but a lower bound, for example, on the efficacy of the measles component of MMR is around 90% (Babad *et al.* 1994). A further reason for improved efficacy relates to the effectiveness of the vaccine in immunizing infants who still have moderate to high titres of maternally derived antibodies to a particular infectious agent. Maternally acquired protection to viruses in the infant typically wanes over a period of 6 months to one year. Immunization in this period often fails to protect a child in later life (Yeager *et al.* 1977; Aaby *et al.* 1989) but high efficacy vaccines may be more effective in immunizing such infants.

An unfortunate characteristic of most vaccines is that in some very small fraction of those immunized, serious complications concomitant with disease may result from the procedure. For example, severe acute neurological illness post pertussis, diphtheria and tetanus vaccination occurs in a very small fraction of immunized children (Miller *et al.* 1993). Complication rates per dose of vaccine administered appear to be related to vaccine efficacy or potency, with high rates occurring (but still a very small risk) with the more potent vaccines (Christensen *et al.* 1983; Vesikari *et al.*

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

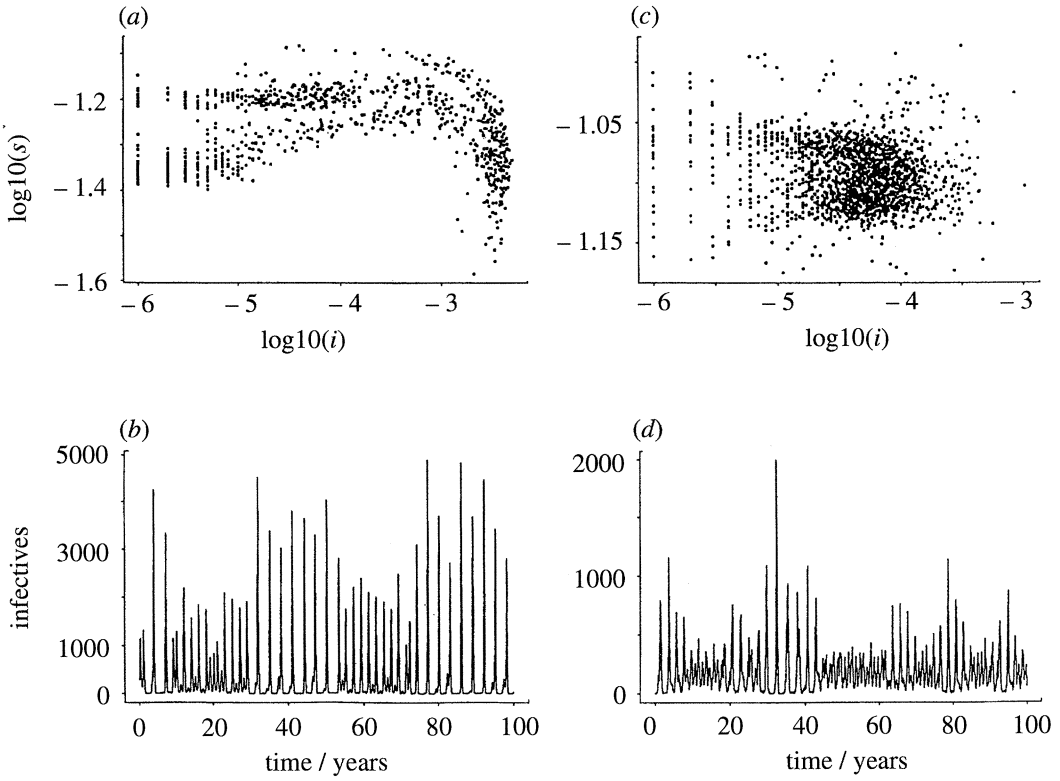

Figure 3. Monte Carlo simulations of two models of the transmission dynamics of the measles virus (Grenfell 1994). SEIR denotes the standard compartmental model (susceptibles, incubators, infected and recovered individuals) (graphs (*a*) and (*b*)) and the RAS model is the realistic age-structured model (graphs (*c*) and (*d*)); in both cases the rate of infection varies seasonally to denote the influence of school terms and holidays on mixing patterns. Graphs (*a*) and (*c*) record the number of susceptibles plotted against the number of infectious individuals at each point in time, while graphs (*b*) and (*d*) record temporal trends in the number of infected individuals (see Grenfell 1994). The SEIR pattern in graph (*a*) shows a chaotic pattern while the RAS model in graph (*c*) shows less structure.

1983; Popow–Kraupp *et al.* 1986). A specific example is provided by the high efficacy Urabe Am 9 mumps vaccine (97.1%–98.7%) and the lower efficacy Jeryl Lynn vaccine (92.2%–95.4%). The upper limits for the complication rate of the former is 1/62 000 while that for the latter is 1/250 000 (Fescharek *et al.* 1990; Furez & Contreras 1990). To complicate matters, natural infection itself carries a risk of serious disease arising during the course of infection. The complication rates from natural infection are again low but significantly greater than those arising from immunization (RCGP 1974; Anderson *et al.* 1987). The optimum strategy for individual parents is either not to vaccinate their child but hope that everyone else has vaccinated theirs, or to opt for the lower efficacy but safer vaccine. What is the optimum policy for the community given the objective of minimizing the total number of complications, whether arising from infection or vaccination?

This problem is ideally suited to the techniques of analysis used in mathematical epidemiology. Appropriate SEIR models with age structure can be adapted to mirror the transmission dynamics of mumps under different assumptions concerning which vaccine is used and what level of coverage (at what age) is achieved (Nokes & Anderson 1991). Without some formal quantitative template upon which to base public policy formulation, it would not be easy to arrive at appropriate quantitative figures for comparative assessment of the different options available. Nokes & Anderson (1991) performed a series of simulations to mimic the impacts of the low and high efficacy vaccines under overall coverage levels of 70%, 80% and 90%, and calculated the predicted accumulated number of cases of complications due to the vaccine and natural infection over a 20 year period (figure 4). The results suggest that if mass vaccination coverage is high enough to block effectively viral transmission (80%–90%), the best policy in the long term is to use the vaccine with the lowest complication rate. At lower levels of uptake (70%) the more efficacious vaccine with the higher complication rate (Urabe) is better at protecting the community against complications due to infection. This is an example of: what is best for the individual (the vaccine with the lowest complication rate) is not necessarily best for the community. In this particular case, the high efficacy vaccine was withdrawn by the manufacturing company due to the higher complication rate. However, in some developing and developed countries with lower vaccine uptake levels than in the U.K. and U.S.A., the best protection to the community is afforded by the high efficacy vaccine. Similar principles apply in the case of measles immunization, particularly in developing countries. Again high efficacy vaccines have been withdrawn

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

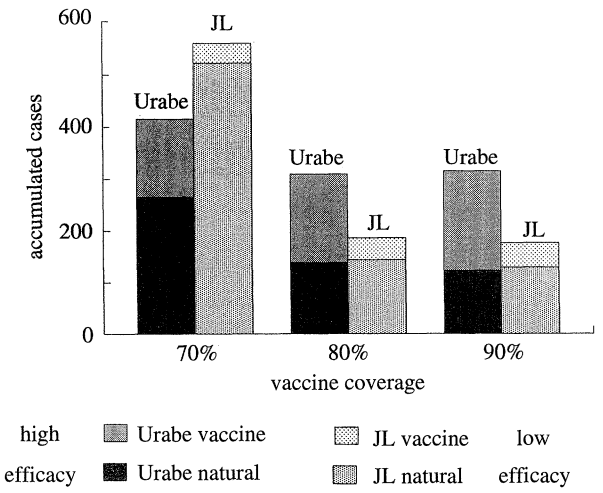

Figure 4. Model predictions of the effect of high (Urabe Am 9) and low (Jerl Lynn) efficacy mumps vaccines on the accumulated cases of serious disease over a 20 year period after the introduction of mass childhood immunization at three levels of coverage (70%, 80% and 90%) (after Nokes & Anderson 1991). Each bar represents estimated numbers of vaccine-induced neurological complications (upper part) based on the most pessimistic estimates of the vaccines' effect, and complications due to natural infection with mumps (lower part) leading to central nervous system (CNS) disorders.

from use due to the high complication rates associated with their use. However, natural infection is not only a direct cause of morbidity and mortality but recent studies suggest that in poor communities it may also lead to excess mortality in children who were exposed to measles early in life (Aaby *et al.* 1993). In such cases, if the aim is to maximize the protection of the community, and if practical issues mitigate against achieving high levels of coverage, the high efficacy/high complication rate vaccine may still be better at minimizing morbidity and mortality. Mathematical models that permit quantitative investigation, if sensibly and cautiously used, will help to sharpen discussion of these issues. They will not resolve the very real conflict between the interests of the individual versus those of the community. This conflict will arise in every successful vaccination campaign since, ultimately, if the disease is close to eradication the risks associated with vaccination will always exceed those associated with infection.

(*b*) *Mass immunization and herd immunity*

An important landmark in attempts to introduce mathematical epidemiological techniques into the framework of public health policy formulation was the generation and subsequent testing of a series of predictions concerning the potential impact of childhood mass vaccination programmes on herd immunity and observed patterns of infection and disease. Simple age-structural deterministic models of the SEIR variety predicted that mass vaccination would act to increase the average age at infection over that pertaining prior to vaccination, to lengthen the interepidemic period and to create a trough of susceptibility to infection in the age classes just above those targeted at the onset of the mass vaccination programme which would persist as a 'ripple' on the surface of herd immunity created by plots of immunity versus age and time (Anderson & May 1982*b*, 1983, 1991; Shenzle 1984). In the U.K. these predictions helped to stimulate the systematic collection of age stratified serological profiles across the entire age range from birth to life expectancy by the Public Health Laboratory Services (Nokes *et al.* 1986, 1990; Morgan-Capner *et al.* 1988). As vaccination campaigns have developed to the state where high levels of coverage pertain in a number of countries in Europe and in Canada and the U.S.A., it has been possible to test these early predictions. All have been supported by observed trends. An example is presented in figure 2 which records the impact of childhood rubella immunization in Finland (started in 1984) and the associate ripple of susceptibility (Ukkonen & Bronsdorff 1988), and a similar profile in England where vaccination was targeted at teenage girls over the period recorded (Nokes & Anderson 1988). Theory predicted that the latter policy would not induce a ripple of susceptibility (Anderson & May 1983).

A rough guide to the fraction of a population, p, that must be immunized by age A to eradicate transmission is provided by simple models (see Anderson & May 1991) where

$$p > \left[\left(1 - \frac{A - M}{L - M}\right)\left(1 + \frac{V}{L}\right)\right]/q. \qquad (1)$$

Here L is life expectancy, V is the average age at vaccination, A is the average age at infection prior to the introduction of mass immunization, M is the average duration of protection provided by maternal antibodies and q is vaccine efficacy. Seroconversion rates (equated with vaccine efficacy, but the seroconversion fraction may overestimate efficacy) with the current components of the MMR vaccine are high: 90–100% for mumps, 94–100% for measles and 99–100% for rubella (Isaacs & Menser 1990). In the use of MMR, the virus with the greatest transmission efficacy prior to mass immunization sets the target for coverage. For MMR this is measles: in urban centres in the developed world the average age A at infection was of the order of 5 years (6–8 years for mumps, 9–10 years for rubella – $R_0 = A/L$). With a life expectancy of around 80 years in western Europe and with the average duration of maternal antibody protection set at six months we can calculate from (1) how p varies with the average age at vaccination and vaccine efficacy as depicted in figure 5. With low vaccine efficacy (i.e. around 95%) and with a medium to high average age at vaccination (2–4 years) the predicted level of coverage to eradicate transmission is in excess of 100% (i.e. it cannot be achieved with a vaccine of that character on the basis of a single dose vaccination programme). This simple calculation illustrates one reason why current levels of coverage in, for example, the U.K. and the U.S.A. have not eliminated local transmission. Vaccine efficacy is not high enough and

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

insufficient attention is paid to lowering the average age at vaccination (i.e. in the U.K. it is around 1.5–2 years at present for MMR). Simple theory suggests that the optimum age to vaccinate is at the trough in the age serological profile between the decay in maternally derived protection and the rise in exposure to national infection (Anderson & May 1991) (figure 6). Immunization at younger ages fails to protect in a significant fraction of infants, due to the presence of maternally derived antibodies. For example, a study in the U.S.A. suggests that vaccine efficacy may fall to around 82% if the vaccine is delivered at round 9–12 months of age, compared with an efficacy of 94% at around 18 months of age (Yeager *et al.* 1977).

Current levels of immunization with MMR are high in the U.K. (figure 7) but with present efficacy levels and the current average age at immunization they are not predicted to be high enough to block transmission (figure 5). Additional problems arise from heterogeneity in vaccine uptake in different spatial locations with low uptake in poor urban centres especially in ethnic minority groups who are often difficult to reach via current health education programmes. Similar problems arise in the United States. These issues have prompted further discussion on how best to improve current immunization programmes either by introducing two dose schedules – one at age 1–2 years and the other at age 4–5 years – or by the introduction of vaccination campaigns where at set times, say at intervals of a few years, education and publicity are geared to a short intensive pulse of mass vaccination across a wide band of age classes (i.e. all primary and secondary school children).

These discussions have also stimulated further mathematical studies to help in deciding which option is best, in terms of reducing incidence and minimizing cost (Agur *et al.* 1993*a*,*b*; Babad *et al.* 1994; Nokes & Swinton 1994). With respect to pulse vaccination campaigns targeted at all age classes, theory based on simple models can help to define (approximately) the optimal interval between the campaigns, τ; where

$$\tau \approx pqA. \qquad (2)$$

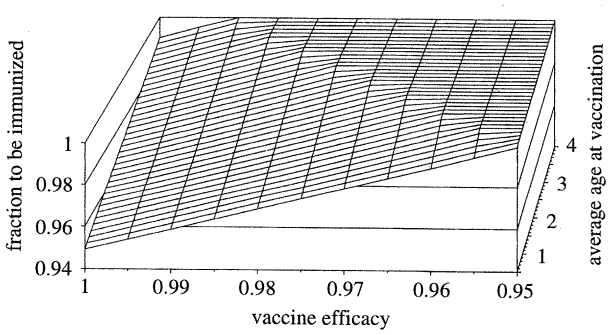

Figure 5. Predicted fraction that must be immunized (based on equation 1 in the main text) to block transmission as a function of vaccine efficacy (q, the fraction successfully immunized) and the average age at vaccination (a). For values greater than unity, a single stage vaccination programme will not be able to eliminate transmission.

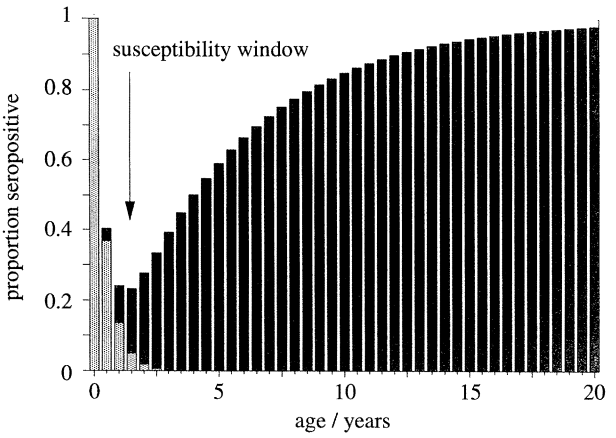

Figure 6. Diagram of a cross-sectional serological profile recording the fraction with antibodies to a particular infectious agent as a function of age. The lightly dotted bars denote the wane of maternally derived protection and the darkly dotted bars denote antibodies resulting from recovery from infection. In this representation the average duration of maternal protection was set at 6 months and the average age at infection was set at 5 years.

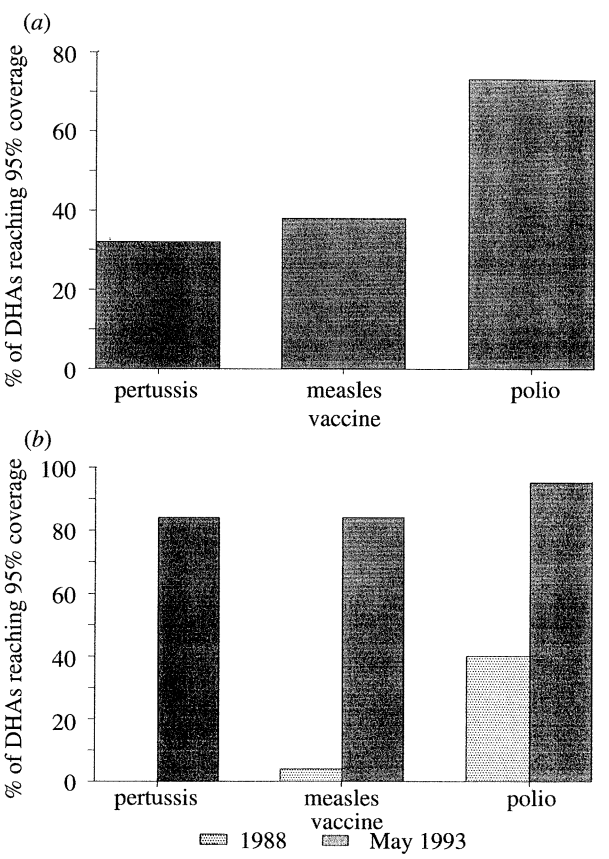

Figure 7. Current levels of vaccine uptake in England. (*a*) Percentage of District Health Authorities (DHAs) that achieved 95% coverage of young children in September 1993 for three vaccines. (*b*) Percentage of DHAs reaching 90% coverage in 1988 (stippled bars) and in May 1993 (shaded bars) for the same three vaccines (pertussis, measles and polio). (Source: Department of Health vaccine uptake statistics.)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

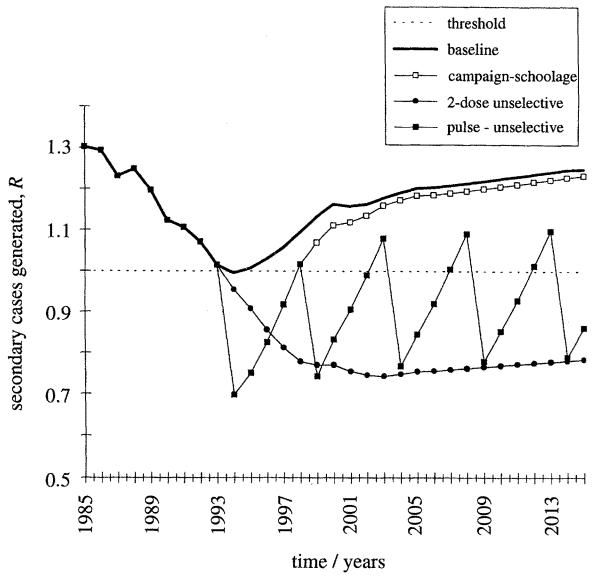

Figure 8. The predicted impact of different vaccination programmes on the average number of secondary cases of measles generated by one primary case in a vaccinated population (R) (from Babad *et al.* 1994) (see text). As defined in the text the graph records changes over time in the value of R for a *baseline* campaign (childhood cohort vaccination as currently in operation in England and Wales), the addition of a single pulse campaign targeted at school-aged children (campaign: school age), a 2-dose unselective (selection not based on previous vaccination history) and a continued pulse campaign at 5 year intervals (see text).

Here p, q and A are as defined for (1). The objective of this approach is to maintain the fraction susceptible below the critical level for the persistence of the infectious agent (i.e. $R_0 < 1$). One such calculation for the population of Israel with an age structured SEIR model suggests that a pulse vaccination campaign every 5 years that successfully immunizes 85% of children in the age range 1–7 will suffice to eliminate local transmission (Agur *et al.* 1993*a*). The approach has been tried in practice to control polio and measles in various regions of the developing world (Nokes & Swinton 1994). However, our understanding of the long term effects of pulse campaigns are limited at present. If the age range that can be reached in a campaign is restricted and vaccine efficacy is not close to 100% then the optimum interval between pulses may reduce to one year such that this approach is in essence the same as a cohort vaccination programme. Furthermore, wide interval (a few years) pulse campaigns may lead to fluctuations in the age distribution of susceptibles. This in turn will influence the ideal age range to target in a particular campaign (Nokes & Swinton 1994).

These problems argue for the augmentation of the cohort programme widely adopted in western countries by a further round of immunization at an older age. A recent study based on mathematical models has attempted to assess what form of two-stage policy is best (Babad *et al.* 1994). Some results from this study are presented in figures 8 to 10. A realistically age structured (RAS) model was used in this study with seasonal and age-dependent transmission (Schenzle 1984). Prior to the introduction of immunization (1956–1965) in England and Wales the model closely mimics observed seasonal trends in the incidence of measles (figure 9) and recorded serological surveys in 1969, 1979–1983, 1986 and 1991 (those after 1967 are post the introduction of mass vaccination) (figure 10). The agreement between

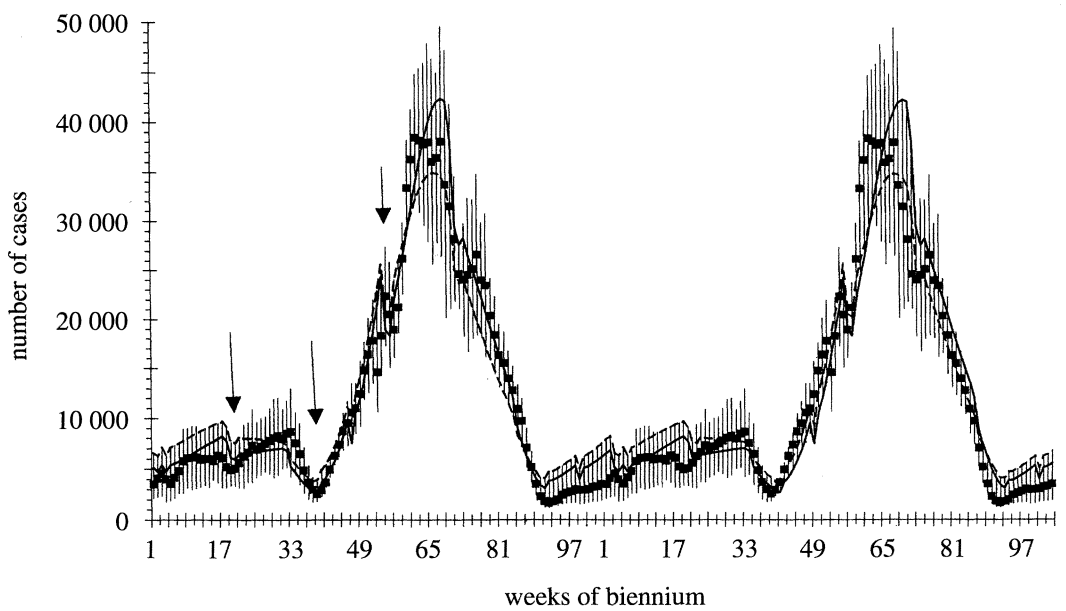

Figure 9. Predictions of a realistic age structured model (RAS) of the transmission dynamics of measles in England and Wales, with seasonal forcing, compared with observed trends (weekly incidence of measles) prior to the introduction of vaccination over the two-year cycle in incidence. The vertical bars denote the range of variation over a series of 2 year periods (1948 to 1967), where the solid squares record the average values. The dashed and solid lines denote predicted trends for two different assumptions concerning the manner in which mixing occurs between different age classes of the population (see Babad *et al.* (1994) for the details).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

prediction and observation supports the application of the model in the assessment of changes in the vaccination programme and future trends in incidence. Four different vaccination strategies were examined: (i) the *baseline* of 90% coverage of boys and girls in the second year of life (as at present) with a vaccine efficacy of 90%; (ii) baseline plus a one-off *campaign* in which 50% of target group (preschool 1–4 year olds, primary school 11–12 year olds, all school age children, 12–16 year olds) are vaccinated over a one month period in 1994, (iii) a *two-dose* policy where baseline is augmented by 50% coverage of target group initiated in 1994 and repeated each year (target group 4 year olds or 13 year olds) and (iv) a *pulse* policy added to the baseline with a campaign every 5 yrs in which coverage is 50% of 5–11 year olds beginning in 1994. Simulated temporal trends for a selection of these strategies are presented in figure 8 by reference to changes over time in the effective reproductive rate R (the number of secondary cases generated per primary case in the vaccinated population). If the value of R is maintained at less than unity local transmission will be blocked. The most effective approach is predicted to be a two-dose policy with the second dose targeted in an unselective way (i.e. not in relation to previous history of vaccination) at 4 year olds (Babad *et al.* 1994). Whether or not to adopt this policy in the U.K. is a matter of discussion at present. In figure 8, the main reason why the current baseline policy fails to eradicate transmission is due to vaccination failure as a consequence of the less than a 100% efficacy of current vaccines. The techniques used to assess the impact of MMR are currently being extended to examine policies for the introduction of new vaccines such as hepatitis B vaccine (see figure 11).

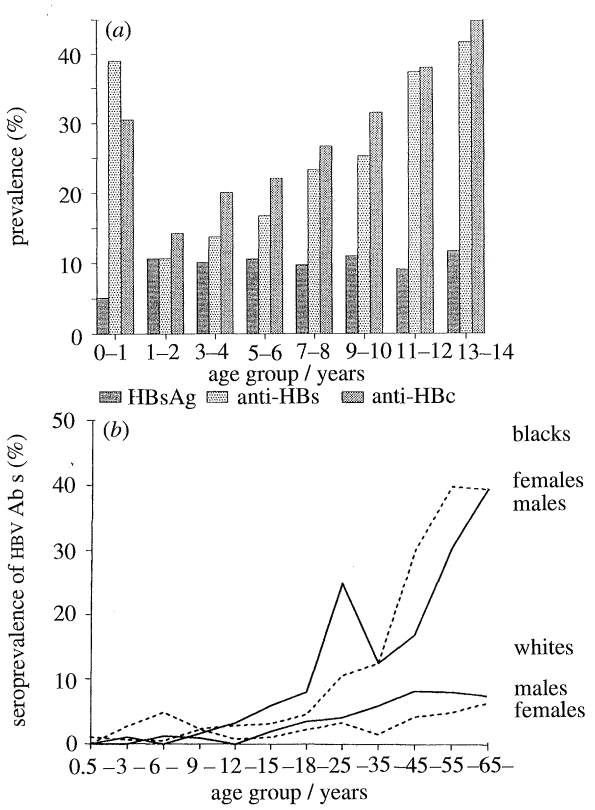

Figure 11. (*a*) The epidemiology of hepatitis B in Taiwan in 1984 (prior to mass immunization) as depicted by cross-sectional studies of the prevalence of antibody surface markers (anti-HBs), antigen surface markers (HBsAg) and antibody core markers (anti-HBC) (from Hsu *et al.* 1986). (*b*) Seroprevalence of hepatitis B virus markers in the U.S.A. during 1976, stratified by age and ethnic group (from McQuillan *et al.* 1989).

3. HETEROGENEITY IN EXPOSURE TO INFECTION

Heterogeneity in exposure to an infectious agent can arise in many different ways. Its presence has important implications both for the interpretation of observed epidemiological pattern and for the design of control programmes. Theoretical studies have focused on spatial (Murray & Cliff 1975; Nold 1980; May & Anderson 1984; Schwartz 1992), seasonal (Aron & Schwartz 1984; Bolker 1993) and age-related heterogeneities (Anderson & May 1983; Schenzle 1984; Diekmann *et al.* 1990). In both age- and season-related trends, the question of who mixes with whom has important implications for patterns of transmission and persistence. Recently, this issue has been brought into much stronger focus by the emergence of the AIDS pandemic which has stimulated a number of mathematical studies of the transmission dynamics of sexually transmitted infections (see Anderson *et al.* 1986; May & Anderson 1987). Early analyses employed a term to describe the per capita rate of infection based on a transmission probability defined per sexual partner (Yorke *et al.* 1978, Hethcote & Yorke 1984). The net rate of transmission was assumed to vary in different segments of the population to reflect heterogeneity in the rate of sexual partner acquisition. An extension of this early work is based on the definition of a probability distribution of the number of different sexual partners per unit of time and proportional mixing between the

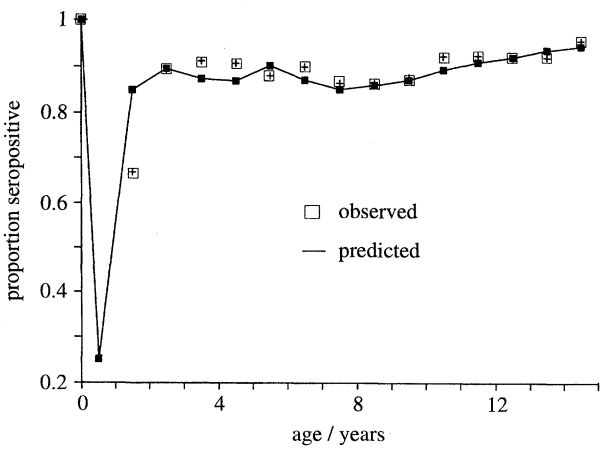

Figure 10. Prediction of the RAS model of the transmission dynamics of measles in England and Wales (solid line and filled squares). The graph records the predicted age-serological profile in 1991 given the reported vaccination uptake, and the observed pattern (from Babad *et al.* 1994).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

different sexual activity classes (Anderson *et al.* 1986; May & Anderson 1987). In this extension the basic reproductive number, R_0, is defined as

$$R_0 = \beta(m + \sigma^2/m)D, \quad (3)$$

where β is the transmission probability per partner, m is the mean rate of partner change per unit of time, σ^2 is the variance of the partner acquisition distribution and D is the average duration of infectiousness. These theoretical studies played a role in stimulating the acquisition of quantitative data from defined populations of the distribution of the sexual partner change rate (Anderson 1988; Anderson & May 1988; ACSF 1992; Johnson *et al.* 1992, 1994). The published surveys of sexual behaviour in large samples drawn from the general population reveal that the mean rate of partner acquisition per unit of time is invariably much smaller than the variance ($\sigma^2 \gg m$) such that most individuals have few partners and a few have many (figure 12). This observation holds for heterosexuals and homosexuals, and within various sectors of the population whether stratified by age, sex, social class or other social or demographic variables (Wellings *et al.* 1994). With low values for the mean over a one year recall period, particularly in heterosexual populations ($m \leqslant 1.0\,\mathrm{yr}^{-1}$ across all sexually active age classes), (3) reveals that the overall magnitude of R_0 is greatly influenced by the prevailing variability in sexual activity (the magnitude of σ^2). This observation is in accord with intuition which suggests that a few highly sexually active individuals may maintain transmission in a much larger community where the majority form few sexual partnerships (the highly sexually active are often referred to as the 'core' group; Hethcote & Yorke 1984).

Equation (3) ignores many important details of human sexual behaviour and the epidemiology of HIV, such as distributed incubation and infectious periods (Anderson 1988; May & Anderson 1989), the formulation of the probability of transmission as per sexual act as opposed to per sexual partner (Kaplan 1990; Blower *et al.* 1991), the formation and break up of partnership (i.e. partnership duration) (Dietz & Hadler 1988), and the heterogeneities introduced by patterns of mixing between different age, sex, sexual activity classes, broader risk groups and individuals living in different spatial locations (Jacquez *et al.* 1988; Gupta *et al.* 1989; Anderson *et al.* 1992; Garnett & Anderson 1993*a*,*b*).

7 In this short commentary on the importance of various types of heterogeneity to the transmission processes, two of the above are examined in more detail.

(*a*) *Mixing patterns between sexual activity classes*

Having stressed the importance of variability in sexual partner change rates to an understanding of transmission success, theoretical studies have also played a role in emphasizing the significance of mixing patterns between sexual activity classes in the interpretation of epidemiological pattern. Mixing may vary on a continuous spectrum from assortative (like with like: in terms of their membership of a

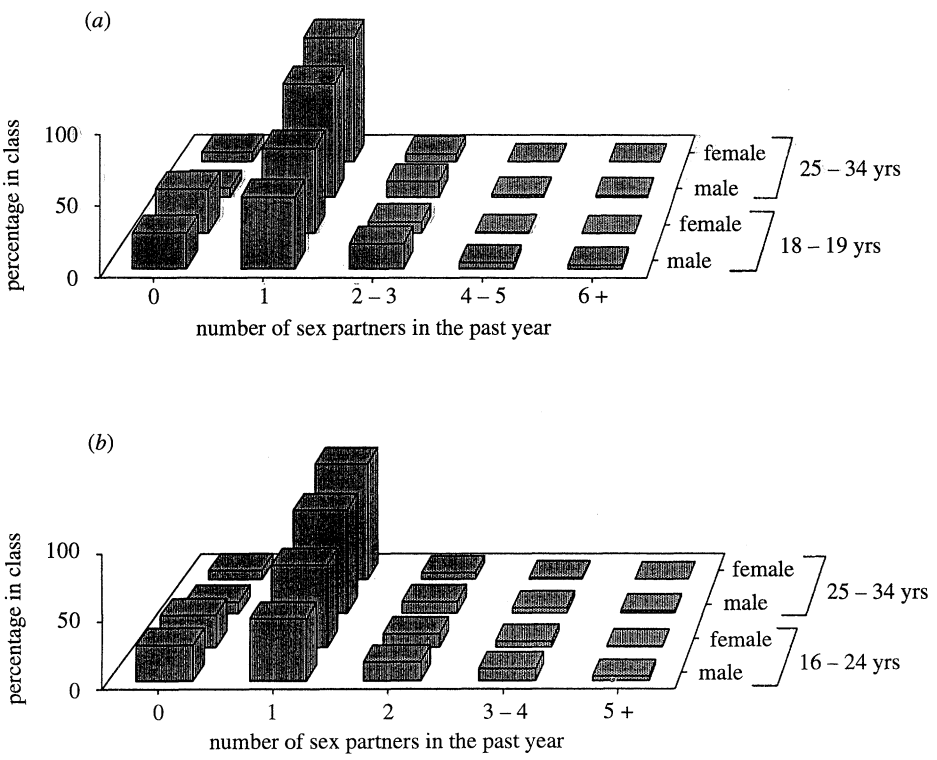

Figure 12. Population based surveys of reported numbers of different sexual partners (for heterosexual males and females) over the previous year prior to interview for two different age groups in (*a*) France (ACSF 1992), and (*b*) England and Wales (Johnson *et al.* 1992). The survey data is recorded as frequency distributions of the percentage of the sample reporting different numbers of sexual partners.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

sexual activity class defined on the basis of partner change rates), via proportional (partnerships are formed in proportion to the fraction in a sexual activity class weighted by the rate of partner change of that class) to disassortative (like with unlike) (Anderson *et al.* 1990). This spectrum may be multidimensional once other strata of the sexually active population are considered (i.e. age, sex, spatial location etc.) (Garnett & Anderson 1993*a*). Once mixing is taken into account in formulating the transmission term the reproductive number must be defined per sexual activity class (i) as the number of secondary cases of infection generated in group i by one case of infection in group j (assuming all in i are susceptible), $R_0(i, j)$, where

$$R(i, j) = c_j p_{ij} \beta D. \quad (4)$$

Here β and D are as defined in (3), c_j denotes the effective rate of sexual partner change of group j (taking account of the mean, m_j and variance σ_j^2) and p_{ij} denotes the probability that a person in group i forms a sexual relationship with someone in group j. There are a series of obvious constraints on the value of p_{ij} (Jacquez *et al.* 1988). Equation (4) makes clear that the infection may persist in the total population if only certain of the values of $R_0(i, j)$ exceed unity in value (Anderson 1991). A series of recent studies have revealed that the values of the elements of the mixing matrix (p_{ij}) are extremely important determinants of the pattern of the epidemic. This is illustrated in figure 13 by two simulations of an AIDS epidemic in a male homosexual community where all parameters were identical in the two projections except the mixing pattern (Jacquez *et al.* 1988; Garnett & Anderson 1993*a*,*b*; Hyman & Stanley 1994). If mixing is highly assortative the epidemic takes off quickly as infection spreads rapidly in the high activity classes. It then moves more slowly through the medium and low activity classes (in a series of waves, that may be distinct if very little mixing occurs between activity classes). The overall magnitude of the epidemic is small by comparison with the proportional mixing example since it is largely restricted to the high activity classes. Conversely if mixing is proportional or disassortative the epidemic takes off more slowly but is much larger in its overall magnitude.

The measurement of mixing patterns is very difficult in practice since a survey must enquire not only about numbers of sexual partners but also about their identity in order to construct a network of sexual contacts. Some studies have attempted to do this (Haraldsdottir *et al.* 1992) in particular risk groups, but more work is needed. The most promising line of investigation is via contact tracing studies in sexually transmitted disease clinics for infections such as gonorrhoea where treatment is available for partner or contact of an infected person (Potterat *et al.* 1985; Rothenberg & Potterat 1988; Garnett & Anderson 1993*b*). More generally, for the spread of HIV in heterosexual populations the mixing matrix must be extended to cover the stratifications of sex, age and spatial location (the inclusion of a term $p_k(i, j, a, a', s, s')$ to denote the probability that someone of sex k, activity class i, age a in residency location s forms a partnership with someone of the opposite sex of sexual activity group j age a' and in residency location s'. Many problems arise once this level of complexity is addressed, both in theoretical work and in empirical studies. In a theoretical context, the problems relate to balancing demand for and supply of sexual partners in the various stratifications of the population, particularly in the context where AIDS-induced mortality acts to change the composition of the population (Blower & McLean 1991; Garnett & Anderson 1993*a*, 1994). Sets of behavioural rules have to be set up to define what happens when supply does not match demand. Empirical studies of these added complexities are very limited but a start has been made in addressing mixing across age classes (Hogsborg & Aaby 1992) and between different residency locations (Garnett & Anderson 1993*a*). Much remains to be done, however, in order to provide precision to the interpretation of observed trends in HIV seroprevalence. The work has obvious application since altering mixing patterns via education (i.e. male contact with female prostitutes) can have a very significant effect on the course of the epidemic in a defined community (Anderson *et al.* 1991).

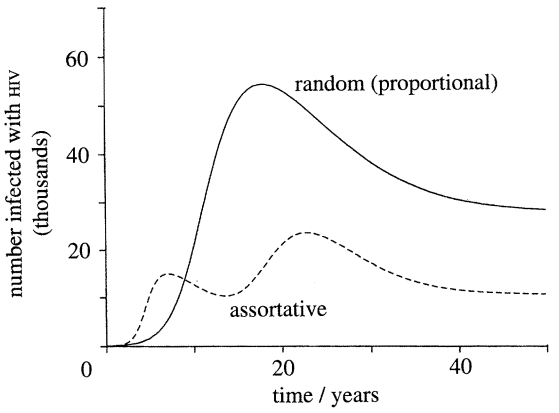

Figure 13. The impact of different mixing patterns between sexual activity classes (defined on the basis of partner numbers per unit of time) on the magnitude of a simulated epidemic of HIV-1 in a male homosexual community (Gupta *et al.* 1989; Anderson *et al.* 1989). In the two projections of changes over time in the number infected with HIV-1 all parameters are the same except for the pattern of mixing: in one case mixing is proportional (= random), while in the other it is moderately assortative (like with like).

(*b*) *Sex acts and sex partners*

In the majority of studies of the influence of mixing patterns the probability of transmission is defined per partnership (although not in all, see Hyman & Stanley 1994). This is a very crude assumption and a further influence of surveys of sexual behaviour has been the realization that the distribution of sexual acts per partner per unit of time is also very skewed in form where the variance is much larger than the mean. Ideally, we need to reformulate the transmission probability (and hence the definition of R_0) in

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

terms of heterogeneity in both partner numbers and sex acts per partner per unit of time, taking account of the covariance between these two variables. In the simplest case, we can express β in terms of the average number of sexual acts (of a defined type, i.e. penal-vaginal intercourse), a, and the probability of transmission per act, γ. If the number of acts per partnership is a Poisson process then after a acts, β is given by

$$\beta = 1 - \exp(-\gamma a). \tag{5}$$

We could define a as the average number of acts per partnership. But following the earlier example of heterogeneity in partner numbers and its influence on the magnitude of R_0 (3), we need to formulate transmission success in terms of the distribution of acts per partner. To proceed, we need to define two unconditional distributions (partners per unit of time, p, and acts per partner, a) and the relationship between acts and partner numbers. Let us assume for simplicity that the distribution of partner numbers per unit of time is gamma in form with mean $m = v/\alpha^2$ where v and α are the two parameters that define the gamma distribution. Let us also assume that the distribution of acts per partner is distributed as a negative binomial with mean $a(p)$ (for someone with p partners per unit of time) and parameter $k(p)$ (the negative binomial clumping parameter). With these assumptions, plus the definition of the functional relationships between $a(p)$ and $k(p)$, and the variable p (partners/unit of time), it is possible to derive an expression for the probability of transmission per partner (May & Anderson 1994), β, where

$$\beta = \frac{\alpha^v}{m\Gamma(v)} \int_0^{\alpha} p^v \, \mathrm{e}^{-\alpha p} \left[\frac{[1 - 1 + \gamma a(p)/k(p)]^{-k(p)}}{[1 - (1 + a(p)/k(p)]^{-k(p)}} \right] \mathrm{d}p. \tag{6}$$

As in (5), γ is the probability of transmission per act. On the evidence provided in surveys of sexual behaviour that record partner numbers per unit of time and acts per partner (Johnson *et al.* 1994; Wellings *et al.* 1994), a suitable form for the function $a(p)$ is

$$a(p) = \sigma + b \exp(-gp), \tag{7}$$

where σ, g and b are constants. The value of $a(p)$ declines as partner numbers rise, to mimic the decreasing average duration of each partnership (figure 14). Furthermore, it is possible to define the relationships between the mean, m, and variance, σ^2, of the partner distribution in terms of a power law in line with observed pattern (Anderson & May 1988) and similarly that of the acts distribution. In the latter case this enables us to define the function $k(p)$ simply in terms of the mean $a(p)$ (i.e. $k(p) = a(p)^2/[ha(p)^r - a(p)]$ where the variance of $a(p)$, $S^2(p)$, is defined as $S^2(p) = ha(p)^r$). Thus with a knowledge of the means of the two distributions (m and $a(p)$), plus a definition of the function $a(p)$, we can calculate the magnitude of β from (6).

We can now assess the influence of heterogeneity in acts per partner on the magnitude of the basic reproductive rate of infection, R_0, where $R_0 = \beta MD$ (here D is the average duration of infectiousness). An illustration is in figure 14 (bottom graph) which illustrates the rise in the value of the basic reproductive number, R_0, as heterogeneity in sex acts per partner rises ($k \to 0$) for a fixed relationship between the mean number of acts per partner and partner numbers (figure 14, top graph). As in the case of heterogeneity in numbers of sex partner per unit of time (see (3)), heterogeneity again acts to enhance the likelihood of the persistence of the infection ($R_0 \geqslant 1$). However, the present case is somewhat more complex since the behaviour of (6) depends on the values of a number of parameters. How heterogeneity influences the magnitude of R_0 is very dependent on the assumption made concerning how the degree of heterogeneity (measured inversely by k) varies according to partner numbers (the function $k(p)$ in (6)). Assuming k to be constant for all values of p is a simplification. It implies that amongst those with high

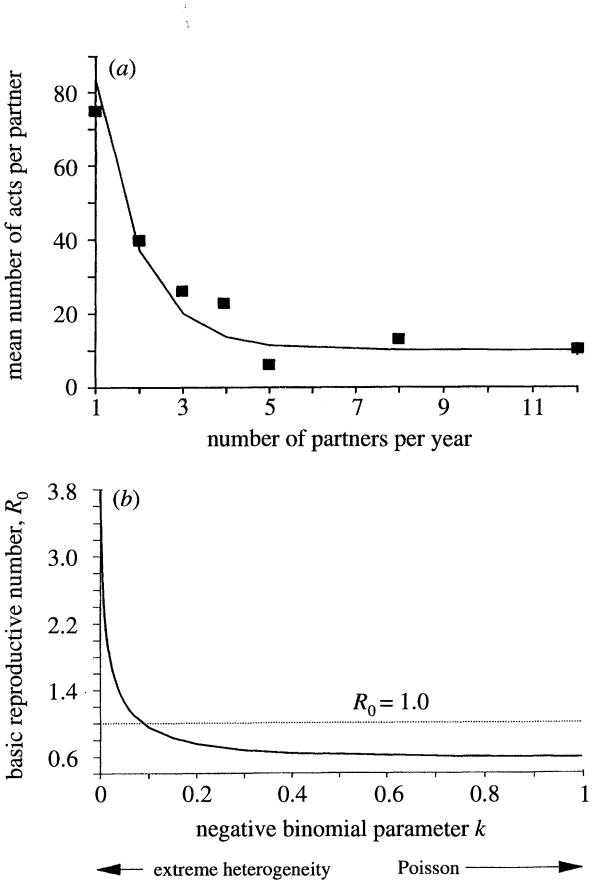

Figure 14. (*a*) Reported relationship between the mean number of sex acts (penetrative penal-vaginal sex) per partner and the number of sex partners over the past year (for heterosexuals) in the national survey of sexual lifestyles in England and Wales (Wellings *et al.* 1990). The squares are the reported values while the solid line is the best fit of the function $a(p) = \sigma + b \exp(-gp)$ where p is partner number per unit of time and σ, b and g are constants. (*b*) The relationship between the magnitude of the basic reproductive rate R_0 and the degree of aggregation or heterogeneity in the distribution of sexual acts per partner (negative binomial in form with clumping parameter k) as predicted by equation (6) in the main text. The magnitude of k was assumed to be constant for all values of the function $a(p)$.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

partner numbers (and hence a low mean number of acts per partner) great heterogeneity still persists. Further empirical studies are required to investigate this more fully but the data base provided by the U.K. National Survey on sexual lifestyles (Johnson *et al.* 1994) provides an opportunity to explore this problem in more detail.

The general point illustrated by the examples in this section is that heterogeneity in transmission plays an important role in observed epidemiological pattern. Models that encapsulate such detail provide a better template with which to interpret the data.

4. THE POPULATION DYNAMICS OF THE IMMUNE SYSTEM

The development of new techniques in the fields of molecular biology and immunology have enabled researchers to classify, more and more finely, the various types of cells and chemicals (e.g. cytokines) that constitute, direct and regulate the human immune system. The system appears to be very complex, containing numerous components, many of which interact with each other, and invading infectious agents, in highly nonlinear ways, such that responses to stimuli from an antigen or other cell types often reveal threshold or saturation effects as the intensity of the stimulus increases. As the ability to classify the numerous components of the immune system becomes more precise there has been a tendency to develop ever more complicated explanations of observed events (Clark & Ledbetter 1994; Schwartz 1994). In these circumstances, as in any other scientific field, mathematical models provide a powerful method for the analysis of particular hypotheses and for the interpretation of observed patterns (Anderson 1991*a*). Unfortunately, however, the mathematical literature on models of the immune system has been somewhat divorced from the experimental research and it is rarely referred to by experimentalists. Indeed, many would argue that theoretical work has had little or no impact on current concepts in immunology. Most immunologists distrust the simplicity of mathematical models, given the all too apparent complexity of the immune system (and, our incomplete understanding of this complexity). This concern is somewhat puzzling, since the approach adopted by the theoretician is to start simply and to slowly build in complexity, in a manner akin to the laboratory scientist who allows one or a few factors to vary as others are held constant within the experimental design. A much more valid criticism has been the tendency for theory to soar free from the constraints of data and for it to be (in general) so distant from the current ideas and concepts that excite the experimental immunologist (Perelson 1990). The fault does not entirely lie with the theoreticians since most experimentalists rarely attempt to quantify the many rate processes (e.g. birth and death rates of cell types) and functional relationships that dictate recorded pattern. The immune system is a very dynamic one and understanding of its responses and capabilities will require quantitative studies of rates of change and the controlling processes. The way forward is clearly a closer collaboration between the two groups of scientists, and recently there have been some encouraging signs of progress (Michie *et al.* 1992; Freitas & Rocha 1993).

In attempts to meld the mathematical and experimental approaches to immunological research, it helps to recognize that much can be learned from related fields such as that of population ecology. Within the immune system, we are dealing with complex 'multispecies' populations of cells that communicate by chemical signals of various kinds. Furthermore, the system is spatially structured with cells circulating (trafficking) via compartmental 'habitats' such as the lymph nodes where B and T cells, for example, mature and come into contact with antigens (or antigen presenting cells). Population ecologists have long been concerned with the dynamics of such multispecies systems in spatial heterogenous environments. Furthermore, there is a long tradition of melding theory with observation and experiment to generate testable hypotheses concerning the processes responsible for observed pattern. The examples of simple models that attempt to mimic the dynamics of the immune system considered in this section are very much derived from thinking about the immune system from the viewpoint of a population biologist.

(*a*) *Regulation of cell proliferation*

Millions of newly formed lymphocytes leave the bone marrow daily. They either multiply, differentiate or die depending on the nature, timing and location of interactions with other cells of the immune system or with foreign antigens. Antigen specific T-cell activation is dependent on interaction between T cells and specialized antigen-presenting cells (i.e. macrophages or B cells). Once active, the T cells promote B-cell activation both by direct contact and via released T-cell derived cytokines (IL-2, IL-4, IL-5). Activation of T and B cells result in the clonal expansion (proliferation) of cells specific to a particular antigen and these in turn, via cytotoxic action or via humoral response (antibody production), attempt to remove the foreign antigen. One of the central problems in understanding the dynamics of immunological responses to foreign antigens (in particular infectious agents that are in essence replicating antigens) is what regulates cell proliferation and, in particular, what is the nature of the functional relationship between the rate of proliferation and antigen concentration. In the absence of regulating processes (density-dependent regulation in the jargon of the population ecologist), our circulatory systems would become a dense soup of dividing lymphocytes. Early experiments of T-cell proliferation rates as a function of antigen concentration revealed a convex nonlinear relationship, where the rate changes from low to high, to low again, as antigen concentration increases (Knight 1987) (figure 15). The precise form of the function depends on other variables such as the concentrations of antigen

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

presenting cells (APCs) and IL-2 (Suzuki *et al.* 1988; Jenkins 1992). Many studies reveal saturation in the rate of proliferation as antigen concentration rises, sometimes of a sigmoidal form. These suggest that there is a maximum possible rate of proliferation, irrespective of antigen or antigen presenting cell concentrations (Torsella *et al.* 1992). What induces convex curves (figure 15) is less certain but they suggest a form of anergy or T-cell tolerance. Both mechanisms may have evolved to prevent destruction of self-antigens (Jenkins 1992) in combination with the process of eliminating lymphocytes that recognize the body's own antigens (clonal deletion). Although detailed studies of the mechanisms that induce tolerance and anergy are in their infancy, one consequence of the patterns depicted in figure 15 is the regulation of cell proliferation.

A simple model of T-cell proliferation in response to antigen of concentration A, is of the form

$$dI/dt = \Lambda - \mu I - \beta IA, \tag{8}$$

$$dT/dt = \beta IA + f(T, A) - \mu T, \tag{9}$$

where I denotes immature T cells and T denotes activated cells. The parameter Λ represents recruitment of new cells from the bone marrow, μ is the death rate, β is the contact rate (between cells and antigen) and the term f denotes the proliferation response as a function of cell density, T, and antigen concentration, A. When the model is extended to mirror immunological responses to an infectious agent (replicating antigen) a third equation may be added as follows:

$$dA/dt = rA - g(T, A), \tag{10}$$

Here r denotes the replication rate of the pathogen and the term g denotes the killing of the pathogen by the immune system which is assumed to be a function of activated T cell and pathogen abundance. A number of recent theoretical studies have revealed that even this very simple interaction of a pathogen with the immune system may exhibit complicating dynamics (Kevrekedis *et al.* 1988; McLean & Kirkwood 1990; Schweitzer & Anderson 1992; Schweitzer *et al.* 1993). For example, the system may exhibit different patterns of behaviour depending on initial conditions where two or more equilibria exist depending on the precise forms of the functions f and g (in (9) and (10)). An illustration of the dynamical properties of one such model is presented in figure 16. This example is intended to mirror the dynamics of a replicating pathogen within the host and the two classes of behaviour representing pathogen 'runaway' and depressed T-cell responsiveness (= tolerance), and pathogen elimination and elevated T-cell responsiveness, are separated by an unstable boundary. Initial conditions determine which state pertains, with high exposure early in life leading to tolerance concomitant with persistent infection, while low exposure results in recovery and lasting immunity (Schweitzer & Anderson 1992). This may be an appropriate paradigm for the induction of tolerance and persistent infection in some fraction of individuals exposed to repeated helminth infection or to the *Leishmania* parasites. The model generates the prediction that high exposure early in life (or indeed even prior to birth via infection in the mother) may induce tolerance and parasite persistence. Appropriately designed field studies could test this prediction.

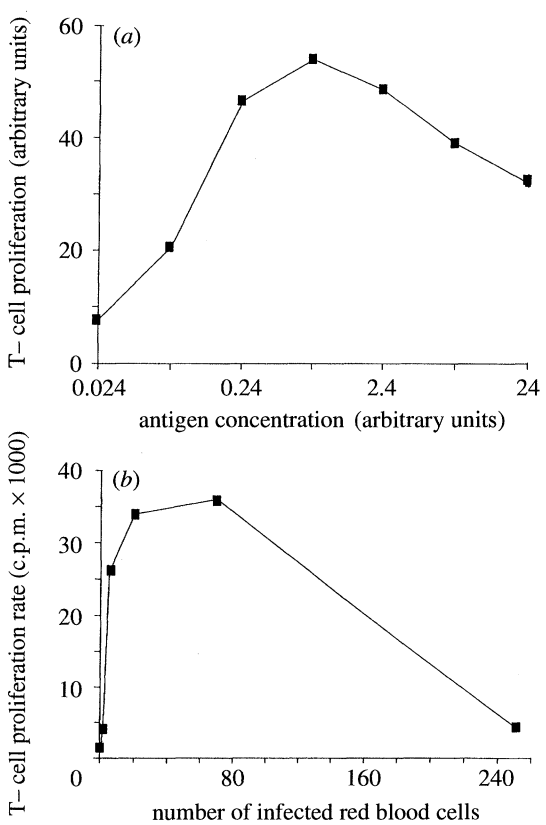

Figure 15. Dose response curves of the influence of antigen concentration ((*a*) pigeon cyctochrome *c*, Matis *et al.* (1982); (*b*) number of *Plasmodium falciparum* infected red blood cells, Jones *et al.* (1990)) on T-cell proliferation rates.

Greater realism can be incorporated in these very basic models to present Th1–Th2 type CD4+ T-cell interactions, antigen presenting cells, cytokine action and the separate components of T- and B-cell mediated immunological attack. However, a major hindrance to such work is the virtual absence of quantitative estimates of either the many rate parameters involved or the functional forms of the relationships between proliferation and pathogen abundance/clearance for any given interaction (whether involving humans or laboratory mammals). There is a clear need for mathematical immunologists to persuade the experimentalist to carry out the necessary experiments, which by today's standards of molecular sophistication are relatively simple in design and aim. This latter factor is often a deterrent: the research is not deemed to be sophisticated enough to attract the experimental immunologist. The counter argument is, of course, that without such quantitative information, the dynamical complexities of these systems, however simple, will not be revealed.

A different approach to that outlined above, is to use simple models to test ideas concerning the biological mechanisms that induce tolerance (or

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

sigmoid or convex proliferation-antigen concentration curves). This is an appropriate role for theory since there exist a number of hypotheses about the mechanism. One example is the two-signal activation hypothesis. Helper T cells appear to need more than just the signal that comes from recognizing an antigen (via presentation by an APC) to become activated in preparation for proliferation. Available evidence suggests they require a second molecular signal from the APC (Mueller *et al.* 1989; Jenkins 1992; Ucker *et al.* 1992; Freeman *et al.* 1993). If this is not received the cell becomes anergic. A number of candidate molecules have been proposed for the role of costimilator, the most popular at present being the B7-2 surface molecule found in antigen presenting cells interacting with the T-cell surface molecule CTLA-4 (Freeman *et al.* 1993).

Verbal arguments about the ability of this hypothesis to explain tolerance at high antigen concentrations can be tested by suitable mathematical models. Recent work by Swinton *et al.* (1994) has tackled this problem with a simple model (to mimic experimental studies of T-cell proliferation) of the form

$$\mathrm{d}T_0/\mathrm{d}t = -\beta(MA)T_0, \tag{11}$$

$$\mathrm{d}T_1/\mathrm{d}t = \beta(MA)T_0 - \sigma M T_1, \tag{12}$$

$$\mathrm{d}T_2/\mathrm{d}t = \sigma M T_1 + \lambda T_2 - \gamma(MA)T_2. \tag{13}$$

Here T_0, T_1 and T_2 are naive unstimulated cells, cells that have received one stimulus and cells that have received both stimuli, respectively. The term MA denotes the combined concentration of macrophages (M) and antigen (A) while β, σ and γ are contact rates and λ is the cell proliferation rate. The key assumptions are that a second signal from the macrophage is required to induce proliferation in class T_2 and that if a further stimulus is received in the dividing class the cells become unresponsive.

The model is designed to mimic an experimental design in which varying concentrations of antigen (A) and antigen presenting cells (M) are added to wells containing naive T cells. It is linear in structure and the proliferation rate, λT, can be derived in terms of A and M where the rate of proliferation R, for an initial concentration of T_0 of τ, is given by

$$R = \frac{\lambda ba\tau}{(d+a)(d+b)}\mathrm{e}^{dt} + \frac{\lambda ba\tau}{b-a}\left[\frac{\mathrm{e}^{-bt}}{d+b} - \frac{\mathrm{e}^{-at}}{d+a}\right] \tag{14}$$

Here $a = \beta MA$, $b = \sigma M$ and $d = \lambda - \gamma MA$. The predictions of (14) provide an excellent mirror of experimental results of *in vitro* systems recorded by Knight (1987) and Matis *et al.* (1983), in which proliferation rates of T cells were measured for varying concentrations of antigen and macrophages at various points in time. It produces a convex curve, and mimics various other features of the experimental data (Swinton *et al.* 1994), which provides support for the basic assumptions of the model and the two-signal hypothesis.

A slightly different, and somewhat more complicated approach, is to construct models that mimic a distribution of contacts between T cell and antigen (or APC) of defined type. One reason for taking this approach is the observation that when both the antigen-specific and the costimulatory signals are presented to a T cell, it synthesizes IL-2. This

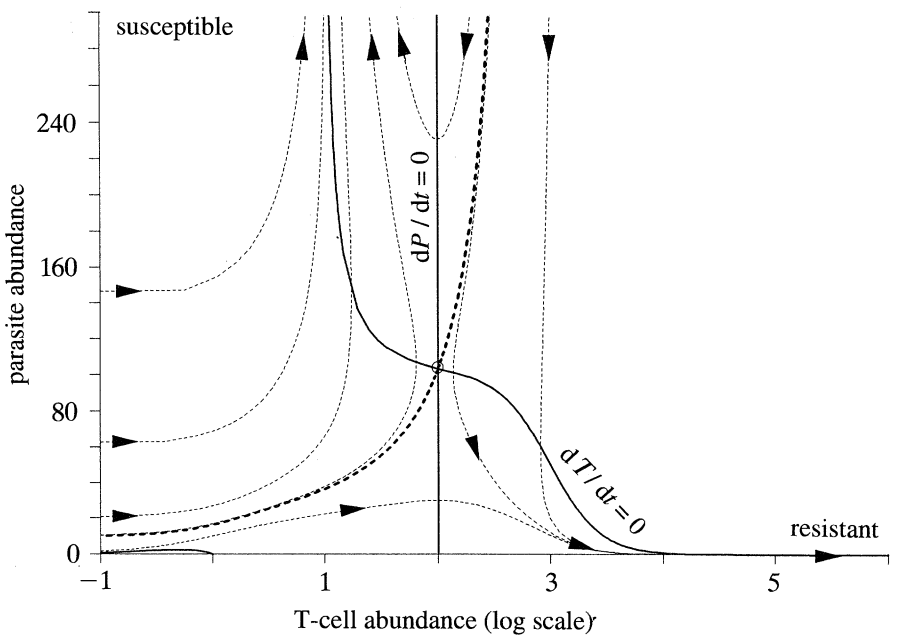

Figure 16. Phase plane derived from a two equation model of parasite abundance (P) and T cell abundance (T – T cells specific to parasite antigens) described by Schweitzer & Anderson (1992) for a parasite that replicates within its host. The solid lines represent isoclines ($\mathrm{d}P/\mathrm{d}t = 0, \mathrm{d}T/\mathrm{d}t = 0$), while the heavy broken line represents a manifold between two areas of attraction (i.e. susceptibility or immunity). The lightly broken lines record numerical solutions of the model with varying initial conditions (i.e. values of P and T at time $t = 0$) (see Schweitzer & Anderson (1992) for details).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

cytokine binds to receptors on the T cells surface and triggers cell division (autocrine proliferation). Activation of a cell to proliferate may be quantified by measuring the expression of markers directly associated with the capacity to divide (such as IL-2 receptors) or known to be upregulated during activation (i.e. CD2 and CD25 on HA1.7 T cells; Cerottini & MacDonald 1989; O'Hehir & Lamb 1990). Of greatest significance, however, is the fact that receptor expression (either of CD2 or IL-2 receptors) tends to vary from cell to cell (Hetzel 1993), with the overall distribution of receptors expressed per cell being highly heterogeneous (i.e. variance $\gg$ mean) for fixed exposure conditions (figure 17). This heterogeneity may have important implications for proliferation rates if the number of receptors expressed is positively associated with the likelihood that an activated cell divides. This is an area for future theoretical studies using mathematical frameworks that record the numbers of contacts with antigen or APCs and assume that proliferation rates depend not just on the receipt of two signals but on the distribution of signals received by the population of cells. Experimental studies are beginning to provide information on which to initiate the investigation of such models (Hetzel 1993).

(*b*) *Malaria and the human immune system*

A slightly different approach to that outlined above is to consider in detail one specific infection for which data exists on the time course of both infection and immunological responses in individual hosts. Mouse-malaria models provide such an opportunity, particularly in the light of recent work using mice with mutations that knock out or impair particular arms of the host's immunological defences. Since virtually all compartments of the mouse immune system are activated during infection with malaria (e.g. *Plasmodium chabaudi chabaudi*), the ability to knock out or delete/deplete particular facets of the response provides a powerful tool for assessing the components that have the greatest impact on the outcome of infection (Langhorne *et al.* 1990). CD4+ T cells are thought to play the major role in protective immunity against the erythrocytic stages of the parasite (Weid & Langhorne 1993).

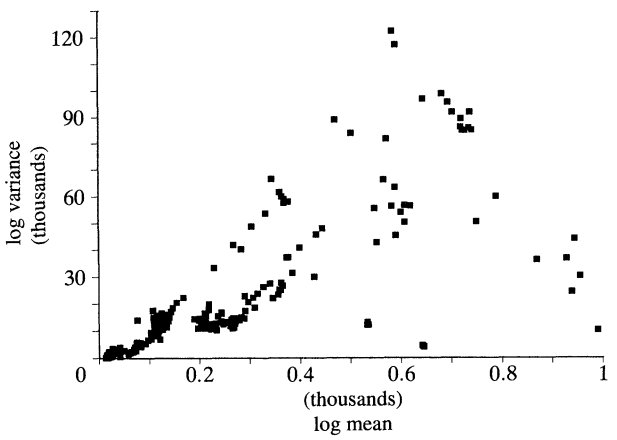

Figure 17. Recorded distributions of CD2 T cell receptors expressed on the surface of the cell represented as a log variance-log mean plot (from Hetzel 1993).

Before turning to the experimental studies, a simple model of merozoite invasion of enthrocytes can be formulated as follows (Anderson *et al.* 1988; Hetzel & Anderson 1994). We define X as the density of uninfected cells, Y as the density of infecteds and M as the density of merozoites. The model takes the form

$$dX/dt = \Lambda - \mu X - \beta XM \tag{15}$$

$$dY/dt = \beta XN - \alpha Y \tag{16}$$

$$dM/dt = \alpha rY - \delta M - \beta(X + Y)M \tag{17}$$

Here β is the per capita rate of erythrocytre invasion, Λ is the production rate of uninfected cells by the bone marrow, μ is their natural death rate, α is the death rate of infected cells ($\alpha > \mu$) due to parasite-induced rupture (calculated from the reciprocal of the length of the schizogenic cycle), r is the average number of merozoites released per infected cell and δ is the mortality rate of merozoites (note that unlike the model of Anderson *et al.* (1988), it is assumed that merozoites can also invade already infected cells (Hetzel & Anderson 1994)). The system has two stable points, one at Λ/μ with no parasites present ($\bar{M} = \bar{T} = 0$) and one where the parasite persists at

$$\bar{X} = \frac{\beta r + \alpha\delta}{\beta(\mu + \alpha(r-1))}, \tag{18}$$

$$\bar{Y} = \frac{\beta\Lambda(r-1) - \mu\delta^2}{\beta(\mu + \alpha(r-1))}, \tag{19}$$

$$\bar{M} = \alpha r\bar{Y}/[\delta + \beta(r-1)]. \tag{20}$$

The non-trivial point with the parasite persisting is locally stable provided

$$\Lambda/\mu > \delta/[\beta(r-1)]. \tag{21}$$

The basic reproductive number of the parasite (= average number of secondary cases of infected cells generated by one primary case in a population of susceptible cells) is

$$R_0 = [\beta(r-1)\Lambda/\mu]/\delta. \tag{22}$$

In the absence of an immunological response the critical parameters that determine persistence are the life expectancy of the merozoite ($1/\delta$), the rate of infection (β) and the number of merozoites produced per infected cell (r). For a mouse-malaria laboratory system (*P. berghei*) it is possible to assign parameter values for all the rate constants (Hetzel 1993) and hence numerical studies can generate a predicted temporal pattern in the numbers (or percentage) of infected cells in a mouse over time post infection (figure 18). Given the simplicity of the model there is surprising agreement in qualitative terms between the pattern generated by the model and that observed in *in vivo* time series (Hetzel 1993). This suggests that the initial dynamics of infection in experimental systems using naive mice is largely driven by cell infection processes with immunological responses playing a subordinate role. The immune clearance of para-

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

sitized cells or merozoites is not required to explain the initial oscillation in infected cell density. However, shorter and longer term there are important differences between theoretical prediction and observation, such as the overall percentage of cells infected (the model predicts higher values than those observed) and the timing and magnitudes of the peaks and troughs in the damped oscillations to a stable point.

The model can be extended to include a simple representation of the host's immunological attack on the parasite. For example, the terms γYT and σMT can be subtracted, respectively, from (16) and (17), to mimic CD4+ T-cell mediated removal of infected cells and merozoites. These cells are believed to be an essential component of the protective immune response to mouse malarias (Cavacini *et al.* 1986; Suss *et al.* 1988; Langhorne *et al.* 1990). An additional equation is needed to mimic T cell activation and proliferation and a very simple structure is assumed to keep the complexity of the four-equation system to manageable levels (Hetzel 1993):

$$dT/dt = aM + bY + f(T). \quad (23)$$

Here a and b denote activation rates from contact with merozoite and infected cell antigens (respectively) and $f(T)$ denotes a saturation function to prevent the runaway of cell proliferation.

Numerical studies of this expanded system reveal some interesting properties as depicted in figure 19. For certain parameter assignments (little is known about the magnitudes of those defined in (23)) the model closely mirrors observed trends both in normal mice with a full complement of T cells and in T cell-depleted mice (Langhorne *et al.* 1990). Most significantly, the model with T-cell mediated immunity mimics a rapid rise and fall in infected cell abundance and small 'recurrences' following apparent recovery: a pattern in excellent agreement with observation (figure 19). The model without T-cell mediated attack predicts a rapid rise, a less rapid fall and stabilization at moderate levels of cell infection. This is again in general agreement with observation in T cell-depleted mice. Interestingly, the theoretical predictions were made a few years before the experimental work was published (Anderson *et al.* 1989; Langhorne *et al.* 1990).

This is a nice example of a good accord between theory and experiment in highlighting the importance of cell mediated immunity in the temporal dynamics of infection within the host. It may seem surprising that such a simple model captures the observed pattern so well, but this example does highlight the significance of dynamical studies to the interpretation of experimental studies of infection and immunity. Prediction can be extended beyond the range of

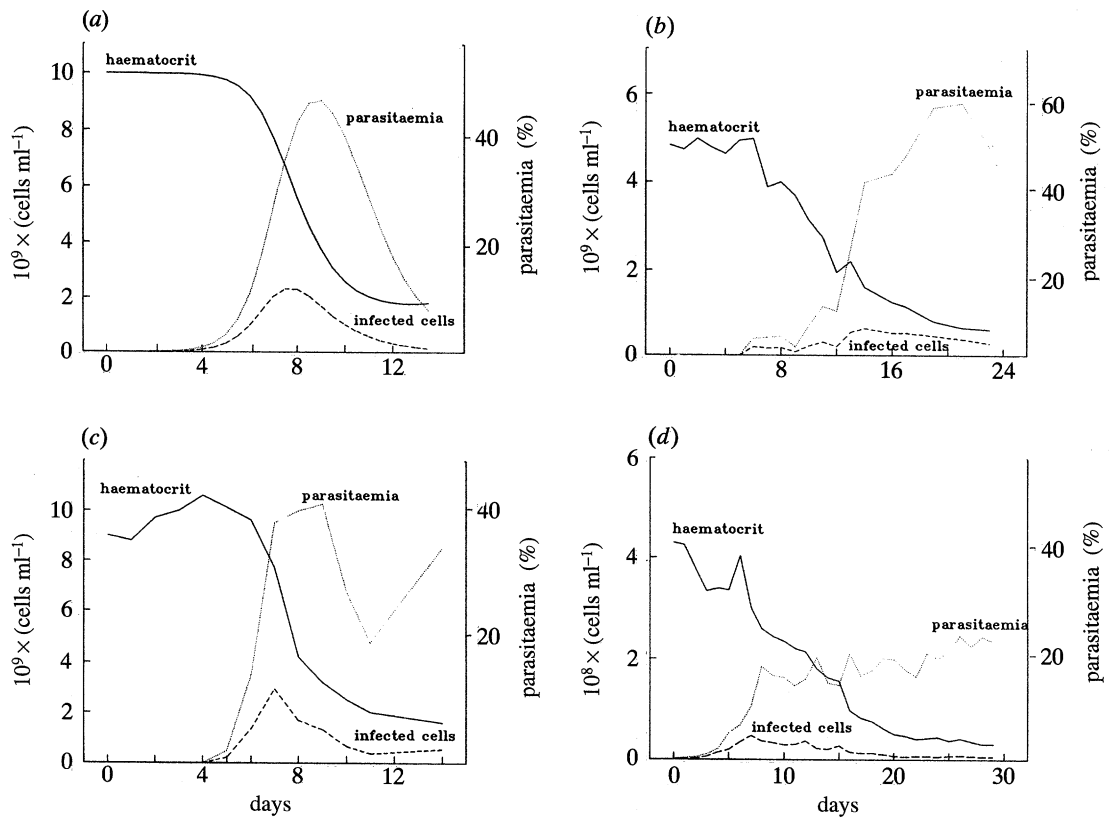

Figure 18. Predicted (graph (*a*)) and observed (graphs (*b*), (*c*) and (*d*)) time courses of *Plasmodium berghei* infection in mice, recorded as temporal changes in the haematocrit (density of red blood cells), the number of infected cells and the percentage of infected cells. Graph (*a*) records the predictions of the model defined by equations (15) to (17) (Hetzel 1993) while the data plotted in graphs (*b*) to (*d*) is from Roth & Herman (1979), Eling *et al.* (1977) and Dearsley *et al.* (1990), respectively. The model simply captures cell infection and death; no immune responses are incorporated.

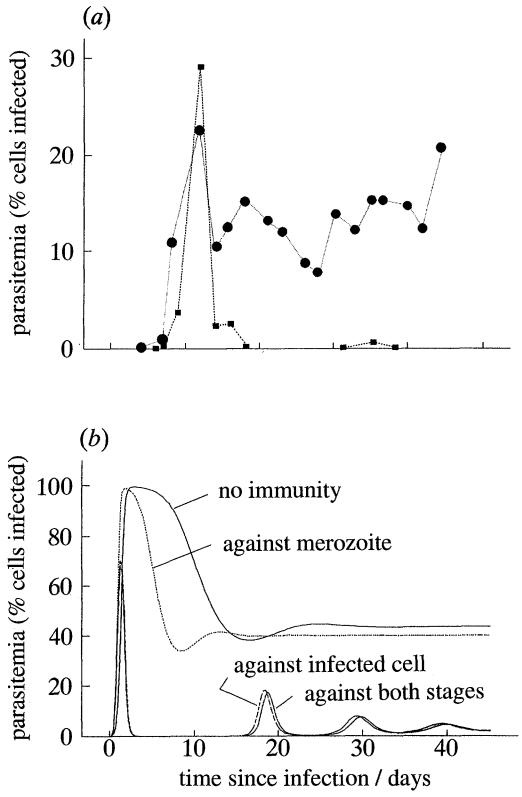

Figure 19. (*a*) Course of *Plasmodium chabaudi* infection in normal C57BL/6 mice (solid squares) and in mice depleted of T cells on day 7 (solid circles) (from Langhorne *et al.* 1990). (*b*) The predictions of the model defined by equations (15), (16), (17) & (23) where the target of the immune response is altered in the different simulations (Hetzel 1993; Anderson *et al.* 1988).

current experimental results to examine the relative effects of immunity directed at merozoite antigens versus that directed at infected red blood cells. As illustrated in figure 19, the model tentatively suggests that the latter is much more important in parasite clearance (Anderson *et al.* 1989). This prediction has obvious implications for research on the development of vaccines to protect against malaria.

(*c*) *Antigenic variation within the host*

Many infectious agents attempt to evade the host's immunological defences by the production of variants that differ antigenically from their predecessors (i.e. repeated changes in 'appearance') during the course of infection in an individual host. As a strategy for evading host immune responses it depends on the variation occurring in antigens whose recognition is involved in protection. At the molecular level three main mechanisms induce variation, namely: mutation (e.g. in HIV), recombination (e.g. malaria and some viral infection) and gene switching (e.g. African trypanosome (*Trypanosoma gambiense*), gonorrhoea (*Neisseria gonorrhoea*) and malaria (*Plasmodium falciparum*) (Vickerman 1986; Roberts *et al.* 1992)). At the between-host level antigenic variation may permit the repeated infection of individual hosts (discussed in the following section). At the within-host level, however, it acts to prolong the typical duration of infection which may facilitate transmission between hosts (i.e. reproductive success).

The past few years have seen a growth in interest in models of multi-strain pathogens and their interaction with the human immune system (Agur *et al.* 1989; Reibnegger *et al.* 1989; Nowak *et al.* 1990, 1991; Antia *et al.* 1993, 1994*a*,*b*). In this section, three particular problems are examined, namely, the interaction with specific and non-specific immunological responses (to particular antigenic variants), the evolution of virulence, and the impact of HIV variants on the human immune system.

(i) *Variant specific and variant cross-reactive immune responses*

The first task in considering the dynamical interaction between many strains and the immune system is to develop a model with variant specific and variant cross-reactive immunological responses. A recent paper by Antia *et al.* (1994*a*) addresses this problem in the following manner. The variables p_i, x_i and z are defined, respectively, as the population size of parasite variant i, T cells specific to the antigens of variant i and T cells cross-reactive to antigens expressed by all variants. Equations for the three variables are constructed as follows:

$$\mathrm{d}p_i/\mathrm{d}t = r_i f_1(P, C) - kp_i x_i - k' p_i z + M(P) \qquad i = 1 \ldots n, \tag{24}$$

$$\mathrm{d}x_i/\mathrm{d}t = sx_i f_2(p_i) - \mu x_i \qquad i = 1 \ldots n, \tag{25}$$

$$\mathrm{d}z/\mathrm{d}t = s' z f_3(P) - \mu' Z. \tag{26}$$

In (24) r_i denotes the replication rate of parasite strain i, k defines the killing rate by specific immunological responses, k' the death rate by cross-reactive responses and f_1 defines a saturation function for the growth rate of parasite strain i which is dependent on the 'carrying capacity' of the host C and the total parasite population ($P = \Sigma p_i$) (e.g. $f_1 = (1 - P/C)$). Similarly in (25) s denotes the proliferation rate of specific T cells, μ is their death rate and f_2 is a saturation function ($f_2 = p_i/(P_i + \phi)$, where ϕ is a constant). The third equation contains a proliferation rate s' for cross-reactive T cells, a saturation function f_3 ($f_3 = P/(P + \phi')$) and a death rate μ' of cross-reactive cells (Antia *et al.* 1994*b*). In numerical studies of the properties of this system of equations new variants are generated by a stochastic term where the probability of the appearance of a new variant q is assumed to be proportional to the total parasite density (the term $M(p)$ in (24)).

The dynamical properties of this system are complex and hence only brief mention is given to some of the most interesting features. In the equilibrium situation with n variants present, the outcome depends on whether or not the number of variants is greater than or less than a critical value n_c given by

$$n_c = [(\mu'\phi')/(s' - \mu')]/[(\mu\phi/(s - \mu)]. \tag{27}$$

When the number of variants is less than n_c the

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

parasite equilibrium (when one exists) is controlled largely by variant-specific immunity. Conversely, when the number of variants exceeds n_c than at equilibrium the parasitaemia is controlled principally by cross-reactive responses. The reason for this is that we assume that the strength of cross-reactive responses is considerably less (per capita) than specific responses. As such the latter work well when diversity is low but as it increases cross-reactive responses come increasingly into play to regulate pathogen abundance (Antia *et al.* 1994*b*).

The transient dynamics to an equilibrium is also of interest. When the rate of generation of antigenic novel forms is high, with only variant specific immunity (i.e. $\acute{s} = 0$), the parasitemia is high and fluctuates as a consequence of the emergence of a new variant and its subsequent control or clearance by specific responses. When both responses are present similar patterns pertain with the rapid generation of new variants, their reproduction and their clearance initially by variant-specific immunity, and later by both forms of immunity. The picture becomes much more complex once the parameters of growth and removal vary between variants. Simulation reveals great variability in the level of infection and its total duration (figure 20). Depending on the rates of generation and loss of variants, the parasite may or may not be cleared from the host. Three points are highlighted by these studies. First, the obvious conclusion that antigenic variation prolongs the typical duration of infection. Second, once chance factors play a role in the time of emergence of new variants and their biological properties such as growth rates within the host, the level of infection and its typical duration may vary widely between hosts (i.e. distributed incubation and infectious periods). Third and finally, as cross-reactive responses build up (more slowly than specific ones) it becomes more and more difficult for new antigenic variants to establish a significant level of infection within the host. These general predictions mirror, in a qualitative sense, what is observed in antigenical variable pathogen populations within the host such as malaria and the trypanosomes (Vickerman 1986; Barry & Turner 1991; Roberts *et al.* 1992).

(ii) *Virulence*

The once dominant belief that successful parasites evolved to be harmless to their host (why destroy your habitat?) has now been replaced in evolutionary circles by the view that selective pressures may drive pathogen virulence in a number of directions depending on the precise relationship between virulence and transmissibility (Anderson 1981; Levin & Pimentel 1981; Anderson & May 1982*b*; Ewald 1983). A key set of observations in support of this view is provided by the history of the interaction between the myxoma virus and its rabbit host, where selective pressures acting on both parasite and host resulted in strains of intermediate virulence becoming dominant both in Australia and England (Fenner & Ratcliffe 1965). In the present context of models of within host dynamics it is simply relevant to note that recent work

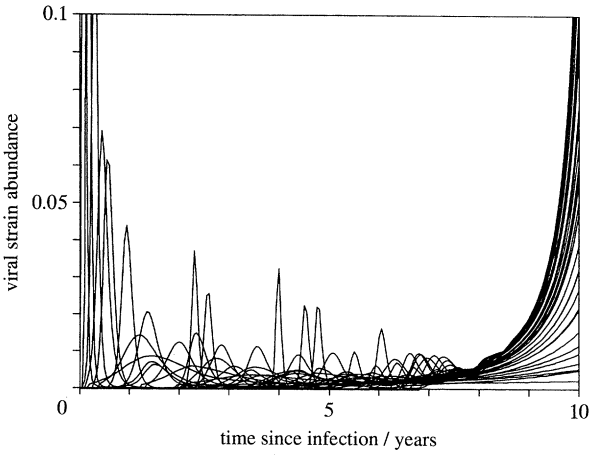

Figure 20. A simulation based on equations. (28)–(30) of changes over time in the abundance of 40 different HIV-1 'escape mutant' variants (arbitrary units) that evolve during the course of infection (from Nowak *et al.* 1990). Note the peaks in viraemia at the beginning and end of the long incubation period of AIDS (see text).

in this field using models that relate transmissibility to the integral of parasite density over its duration of stay in the host, again suggests that transmission success (= fitness) reaches a maximum for parasites with intermediate, rather than high or low virulence (Antia *et al.* 1993, 1994*a*,*b*).

(iii) *The pathogenesis of HIV*

A specific and very typical example of antigenic variation within the human host is provided by the human immunodeficiency virus (HIV-1). The tremendous antigenic diversity of the virus (Holmes *et al.* 1992), particularly in the V3 loop of the gp120 envelope protein (the major region susceptible to neutralization by antibodies), has been a major stumbling block in the development of effective vaccines. This is a consequence of the high mutation rate associated with viral replication: the rate of nucleotide misincorporation is of the order of more than 10^{-4} per base per cycle and for a genome of 10^4 bases this implies an error rate of more than 1 base per genome per metamorphic cycle (Dougherty & Trenin 1988). Errors are more likely to occur in some parts of the genome ('hot spots') such as the gene encoding for gp120.

One of the major puzzles of HIV infection and the induction of serious disease (AIDS) is the very long and variable incubation period from infection to the diagnosis of AIDS. The median period is of the order of 10 years in developed countries (Hendriks *et al.* 1993) and perhaps somewhat less in poor communities in developing countries (Plummer *et al.* 1993). Over this period viraemia is initially high (the primary HIV infection phase) for a period of a few weeks to a few months, falls to a very low level for many years before rising to high levels once serious immunodeficiency is apparent (Levy 1993). Recent studies, using sensitive molecular techniques, have highlighted the fact that even during the asymptomatic phase there is a high burden of HIV in the lymphoid tissue both as

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

extracellular (trapped in the follicular dendritic cell network of the germinal centres) and intracellular virus (Embretson *et al.* 1993; Pantaleo *et al.* 1993). HIV is persistently replicating even over the asymptomatic phase, but as disease progresses viral burden and levels of replication increase (Fauci 1993*a*).

A number of hypotheses have been put forward to explain why HIV induces the collapse of the immune system. These include, the cytopathic properties of the virus (i.e. direct killing of CD4+ cells), the role of cofactors (which in conjunction with HIV induce immunodeficiency), induction of apoptosis of CD4+ cells, molecular mimicry (generating autoimmune response to key molecules such as HLA regions), autoimmunity in general (presence of autoantibodies in HIV infection) and progressive imbalance in Th1-Th2 type immune responses (all recently reviewed by Fauci (1993*a*) and Levy (1993)).

Probably the least popular of these (perhaps because of its simplicity) is the notion that killing of CD4+ combined with antigenic variation is sufficient to explain the observed trends in viraemia and CD4+ cell decline in HIV infected patients.

To examine this problem in more detail we require a particular case of the model defined by (24–26) (a simpler version with no saturation on virus or CD4 cell replication) in which the pathogen kills (and replicates within) the immune system cells (CD4 + T cells) responsible for orchestrating the immunological defences against the pathogen. A series of recent papers have addressed this problem (Nowak *et al.* 1990, 1991; Nowak & May 1993). A simple but appropriate model has four variables: v_i, y, x_i and z, which denote the densities of virus strain i, total CD4+ cells specific to strain i (mounting specific responses to strain i), and CD4+ cells that mount cross-reactive responses to all strains, respectively. Where V is the total virus population ($V = \Sigma v_i$), the rates of change of each variable with respect to time ($t = 0$ being the point when an individual acquires infection) are then:

$$\mathrm{d}v_i/\mathrm{d}t = f_i(v_i, y) - v(k_i' + k_i x_i) \qquad i = 1, \ldots n, \qquad (28)$$

$$\mathrm{d}x/\mathrm{d}t = \gamma v_i - \mathrm{d}x_i - uVx_i \qquad i = 1, \ldots n, \qquad (29)$$

$$\mathrm{d}z/\mathrm{d}t = \gamma' V - bz - uVz. \qquad (30)$$

The term $f_i(v_i, y)$ denotes the reproductive rate of viral strain i (e.g. $f_i(v_i, y) = (\acute{r}_i + r_i y)v_i$ to denote replication by strain i at a per capita rate $r_i y$ arising from infection with CD4+ cells (y), along with a constant background replication rate $\acute{r}_i$ to denote replication of the virus in cells other than of the CD4+ type (such as macrophages and follicular dendritic cells)). As before the term $k'v_i z$ represents killing by cross-reactive responses and the term $kx_i v_i$ represents killing by specific responses. For the total population of CD4+ cells (y) we may define K as the recruitment rate of CD4+ cells, d is the per capita death rate and uVy denotes the rate at which cells are killed by any member of the total virus population. Equivalently, the recruitment term in (29) and (30) ($\gamma v_i, \gamma' V$) denote activated cells joining the strain-specific and cross-reactive CD4+ cell population. These cells are killed by the virus at net rates uVx_i and uVz and a natural mortality rate of d also removes cells. The total number of virus strains, n, is not constant, because replication errors generate new mutants that escape ('escape mutants') the current strain-specific immune responses and persist in the presence of cross-reactive responses. This introduces the stochastic element where the probability that a new strain is generated in the time interval t and $t + \mathrm{d}t$ is given by $qv(t)\mathrm{d}t$ (Nowak *et al.* 1990). The key properties of this model can be understood by analytic and numerical studies. Specifically, from (28) to (30) we may obtain an equation for the rate at which the total virus population changes over time (ignoring reproduction of virus within cells other than the CD4+ type):

$$\mathrm{d}V/\mathrm{d}t = V\left[r - \frac{k\gamma V}{(d + uV)}D - \frac{k'\gamma' V}{(d + uV)}\right]. \qquad (31)$$

To obtain this equation, Nowak & May (1993) assumed that x_i and z converge to their steady state on time scales short compared with those on which the total virus population changes. The term D denotes the Simpson index ($D = (v_i/V)^2$, which is an inverse measure of viral diversity where $D = 1$ if only one strain is present and $D = 1/n$ if n strains are present in equal abundance. D is always between 0 and 1; it is the probability that two viruses chosen at random belong to the same strain. The model has three distinct parameter regions corresponding to three qualitatively different courses of infection. These are as follows. (i) No asymptomatic phase and the virus population rapidly expands to very high levels ($ru > k'\gamma' + k\gamma$). Here the virus 'outruns' both the specific and non-specific responses by its high replication rate (r) and high cytopathic properties (u). (ii) Chronic infection but with no collapse of the immune system ($k'\gamma' > ru$). Here the strain specific responses are sufficient to control replication but the total viral population is not eliminated but persists at low levels over the life span of the host (with low levels of CD4+ suppression). (iii) Chronic infection and the collapse of the immune system (CD4+ cells $\rightarrow 0$) after a long incubation period ($k'\gamma' + k\gamma > ru > k'\gamma'$). In this case the strain-specific responses play an important role but they are unable to control replication by themselves, requiring the help of the cross-reactive responses that build up more slowly over time. Eventually, however, the slow loss of CD4+ cells (the cytopathic properties of the virus) results in all variants being able to eventually grow once the viral diversity breaches a critical value D_c where

$$D_c = (ru - k'\gamma')/(k\gamma). \qquad (32)$$

This is the viral diversity threshold theorem put forward by Nowak *et al.* (1990, 1991) to explain the pathogenesis of HIV and the long and variable incubation period of AIDS.

An illustration of the outcome under the conditions defined under (iii) above is presented in figure 21. Since stochastic elements are involved in the timing of

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

the emergence of new variants, the model generates a variable incubation period (viraemia attains a defined very high level in the patient and CD4+ cell abundance falls below a defined level) as observed, with the gamma distribution as the theoretical prediction of the distribution of times to AIDS (Nowak & May 1993). A further prediction of interest concerns the viral diversity over the incubation period. Due to the continual emergence and establishment of new escape mutants, the absolute number of strains is predicted to rise with time from infection. However, once the diversity threshold is exceeded, strains with faster replication rates would predominate in the rapidly growing population and hence the model predicts that diversity as measured by the Simpson's index will decline as AIDS develops due to sampling issues (Nowak *et al.* 1991). Before turning to the degree of agreement between prediction and observation, it is important to consider precisely what is implied by the diversity threshold. At face value the implication is clear: antigenic diversity is a cause not a consequence of the development of AIDS. However, this notion must be interpreted with care. The diversity is in essence a measure of the impact over time of the total viral population on the abundance of CD4+ cells. As more and more escape mutants arise over time they progressively deplete CD4+ cells. Hence breaking the diversity threshold can also be interpreted as CD4+ cell abundance falling below the desired level at which the specific and cross-reactive responses are able to control or suppress the replicative abilities of this antigenically diverse population.

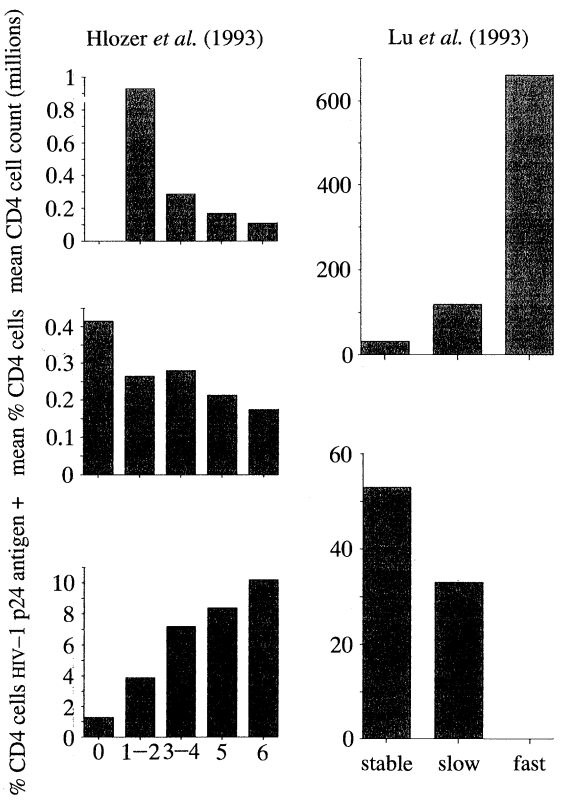

Figure 21. Left-hand side: changes in mean CD4 cell count, mean % of CD4 cells and the % of CD4 cells with HIV-1 p24 antigen (% infected) in patients at different stages of the incubation period of AIDS as defined by the Walter Reed stage of HIV-1 disease (from Holzer *et al.* 1993). Right-hand side: the relationship between the density of HIV-1 infected cells and serum neutralizing activity to virus antigens, and the rate of decline in CD4 cell abundance over a 2 year period (from Lu *et al.* 1993).

The model defined by (28) to (30) has a rich array of possible dynamical behaviours, despite its all too obvious simplicity. How well does it mimic observed patterns in infected patients? All three patterns of behaviour can be seen, not only in HIV-1 infected patients but also in SIV infection in various non-human primates (Nowak & May 1993). Focusing only on humans and HIV-1, what is clear at present is that some patients have not progressed to AIDS despite very long periods of infection (15 years), others have progressed very rapidly where viraemia fails to fall significantly after primary HIV infection, while the majority seem to progress to AIDS on an average time scale of 10 years (Fauci 1993*a*,*b*; Levy 1993). A large number of studies suggest that the biological properties of the viral strains present in a patient are associated with the rate of progression to AIDS (Levy 1993). The notion that disease is associated with viral virulence (i.e. replication rate or cell invasion rate) is again in accord with the prediction of the model. Under case (i) above, if the replication rate is high enough, the total virus population outruns the immune system and severe immunodeficiency will rapidly appear. The timing of the emergence of these highly virulent (fast replicators) escape mutants (sometimes referred to as syncitia inducing variants) will be of major significance. As such as alternative view of the properties of the model is that, irrespective of the degree of viral diversity, the development of AIDS is simply related to the time that elapses from initial infection to when a viral variant emerges with a sufficiently high replication rate (r) or cell killing rate (u) to satisfy the condition defined in case (i) above (i.e. $ru > k'\gamma' + k\gamma$). Furthermore, the parameters k', γ', k and γ will be influenced by host genetic background. This may explain what seem to be differences in the median incubation period in Africa and in Europe plus North America (it could also be related to the viral quasispecies in these different areas).

Other predictions of the model mirror observed patterns. For example, the model predicts that CD4+ cell abundance and its rate of decline will be directly related to viral load with the total density y decaying slowly over time. Recent work by Connor *et al.* (1993), Holzer *et al.* (1993) and Lu *et al.* (1993) supports this prediction. Yet others are as yet less well supported by data, in particular the prediction that viral diversity (as measured by an index such as Simpson's D, which takes account of frequency and the likelihood of sampling) will slowly rise and then decay as AIDS develops. Some data sets are supportive; others are not (Nowak *et al.* 1990; Nowak & May 1993). Acquiring such information over many years,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

and with sufficient replication of sampling at one point in time is difficult. However, a recently developed technique, namely, that of the DNA heteroduplex mobility assay, offers much hope for speeding up analyses of viral diversity both within and between patients (Delwart *et al.* 1993). In recent, as yet unpublished, studies there is a hint that viral diversity is low in those who progress rapidly to AIDS and relatively higher in the non-progressors. If those early results are confirmed two different explanations may underpin the observations. One is simply related to sampling. If a fast replicating strain dominates in fast progressors, simple sampling theory would indicate low diversity unless very large numbers of samples were analysed. Conversely, in slow progressors where no fast replicating strain dominates, the likelihood of sampling many strains is higher. An alternative explanation rests on the idea of breaching the condition defined under case (i) above, irrespective of viral diversity. The emergence of a fast replicator which satisfies the condition would lead to AIDS, irrespective of how many other strains are present.

More broadly, the need at present is for more detailed studies of viral diversity, and the associated biological properties of the variants present, over the long incubation period of AIDS. On the theoretical front, there is also a need for further refinement and elaboration. An obvious area is a more careful analysis of the impact of HIV infection on antigen presenting cells such as macrophages and follicular dendritic cells. A preliminary study by Frost & McLean (1994) suggests that a simple model of HIV-induced destruction of follicular dendritic cells can mimic observed patterns of viraemia, CD4+ cell decline and the long incubation period of AIDS. This idea needs to be investigated more carefully with models that also address antigenic variation and HIV-induced killing of macrophages.

5. PATHOGEN GENETIC VARIATION AND BETWEEN-HOST TRANSMISSION

Much recent attention has been focused on parasite genetic variability within human communities (Forsyth *et al.* 1988; Delwart *et al.* 1993; Moxon *et al.* 1994; Gupta *et al.* 1994*a*). In part this is a consequence of a variety of new molecular techniques that facilitate the study of infection agent abundance and the strains present in different hosts (e.g. polymerase chain reaction, and DNA heteroduplex mobility assays). In addition to the collection of a burgeoning data base on pathogen variability, these advances have stimulated a natural extension of mathematical studies in epidemiology to address the transmission dynamics of multiple strains of a pathogen (eliciting specific and cross-reactive immunological responses) within given host communities (e.g. Anderson 1981; Levin & Pimental 1981; Anderson & May 1986; Nowak & May 1993, 1994*a*,*b*; Gupta *et al.* 1994*b*). A particular focus has been the evolution of virulence and the factors that determine the concomitant persistence of many strains within the host population (May & Anderson 1979; Bremermann & Thieme 1989; Seger 1988; Seger & Hamilton 1988; Frank 1992; Lenski & May 1993). The early models tended to exclude the possibility of superinfection, where an already infected host can be infected by a different strain. A few studies have addressed this problem (Levin & Pimental 1981; Levin 1983; Anderson & May 1986) in the context of two strains where the more virulent strain can take over the host infected by a less virulent one. The definition of virulence in some of these studies is vague but it normally refers to replication rate within the host which is assumed to be positively correlated with pathogen-induced host mortality and transmissibility.

Recently, however, a major advance in the theoretical work has been the study of models that address the dynamics of many strains circulating in the host population at the same time (Nowak & May 1993). In this brief discussion of the problem, a general model of Nowak & May (1993) is first examined before moving to the specific examples of HIV infection and falciparum malaria in human communities.

Before turning to the general case, the predictions of simple models with no superinfection are worth reiterating. Without superinfection these models represent simple competition for susceptible hosts. A number of studies have shown that the parasite with the highest basic reproductive rate, R_0, always wins (Anderson & May 1986; Diekmann *et al.* 1990). This suggests that evolution will tend to maximize R_0. In the absence of any association between pathogenicity (the pathogen-induced host death rate) and transmissibility, the later will be maximized and the former minimized (i.e. the conventional wisdom that evolution will select for parasites that do not harm their hosts). However, as noted earlier once transmissibility is linked with pathogenicity, there will be an evolutionary stable degree of virulence (= pathogenicity) which maximizes R_0 at some intermediary value between high and low virulence (Anderson 1981; Levin & Pimental 1981; Anderson & May 1982*a*, 1983; Ewald 1993).

(*a*) *Superinfection and virulence*

To mirror the transmission of genetically variable infectious agents (e.g. Hepatitis B; *Plasmodium* sps., *Trypanosome* sps., HIV-1, HIV-2; *Neissera gonorrhoea* etc.) we need a theoretical framework which handles many different strains with different pathogenicites (virulence) and transmissibilities in a context where coinfection can occur. Allowing for this increases the complexity of the problem considerably. Hence a useful point of departure is to consider the case of superinfection where a more virulent strain (greater pathogenicity) can 'take over' a host who is already infected with a less virulent strain, but the host will, in effect, only contain one strain (the more virulent) at any one time. This is the approach adopted by Nowak & May (1994*b*). For a directly transmitted infection which does not induce fully protective immunity the model of the dynamics of susceptible hosts, x, and hosts infected with strain i, y_i is as follows:

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

$$dx/dt = k - ux - x\sum_{i=1}^{n}\beta_i y_i, \tag{33}$$

$$dy_i/dt = y_i\left[\beta_i x - u - \alpha_i + s\beta_i\sum_{j=1}^{i-1} y_j - s\sum_{j=i+1}^{n}\beta_j y_j\right]$$

$$i = 1, \ldots n. \tag{34}$$

To simplify matters host population growth is taken to be an immigration (rate k) − death (rate m) process. The term α_i denotes the pathogenicity of strain i (additional per capita death rate of infected host), β_i is transmission probability for strain i and s denotes the rate at which superinfection occurs relative to infection of uninfected hosts. If current infection induces a degree of immunity to reinfection (cross-reactive immunological responses) then $s < 1$, while if current infection enhances the likelihood of superinfection then $s > 1$. Note that no recovery from infection takes place. Nowak & May (1994*b*) assume a specific relation between transmissibility and pathogenicity of the form $\beta_i = a\alpha_i/(c + \alpha_i)$ such that at low virulence infectiousness increases linearly while at high virulence the infectiousness saturates. This gives a basic reproduction rate for strain i of

$$R_{0i} = ak\alpha_i/[m(c + \alpha_i)(m + \alpha_i)]. \tag{35}$$

Nowak & May (1994*b*) produce a series of analytic and numerical studies to reveal the very complex dynamical properties of this model. They show that superinfection has two major effects, namely: (i) it shifts average parasite virulence to higher levels than that which maximizes R_0 (i.e. $\alpha_{opt} = (cu)^{1/2}$); and (ii) it leads to coexistence between a number of different parasite strains with a range of pathogenicities (α min to α max) (figure 22). Interestingly, the latter observation is very much in accord with observed patterns, as for example, in the cases of the myxoma virus of rabbits (Fenner & Ratcliffe 1965), the influenza viruses of humans (Stuart-Harris 1982), and probably HIV-1 (Delwart *et al.* 1993). The former prediction can be explained intuitively as superinfection leading to intra-host competition among strains resulting in increased levels of virulence over that which would pertain to maximize R_0 in the absence of superinfection. Other properties of the model of general interest include the following: superinfection can maintain strains with very high virulence; it can generate very complicated dynamics such as heteroclinic cycles and it can induce sudden changes in the average level of virulence (Nowak & May 1994*b*).

The multi-strain framework is an important step forward in understanding the transmission dynamics of genetically variable infectious agents. Caution must be exercised, however, in the blanket application of the concepts that have emerged from the analysis of equations (33) and (34). Of greatest importance in this context is the interpretation of superinfection as a form of competitive dominance hierarchy among strains within the host where the most pathogenic always wins (Nowak & May 1994*a*,*b*). In addition, but of less general significance, a particular assumption was

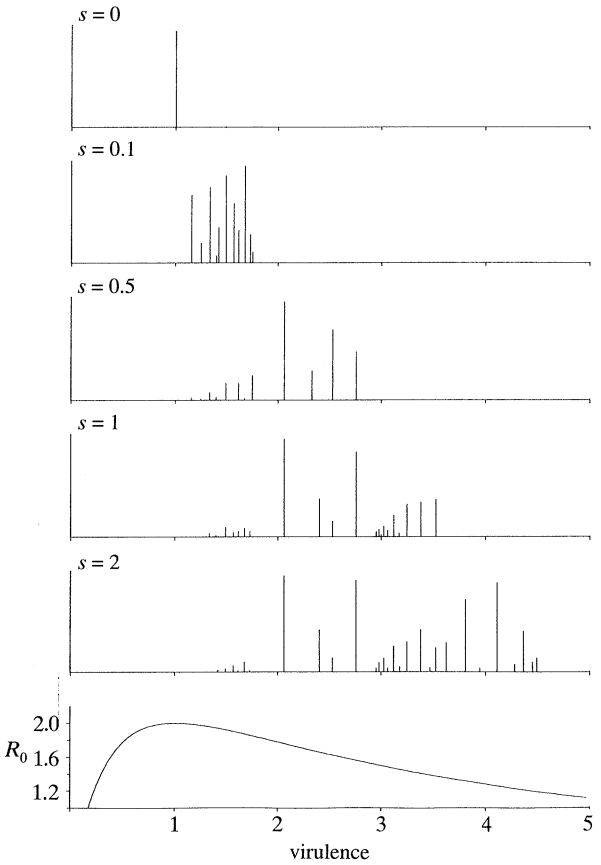

Figure 22. The equilibrium distribution of parasite strains with different virulences (from Nowak & May 1994*a*) generated by the model defined by equations (34) and (35). In the absence of superinfection ($s = 0$) the strain with maximum basic reproductive rate, R_0, is selected. With superinfection ($s \neq 0$), coexistence of many different strains is possible. The bottom graph records the assumed relationship between R_0 and virulence used to generate the equilibrium distribution of parasite strains. Note that when superinfection is possible selection may not maximize the value of R_0.

made relating pathogenicity and virulence. Lastly, for many infections of relevance in an epidemiological context, immunity to reinfection (a recovered class) is of relevance to the interpretation of observed patterns of infection in human communities. With this last issue in mind Swinton & Anderson (in preparation) have begun the study of a system of equations that keep track of multiple strains of an infectious agent in an age-structured human community. A particular focus of this study was the build-up of cross-reactive immunological responses as a person ages and accumulates experience of exposure to many strains of a particular parasite (Anderson 1994). The issue of virulence to the human host was not addressed.

In order to allow individuals to build up 'immunological memory' via cross-reactive responses we must keep track of how many times they have experienced infection with each strain, j. A system of equations can be formulated to represent chances over time, t, and with age, a, of the number of people who have been infected with strain j exactly i_j times, $y(t, a, i_1, i_2, \ldots i_n)$ (this can be simplified by writing

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

$\underline{i} = i_1, i_2, \ldots i_n$ with $y(t, a, i)$. Defining $\lambda(t, i)$ as the force of infection for people of type $\underline{i}$ with strain j then

$$\partial y/\partial t + \partial y/\partial a = \sum_j \lambda_j(t, \underline{i} - e_j) y(t, a, \underline{i} - e_j) - \sum_j \lambda_j(t, \underline{i}) y(t, a, \underline{i}) - \mu y(t, a, \underline{i}), \quad (36)$$

where e is a vector with 1 in the jth position and zero elsewhere (Swinton & Anderson, in preparation). This structure is somewhat similar to that employed by Kretzschmar (1989) for macro parasitic infections but with important differences in that the force of infection depends on the type of individual, $\underline{i}$. The force of infection is defined as

$$\lambda_j(\underline{i}) = p_j(t)\beta_j f\left(\sum s_{jk} i_k\right). \quad (37)$$

Here, $p(t)$ is the prevalence of people infected with strain j, β_j is the transmission probability of strain i and the function f defines the action of cross-reactive and specific immunity where s_{jk} defines the relative contribution of strain k in the strength of immunity against strain j (a linear decay or exponential decay function could be assumed). Introducing the generating function

$$u(a, \underline{z}) = \sum_{\underline{i}} \underline{z}^{\underline{i}} y(a, \underline{i}), \quad (38)$$

where $\underline{z}^{\underline{i}} = z_1^{i_1}, z_2^{i_2}, \ldots z_R^{i_n}$ then for a linear structure of the function f in (37) we obtain a partial differential equation for the generating function of the form

$$\partial u/\partial t + \partial u/\partial a + \sum_k g_k \partial u/\partial z_k + hu + \mu u + \sum_{jk} f_{jk} \frac{\partial^2 n}{\partial z_j \partial z_k} = 0. \quad (39)$$

Here

$$g_k(t, \underline{z}) = z_k \sum_j s_{jk} p_j(t) \beta_j (z_j - 1), \quad (40)$$

$$h(t, \underline{z}) = \mu - \sum_j p_j(t) \beta_j (z_j - 1). \quad (41)$$

Equations (39) to (41) enable explicit solutions to be obtained for special cases and for particular variables of epidemiological interest (i.e. the age distribution at equilibrium of the fraction never infected with any strain, or the total number of infections that have occurred in the population at time t). However, analysis of the properties of the model is incomplete at present and for epidemiological studies it is helpful to consider particular cases with a restricted number of strains.

(*b*) *The transmission dynamics of multiple strains of* HIV

HIV is a topical area to address with these multiple strain transmission models, particularly in the context of the evolution of virulence. A key question of public health significance is whether evolutionary pressures will drive the virus to higher or lower pathogenicity than that currently observed. The question is very complex given the different selective pressures that may operate on intra-host dynamics and inter-host transmission. Both are obviously linked but it is very unclear at present whether a short incubation period (2–3 years) to AIDS (i.e. the emergence of fast replicating strains that rapidly depress CD4+ cell abundance) is linked to higher transmission efficacy by comparison with a much longer incubation period (10–12 years). Transmission efficacy as measured by the basic reproductive rate, R_0, depends on the duration of the incubation period (D) times the average transmission probability (defined per partner) over that period (β). For simple models (not incorporating heterogeneity in sexual behaviour or variable infectiousness over the incubation period, Anderson *et al.* (1986), May & Anderson (1987), Anderson (1988))

$$R_0 = \beta m D, \quad (42)$$

where m is the mean rate of sexual partner acquisition. As illustrated in figure 23, rapid progressors (assuming that progression depends on the most virulent strain of the virus present in the host) would have to be very infectious over the much shortened incubation period, if the generation of secondary cases from such individuals is to exceed that arising from a slow progressor. It might seem reasonable to assess

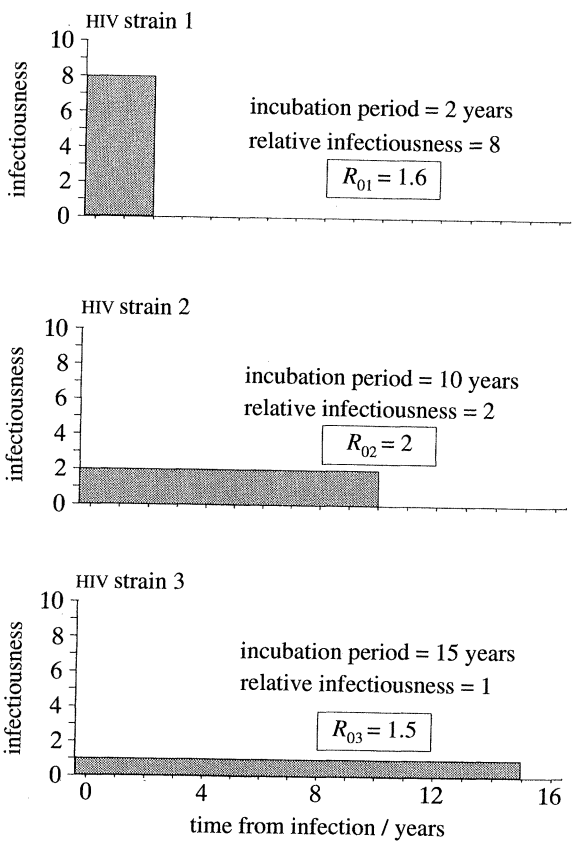

Figure 23. Diagrammatic representation of the interplay between the magnitude of R_0 and the degree of infectiousness (i.e. viraemia) over the incubation period plus the duration of the incubation period (short periods represent high virulence; long periods denote low virulence).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

average infectiousness over the incubation period as the integral of some measure of viral abundance in the host. However, even this assumption is open to question since infectiousness may depend not on viral abundance but on which genetic strains of the virus are present, and in what abundance. In other words if a person has n strains each of abundance $v_j(t)$ for strain j at time t since infection (assuming a single exposure event resulted in infection) and with relative infectiousness γ_j, then the overall transmission probability, β, of that person over an incubation period of T time units, will be

$$\beta = A \int_0^{T_n} \sum_{j=1}^{n} \gamma_j v_j(t)\, dt, \tag{43}$$

where A is a scaling constant. In other words as biological and epidemiological information on the properties of different viral strains grows it will be necessary to meld models of within host dynamics with those that address transmission between hosts in order to interpret observed patterns of infection and disease.

However, to start simply, consider a two-strain superinfection model where the more pathogenic strain (shorter incubation period) will replace the less virulent strain if superinfection occurs. If we ignore evolution within the patient and just focus on the transmission of the two strains a suitable model is as follows:

$$dx/dt = \Lambda - \mu X - \beta_1 y_1 X - \beta_2 y_2 X, \tag{44}$$

$$dY_1/dt = \beta_1 y_1 X - \beta_3 y_2 Y_1 - (\alpha_1 + \mu) Y_1, \tag{45}$$

$$dY_2/dt = \beta_2 y_2 X + \beta_3 y_2 Y_1 - (\alpha_2 + \mu) Y_2. \tag{46}$$

Here X denotes the density of susceptibles, Y_1, is the density of individuals infected with the less virulent strain, Y_2 is the density of people infected with the virulent strain and y_1 and y_2 are the proportions infected with each strain. The average time to death of people with strain 1 is $1/(\alpha_1 + \mu)$ and that of people with strain 2 is $1/(\alpha_2 + \mu)$, where $(\alpha_2 + \mu) > (\alpha_1 + \mu)$. The parameters β_1, β_2 and β_3 denote the transmission rates, (transmission probability $\times$ the mean rate of sexual partner change) of strain 1 to susceptible people, strain 2 to susceptible people and strain 2 to people infected with strain 1, respectively. The parameter Λ records the immigration rate to the sexually adult age classes and $1/\mu$ is the average duration of sexual activity. Looking at the equilibrium properties of this model two situations arise. First, if the more pathogenic strain has a lower R_{oi} than the less virulent strain, then coexistence is possible provided

$$\beta_3 > (\alpha_2 - \alpha_1) > (\beta_2 - \beta_1). \tag{47}$$

In other words, the transmission rate from strain 2 infected persons to strain 1 infecteds must be sufficiently large (but it can still be smaller (due to cross-reactive immunological responses) than β_2 and β_1). If (47) is not satisfied then the less virulent strain will persist and the highly virulent one will go extinct. However, in this latter case, the transient dynamics are revealing, given appropriate parameter assignments, as recorded in figure 24. In the two simulations shown in this figure, the top graph records the outcome (when $R_{02} < R_{01}$ and condition (47) is not satisfied) when both strains are present at $t = 0$. Note that the time to extinction of strain 2 is 150 years. Similar conditions prevail in the lower graph, but strain 1 is present first and strain 2 attempts to invade at year 20. The time to its extinction is shorter, but it still persists for a period of 80–100 years. A further simulation is recorded in figure 25 but with the condition for coexistence satisfied. Note that even small differences in the values of the basic reproductive rates can lead to wide differences in the prevalence of each infection. Furthermore, these will only emerge over very long periods of time (i.e. hundreds of years).

The second situation arises if the more pathogenic strain has a bigger reproductive rate ($R_{02} > R_{01}$). In this case coexistence can again occur provided

$$\beta_3 < (\alpha_2 - \alpha_1) < (\beta_2 - \beta_1). \tag{48}$$

If the rate of superinfection is zero, both strains can persist (in contrast to the situation for models of directly transmitted viral infections (Anderson & May 1986)), because the transmission term is a product of the numbers of susceptibles times the fraction infected. Once superinfection occurs they will continue to coexist provided the rate of superinfection is relatively low. If the input of new susceptibles in (44) is

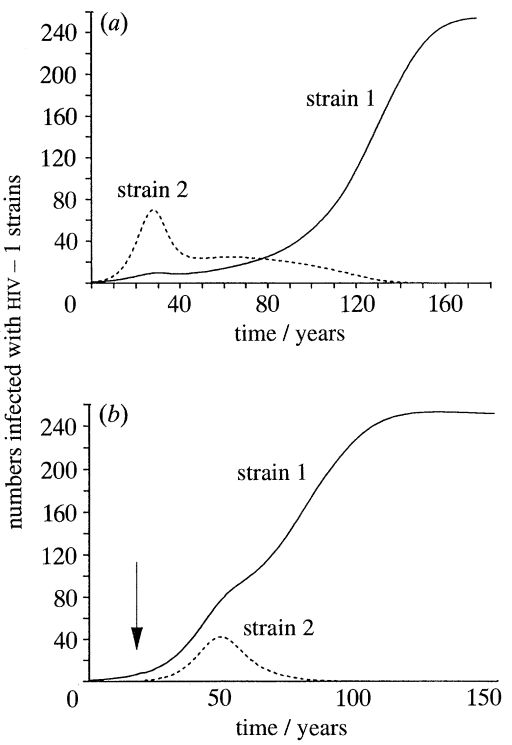

Figure 24. Simulations of the transient dynamics of competition between two HIV-1 strains (labelled 1 and 2), where strain 2 represents a highly virulent strain (but with $R_{02} < R_{01}$). In (*a*) both strains are present at $t = 0$, whereas in (*b*) strain 2 arrives at time $t = 20$ years. In both cases strain 2 eventually becomes extinct by year 150–180 (see text).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

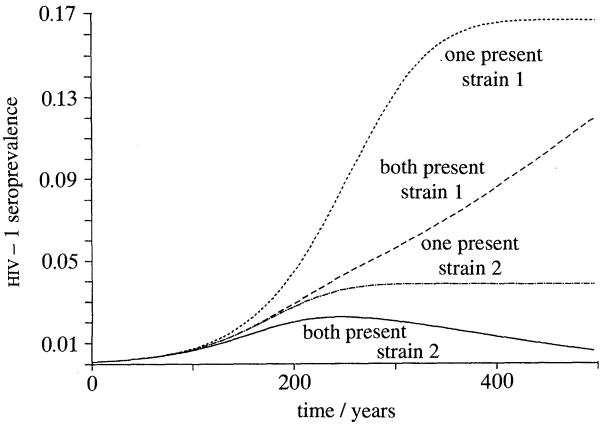

Figure 25. Similar to figure 24 but with the condition for coexistence satisfied. The prevalences of strains 1 and 2 are plotted when both coexist and when each is present on its own (see text).

replaced by a new birth term, aX, where a is the per capita birth rate (weighted by the fraction of newborns who survive to sexual maturity) then further conditions pertain if the two strains are to coexist (Anderson & May 1994). Specifically for the first case of $R_{01} > R_{02}$, condition (47) must be met but also the following condition

$$(\beta - \alpha_1) > a > (\beta_2 - \alpha_2). \quad (49)$$

For the second case, $R_{02} > R_{01}$, condition (48) must be satisfied and also

$$(\beta_1 - \alpha_1) < a < (\beta_2 - \alpha_2). \quad (50)$$

An interesting aspect of this model is that if coexistence occurs the system is neutrally stable. A special case of this system recently analysed by Lipsitch *et al.* (1994) is the situation in which superinfection does not occur ($\beta_3 = 0$). In this case (as suggested in (50)), population growth rate influences the outcome. Rising population growth rates increasingly favour the more virulent strain, moving from only strain 1 persisting, via coexistence, to only strain 2 (the more virulent) persisting.

An added complication, which is very relevant to the evolution of HIV, concerns the impact of vertical transmission from mother to newborn infant. In the case of HIV-1 the rate is high (of the order of 20%–35% of all offsprings born to infected mothers) in many developing countries (Ryder & Timmerman 1991; De Cock *et al.* 1994). A variety of theoretical studies have suggested that if a significant component of transmission is via the vertical route, then selective pressures should drive the pathogen to reduced virulence (Anderson & May 1981; Busenberg & Cooke 1993). However, it may lead to complicated dynamical patterns (for directly (non-sexually) transmitted infections) in which the value of R_0 is not optimized (Nowak 1991). In the context of HIV and sexually transmitted infections in general (which are often vertically transmitted) this is an area in need of further theoretical study.

Having noted that data on infectiousness and virulence of different HIV-1 strains is extremely limited, is it possible to compare the predictions of these very simple models with observed trends? The concomitant transmission of HIV-1 and HIV-2 in parts of West Africa provides an interesting example of multi-strain transmission (via the same route) and coinfection. In recent times the spread of HIV-1 through the area seems to have been much more efficient than that of HIV-2 (De Cock *et al.* 1994). The viruses are only distantly related with about 40% amino-acid homology in sequence information (the most conserved regions are the *gag*- and *pol*-encoded proteins). However, viral isolates of both viruses show great diversity in infectivity and replicatory ability in cell culture conditions. There are significant differences in the epidemiological patterns of infection of the two strains. In particular, cross-sectional studies reveal much higher prevalences of infection with HIV-2 in older age-class than HIV-1, with convex patterns over age classes for HIV-1 with peak prevalence in the 20–29 year old age classes and more monotonic growth to a stable prevalence in 55–59 year olds in the case of HIV-2 (figure 26). This could be a consequence of various factors including longer periods of spread of HIV-2 in West Africa by comparison with HIV-1 in East and Central Africa, low transmission efficacy compared with HIV-1 and reduced pathogenicity (either a long incubation period to AIDS, or a smaller fraction of infecteds developing serious disease, or both). The last factor

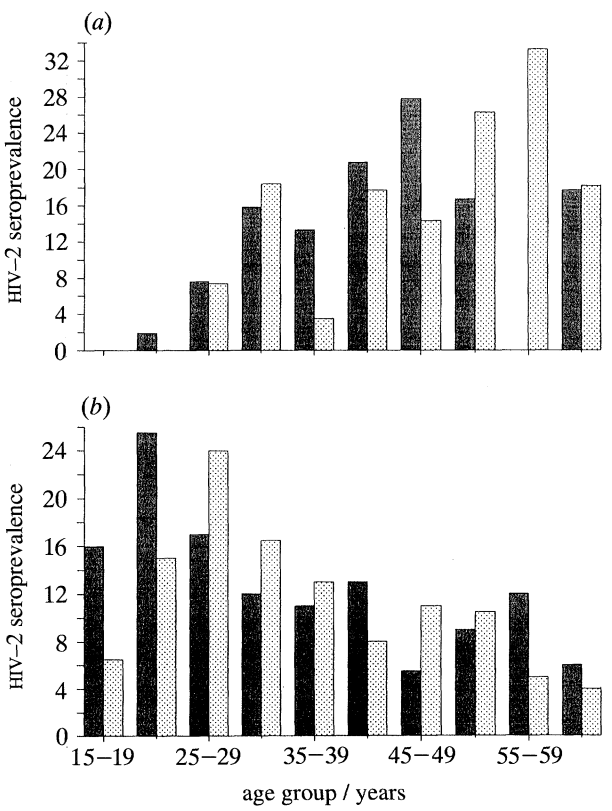

Figure 26. Age and sex stratified cross-sectional surveys of (*a*) HIV-2 seroprevalence in Guinea–Bissau (1987), and (*b*) HIV-1 seroprevalence in Uganda (shaded bars, females; stippled bars, males) (National Survey 1987–1988). (From U.S.A. Bureau of the Census HIV/AIDS database, 1993.)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

seems very likely since other evidence points to HIV-2-associated disease being slower to develop and less aggressive than that associated with HIV-1 (De Cock & Brun-Vezinet 1989). As yet no precise data is available on the distribution of the incubation periods of AIDS in HIV-2 infected persons but evidence is accumulating to suggest it is much longer in those who eventually develop disease (Ancelle *et al.* 1987; De Cock *et al.* 1994). Little is known about transmission efficacy, but virus abundance seems low in the HIV-2 infected persons by comparison with those infected with HIV-1. Such observations are tentative given the difficulty of assessing an 'average' variant abundance over a very long incubation period (Poulsen *et al.* 1989; Pepin *et al.* 1991). A very recent study of vertical transmission of HIV-1 and HIV-2 in Abidan, Cote d'Ivoire by De Cock *et al.* (1994) does point to a marked difference between the two viruses with the rate for HIV-2 being much less than that for HIV-1.

The pattern of coexistence of these two viral strains in West Africa is unfolding at present. If the theoretical prediction displayed in figures 24 and 25 bear some relation to reality, it may be a very long time before we fully understand the competitive interaction between the two viral strains. However, the early picture suggests that HIV-1 seems to be replacing HIV-2 in areas where the transmission of both is moderate to intense. This is illustrated in figure 27 from studies of the epidemiology of both viruses in high risk groups in Abidan, Cote d'Ivoire by De Cock and colleagues (US Bureau of the Census 1993). However, great care must be exercised in the interpretation of these patterns since in the early stages (i.e. many decades) of the epidemic the most virulent virus will spread most quickly in those with many sexual partners (because of a high transmission probability) even if longer term it is to be out competed by a less virulent virus with a greater net transmission potential (i.e. a bigger R_0) (see figure 24). Some idea of the eventual outcome may emerge once we know more about the transmissibility of both viruses in dual infections. In the case of vertical transmission the available data suggests that HIV-1 is much more efficiently transmitted, but there is no data as yet on horizontal transmission via sexual contact (De Cock *et al.* 1994).

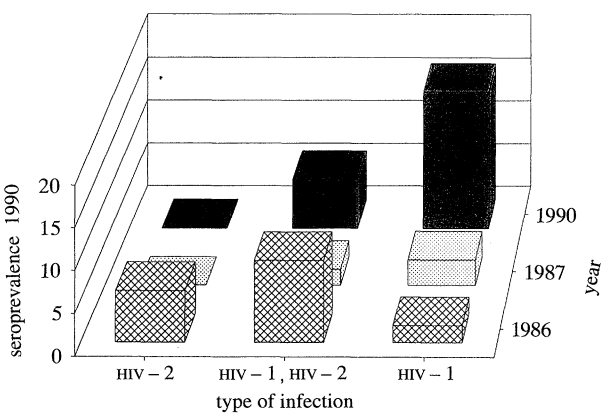

Figure 27. The prevalence of HIV-1 and HIV-2 (alone or in combination) in patients attending sexually transmitted disease (STD) clinics in Abidan, Ivory Coast in 1986, 1987 and 1990. (Data from U.S.A. Bureau of the Census, HIV/AIDS database, 1993.)

(*c*) *Antigenic variation and the transmission dynamics of malaria*

Multi-strain transmission models are also of great relevance in the interpretation of the epidemiology of the most pathogenic of the human malarial parasites, *Plasmodium falciparum*. The severe form of the disease induced by this infection is an important cause of child-mortality in many developing countries with estimates of childhood deaths between 0.5 to 2 million in Africa alone (Bruce-Chwatt 1985; Knell 1991; Greenwood *et al.* 1992). However, death due to malaria is a relatively rare event at the population level in relation to the frequency of infection in endemic areas and its diagnosis in the absence of clinical or parasitological investigations can be difficult (Snow *et al.* 1993). Recently, evidence has emerged of space-time clustering of severe childhood malaria in defined localities (Snow *et al.* 1993), and this may be due to new epidemics arising from the arrival or emergence of antigenically distinct parasite strains that have not been a source of infection before in a particular community. Much current research is focused on the study of parasite genetic variation and its association with severe morbidity and mortality. The variation in malaria parasites arises both within the host (mutation in asexual reproduction) and/or variant surface antigen expression in clones (Brown & Brown 1965) within and between hosts. At least three genetic mechanisms play a role in the generation of between host variation, including recombination events (such as unequal crossing-over in genes that encode polypeptides that contain tandem repeats), Mendelian segregation and recombination during the sexual phase and the generation of novel molecules by intragenic recombination during meiosis (Day *et al.* 1993; Roberts *et al.* 1993). Although a better understanding of the molecular basis of antigenic diversity has emerged over recent years, study of this diversity in human communities, and its implication for observed epidemiological pattern remains limited. Undoubtedly, antigenic variation within the host acts to prolong the average duration of infection in the face of the host's immune defences (this may be of particular importance in areas with highly seasonal transmission, where vector density is very low for certain periods of the year). Circulation of distant antigenic strains (clones?) of the parasite in the human population may help to overcome the influence of herd immunity by facilitating the repeated infection of individuals by different variants. One of the long standing puzzles in the epidemiological study of malaria in human communities is the long period of time required to build up a degree of resistance to infection despite repeated exposure in childhood (figure 28). A possible explanation of this pattern is the circulation of many

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

antigenically distant types or strains of the parasite where recovery from infection by one strain confers good protection against repeated infection (or at least the disease induced by infection) by that strain (specific immunological responses) but weak protection against infection or disease induced by antigenically different strains (weak cross-reactive responses). Immunity may only develop after exposure to many different antigenic types.

Molecular epidemiological studies have begun to address this problem using techniques that identify the cross-sectional and longitudinal trends of exposure to a variety of antigenically distinct strains (Forsyth *et al.* 1988; Day & Marsh 1991; Day *et al.* 1992; Gupta *et al.* 1994*a*,*b*). These studies are addressing a number of questions such as: is serious disease related to infection by particularly virulent strains of the parasite? How do strains with widely different virulences coexist in the same host population? Will the transmission of many distinct strains influence the way in which we estimate the basic reproductive rate or transmission potential of malaria in a defined community?

Two recent studies have addressed the last two of these questions (Gupta *et al.* 1994*a*,*c*). Despite the occurrence of genetic recombination events, evidence is emerging to suggest that a significant degree of clonality may prevail in areas of endemic infection (Day, personal communication; Gupta *et al.* 1994*b*) (recombination may be a rare event, particularly where the intensity of transmission is low or moderate). Working on the hypothesis that malaria can be viewed as the independent circulation of a series of antigenically distant strains (ignoring the occurrence of recombination events) it is possible to formulate a simple model of the transmission dynamics of many strains. Under the assumption that strain specific immunity is lifelong, the average age at first exposure to malaria, A_n, where n strains are circulating is given by (Gupta *et al.* 1994*c*):

$$A_n = L \Big/ \sum R_{0i}, \tag{51}$$

where R_{0i} is the basic reproductive rate of strain i and

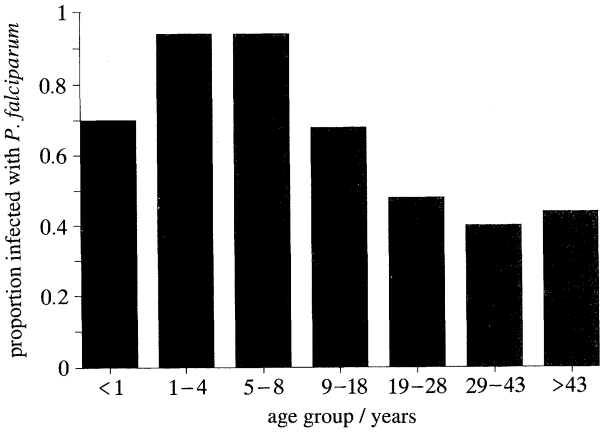

Figure 28. Age stratified survey for *Plasmodium falciparum* infection in northern Nigeria (Molineaux & Gramiccia 1980).

L is life expectancy. If we ignore strain variability and assume that immunity is not lifelong (due to the need to be exposed to many strains) then the equivalent expression is

$$A = (H + D\gamma)/R_0 - 1). \tag{52}$$

Here, H is the average duration of immunity, D is the average duration of infection and γ is the average number of blood meals a mosquito takes during its lifetime. Both (51 and 52) make clear that an early average age at infection may result even when R_0 is small, either due to a short duration of immunity (51) or due to the summation of a series of low R_{0i} values for a large number of strains (50) under conditions of lifelong strain specific immunity (Gupta *et al.* 1994*c*).

Past studies have always assumed that the transmission potential (R_0) of malaria was high due to the low average age of infection in endemic areas (Molineaux & Gramiccia 1980) (figure 28).

One of the most valuable features of the study by Gupta *et al.* (1994*c*) was the collection of field data from Madang, Papua New Guinea and its analysis based on the theoretical framework briefly mentioned above. The field study recorded the rise in exposure to five *P. falciparum* isolates that differ in parasite-induced erythrocyte surface antigens (PIESAs) (figure 29). The proportion exposed to any type rises moderately rapidly with age, but the rise in the proportion exposed to any one PIESA serotype rises only slowly with age. The observed patterns are consistent with the idea that serotype-specific immunity may endure for long periods, but they certainly do not exclude the alternative hypothesis that immunity builds up slowly after repeated exposure to one serotype. Most interestingly, an average R_0 value of between 6–7 provided an excellent fit to the observed proportion who have seen exactly n strains by age a. The method of estimation was based on a binomial probability model assuming complete independence of strain transmission. As recorded in figure 30, the positive binomial provides an excellent fit to the data, but it should be noted that the number of strains examined is small ($n = 5$). Also of interest is the observation that in analyses of the data sets for each isolate, the range of estimated R_{0i} values is small. They all cluster around the average of 6–7.

Before leaving the topic of estimating R_0 values, it is important to note that as recorded in section 2 for sexual transmitted infections, heterogeneity in exposure can complicate both estimation and interpretation. For example, if we take account of heterogeneity in exposure (irrespective of strain type) and assume that the number of exposures in the population at any age a, is negatively binomially distributed (with aggregation parameter k), then simple models of the change with age in the mean number of strains experienced can be used to assess the significance of heterogeneity. As recorded in figure 30, the greater the degree of heterogeneity, the lower the plateau in the mean number of strains experienced. In these circumstances, the actual value of R_0 is likely to be larger than that estimated on the basis

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

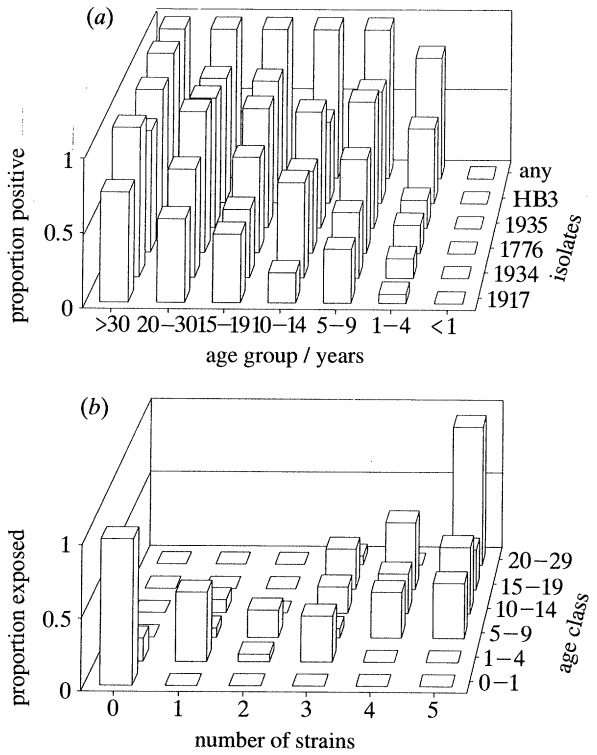

Figure 29. The rise by age in exposure to five *Plasmodium falciparum* isolates, shown as (*a*) the proportion in each age class with agglutinating antibodies to each strain, and alternatively (*b*) as the proportion in each age class exposed to exactly *n* strains (from Gupta *et al.* 1994*c*).

of a model that fails to take account of heterogeneity in exposure (Dietz 1980; Dye & Hasibeder 1986; Anderson & May 1991).

The analyses described above raise the question of why are the R_{0i} values recorded in the particular field

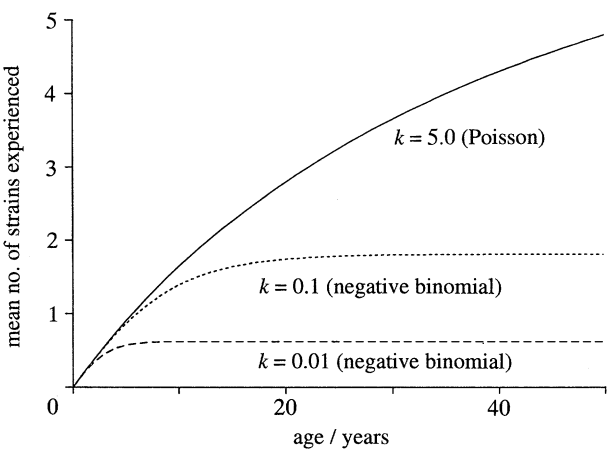

Figure 30. The influence of heterogeneity in exposure to infection within a human community with endemic *P. falciparum* infection. A model of exposure by age to many different strains of the parasite is adapted to mirror heterogeneity in exposure on the assumption that the distribution of people exposed to *n* strains at age *a*, is negative binomial in form with clamping parameter *k*. The graph records the mean number of strains experienced by age *a* under various assumptions concerning the degree of heterogeneity.

study in Papua New Guinea, so similar in magnitude. This may be a fault in the method of estimation where, for instance, independence was assumed in the occurrence of the different strains. This may or may not be true. It seems probable that the slow build-up of cross-reactive immunological responses after repeated exposure will influence the duration of infection, the infectiveness of a person and the occurrence of disease, but they are unlikely to influence exposure. It may be a fault associated with heterogeneity in exposure to infection or it may be due to the small number of serotypes examined.

Let us assume it is correct, and use a multi-strain transmission model to consider this more carefully. This approach has recently been followed by Gupta *et al.* (1994*b*), using the following set of equations to mirror the transmission of two strains within human host and vector populations. For two strains in a two-host life cycle (humans and mosquitos) seven variables may be defined using a novel way of compartmentalizing the various states of infection and immunity in the populations. The human population can be viewed as consisting only of immunes and susceptibles if we assume that the host is effectively immune to further invasion over the duration of the established infection. The proportion of immunes to strain i, x_i will include a smaller proportion of infections with strain i, x_{Ii}. It is assumed that the average duration of infectiousness is $1/s$ time units and that life expectancy is $1/\mu$ (very long by comparison with $1/s$). In the two-strain case, x_3 is defined as the proportion immune to both strains. The proportion of vectors infected by strain i is defined as y_i (no co-infection). The model has the following structure:

$$dx_1/dt = \lambda_1\left(1 - \sum x_i\right) - (c\lambda_2 + \mu)x_1, \quad (53)$$

$$dx_2/dt = \lambda_2\left(1 - \sum x_i\right) - (c\lambda, +\mu)x_2. \quad (54)$$

Here $1/\alpha$ defines vector life expectancy (the same for each strain), c records the degree

$$dx_3/dt = c(\lambda_2 x_2 + \lambda_2 x_1) - \mu x_3, \quad (55)$$

$$dx_{I1}/dt = \lambda_1\left(1 - \sum x_i + cx_2\right) - sx_{I1}, \quad (56)$$

$$dx_{I2}/dt = \lambda_2\left(1 - \sum x_i + cx_1\right) - sx_{I2}, \quad (57)$$

$$dy_1/dt = ax_{I1}\left(1 - \sum y_i\right) - \alpha y_1, \quad (58)$$

$$dy_2/dt = ax_{I2}\left(1 - \sum y_i\right) - \alpha y_2, \quad (59)$$

$$dy_2/dt = ax_{I2}\left(1 - \sum y_i\right) - \alpha y_2, \quad (60)$$

of cross-protection afforded by immunity to one of the strains (again assuming the same for each strain, $c = 1$ denotes no cross-immunity, $c = 0$ denotes total cross protection). The term λ_i denotes the per capita rate or force of infection where

$$\lambda_i = mab_i y_i. \quad (61)$$

Here a is the rate at which a mosquito bites the human host, b_i is the proportion of infected bites that results in infection with strain i and m is the number of female mosquitos per human host (i.e. the Ross-

Macdonald transmission term, Macdonald 1957). The model assumes that immunity to any particular strain is lifelong following recovery from 'infectiousness'. This assumption is in accord with the earlier discussions of the data from Madang, Papua New Guinea (Gupta *et al.* 1994*b*). The novelty in this framework lies in reducing the typical susceptible-infected-immune compartmentalization to one of immune non-infectious – immune and infectious.

The dynamical properties of this complex system of equations can be understood in part by analytic and numerical studies (Gupta *et al.* 1994*b*). In the extreme case of $c = 0$ (total cross immunity) competitive exclusion occurs and the strain with the largest R_0 wins. Similarly, when $c = 1$ (no cross immunity) both strains can always persist since they circulate totally independently. For intermediate values of c (a degree of cross immunity) equilibrium analysis suggests that coexistence is possible provided the following condition is satisfied:

$$c > (R_{01} - R_{02})/[R_{02}(1 - R_{01})]. \quad (62)$$

As the strength of cross immunity increases ($c \to 0$) coexistence is possible provided the R_{0i}s are not too dissimilar in value as illustrated in figure 31. The local stability of the possible equilibrium in the coexistence case has not been derived analytically but numerical studies suggest damped oscillation to a stable point. However, for small c values (high cross immunity) the oscillations have very wide amplitude if the R_{0i} values are not close to each other, and hence extinction may occur. These predictions may have some bearing on the observation recorded earlier in the study of Gupta *et al.* (1994*c*). The small range of the observed R_{0i} values for five strains coexisting together may reflect a moderate to strong degree of cross-reactive immunity. This observation is obviously preliminary since the model is very simple and the field study only examined a small number of strains. The topic, however, seems well worth further investigation.

One further problem of interest in the context of malaria concerns the evolution and persistence of very virulent strains of the parasite. Long term studies of temporal and spatial trends in 'malaria strain' prevalences are underway (Day *et al.* 1993; Snow *et al.* 1993) and it may transpire that strain composition and abundance fluctuate widely, both temporally and spatially, with strains of differing virulence arriving and disappearing on a frequent basis. The wide amplitude oscillations in the two-strain model above, and the very nonlinear dynamical patterns predicted by the Nowak & May (1993) superinfection model may underlie such patterns. Alternatively, the instabilities may arise from spatial structure, other heterogeneities or indeed, chance events in human populations of small size in a given spatial location. The relationship between virulence and transmissibility in falciparum malaria is an unresolved

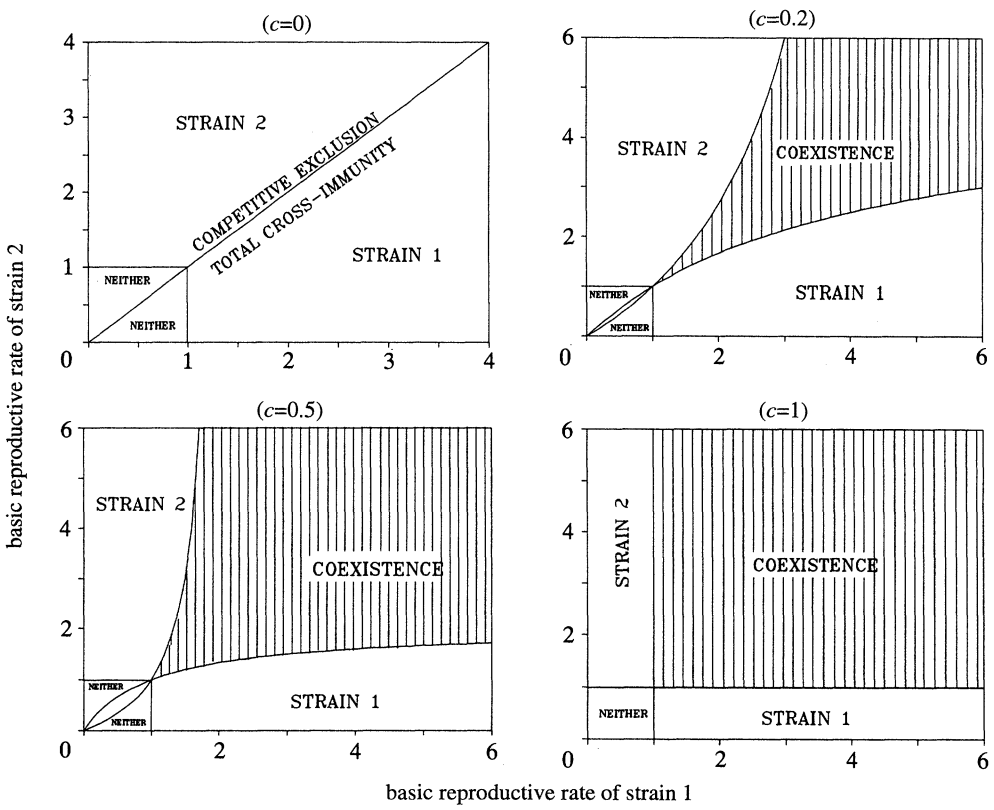

Figure 31. The influence of the degree of cross-immunity ($c = 0$, total cross-immunity; $c = 1$, no cross-immunity) on patterns of coexistence (at equilibrium) of two strains of *P. falciparum* (from Gupta, Swinton & Anderson 1994). The graph records the model predictions of the domains of coexistence and persistence of either strain in the two-dimensional space of values of the respective basic reproductive rates (R_{0i}) of the two strains (see text).

problem since arguments based on biological observation can be advanced to support either a positive or a negative correlation (Day *et al.* 1993; Gupta & Day 1994; Gupta *et al.* 1994*a*). Simple two-strain models without superinfection would suggest that for virulent strains to persist their R_0 values must be close to or greater in value than that of strains that induce mild disease. If serious disease dictates a shorter infectious period, this implies that infectiousness must be raised over that period to obtain equivalence in transmission potential with milder forms. Similarly, if superinfection results in the virulent strain outcompeting the avirulent forms within the human host, then theory suggests that virulence must have some linkage with transmissibility if the highly pathogenic strains are to persist (Anderson & May 1982*a*; Nowak & May 1993). Molecular epidemiological studies should be able to resolve this issue in the not too distant future.

6. INFECTIOUS DISEASE AND HUMAN POPULATION GROWTH

Throughout history, infectious diseases have had a major impact on the rise and fall of human civilizations (McNeill 1977). Many examples spring to mind including the effect of smallpox which, it is argued, was decisive in helping Cortez to conquer the Aztec Empire in the sixteenth century (whose subects numbered millions) with an army of only six hundred men. Many major plague (*Pasteurella pestis*) epidemics have been documented and the best known are the pandemics which, under the name of the Black Death, ravaged Europe in the fourteenth and seventeenth centuries (Henschen 1966). The detail contained in an account of the 1665 plague epidemic in London provides a macabre picture of the suffering and mortality caused by this disease (Defoe 1722).

Locally, some of the worst episodes induced dramatic changes in population size. The parish registration for the village of Eyam in North England, for example, records the loss of 267 people out of a population of 350 during the 1665 plague epidemic. On a larger scale the population of China was halved between 1200 and 1393, largely due the plague pandemic (although some historians argue that the Mongol invasions played a significant role). However, despite these checks on growth and perturbations in human mortality, the world's population continued to expand exponentially and today the current population size of 5.3 billion is projected (under conservative assumptions) to double by the year 2050. In the past four decades more people have been added to the globe than in all of history before the middle of this century. Virtually all of the new growth will take place in Africa, Asia and Latin America where infectious diseases remain the leading cause of human morbidity and mortality (Bongaarts 1994; World Bank 1993).

These observations raise the question of why infectious agents have not had a greater impact on human population growth, both in the past and in the present. Explanations can be sought at two levels. First, abundant evidence points to the existence of polymorphisms at the major histocompatibility and other genes that are important determinants of resistance/susceptibility to infectious agents. High diversity of MHC alleles is apparent in human populations and major classes of alleles that are separated from each other by as many as dozens of amino acid substitutions have persisted for very long periods of time (Gyllensten 1990; Klein *et al.* 1993; Wills & Green 1994). Most of the reported correlations between particular MHC alleles and diseases in humans suggest dominant rather than recessive susceptibility although few of these concern infectious agents. One of the most recent examples is a study by Hill *et al.* (1991) of children in The Gambia with severe falciparum malaria (*Plasmodium falciparum*). Children with severe disease showed an unusually low frequency of a Class I allele and a Class II haplotype that are common in the local population.

The maintenance of high diversity in genes that confer resistance is an issue that has attracted much attention. Various hypotheses have been proposed including overdominance, assortative mating, molecular mimicry, maternal-fetal incompatibility and adaptation to repeated pathogen exposure. The latter explanation is attractive, particularly so in light of much recent evidence which points to great genetic diversity within the pathogen populations themselves (e.g. *Plasmodium falciparum*). Intuition and theory suggest that rapid evolution by the pathogen (in response to the short term effects of the immunological defences of the host and the long term impact of selection for resistance in the host population) will act to maintain genetic diversity within the host population (Anderson 1981; May & Anderson 1983). One of the first to suggest this was J B S Haldane in 1949 (Haldane 1949). The spread and fixation of resistance alleles in the human population is prevented by the rapid evolution of the parasite where host herd 'genetic' resistance acts as a strong selective pressure for change in the pathogen population (Wills & Green 1994). The large difference between the generation time of the pathogen and that of its human host appears at first sight to be advantageous to infectious agents. This is not necessarily the case due to epidemiological and genetic factors. The intensity of transmission is invariably related to host density, hence high mortality during an epidemic will often reduce its likelihood until new births replenish the pool of susceptibles. In the genetic context, resistance to one or more infectious agents may confer susceptibility to others. Hence, with many different pathogens present, the selective pressures exerted by one may be in opposition to those imposed by others.

The notion that genetic diversity within a human population mitigates against high mortality during an epidemic outbreak of disease is supported both by laboratory studies of various infections in rodents (Dickerson 1990) and by studies in isolated and highly inbred human communities. For example, Chagnon records the impact of an influenza epidemic in 1973 in three Yanomamo villages in Venezuela which killed

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

27.4% of the inhabitants (Chagnon & Melancon 1981). He argued that the high level of mortality was in part due to the relative genetic homogeneity of the population. In the villages studied the average relatedness between all individuals was 0:0979 which is a level that falls in a range that covers first and second cousins (0.1250–0.0313).

A different approach to the question of why pathogens do not have a greater impact on human population growth is via an examination of the population biology of the interaction. So as to focus on the key principles involved, genetic diversity in host and pathogen populations is ignored. Ideally both genetic and population dynamic processes should be considered concomitantly (May & Anderson 1983; Beck 1984) but the resultant models tend to be complex in structure (Anderson *et al.* 1989).

(*a*) *Respiratory and gastro-intestinal infections that act to increase mortality*

Most of the infections that have attracted the greatest attention in the historical literature on human demography and disease are epidemic in character where the infection sweeps through a population in the space of a few months or a few years, inducing high mortality. Many were caused by directly transmitted infections such as influenza, measles, smallpox, cholera, and the plague. The commonest 'epidemic' infections tend to be respiratory or gastro-intestinal and in the case of viral infectious agents, the duration of infection is typically short (a few days to a few weeks) where the host either develops lasting immunity to reinfection by the same strain of the pathogen or dies as a consequence of infection. To examine their potential impact on human populations a model that combines both epidemiological and demographic processes is ideally required (Anderson *et al.* 1988; McLean & Anderson 1988; Anderson & May 1991).

A simple example is that of a directly transmitted virus which induces life long immunity to those who recover but adds an extra per capita mortality rate of α during the stage of infection. In a human population with a per capita growth rate r and average death rate μ (ignoring age-related trends in mortality and fertility) the infection will stably regulate human population growth provided

$$\alpha > r. \tag{63}$$

Conversely, if (63) is not satisfied, the system is predicted to settle to a state in which the human population continues to grow exponentially but at a reduced rate ρ (Anderson & May 1991) where

$$\rho = [B^2 - \mu(\alpha - r) + rv]^{1/2} - B. \tag{64}$$

Here $B = \frac{1}{2}(\alpha + \mu + v - r)$ and $1/v$ is the average duration of infection (= infectiousness in a simple model with no latent period). In either event the infection is unable to establish itself so long as population size is below a critical threshold value, $N_T = (\alpha + \mu + v)/\beta$, where β is the probability of transmission.

What (64) implies in terms of a typical directly transmitted viral infection is illustrated in figure 32 where the human population growth rate (% per annum) is plotted against the case mortality rate (%) induced by the infection (= $[\alpha/(\alpha + \mu + v)] \times 100$). In this particular example the average duration of infection was set at 15 days and the pristine population growth rate was fixed at 3% per annum (in line with many developing countries at present) in the absence of the infection. Note that to stably regulate population growth the case mortality rate would have to be in excess of 60%. Even case mortality rates of 20% would only reduce net growth to 2% per annum. In other words to have a major impact on net growth the infectious agent would have to be very pathogenic. This rarely occurs in practice, probably as a consequence of genetic diversity in resistance/susceptibility even within small relatively inbred populations. For example, in the epidemic of influenza in the Yanomamo Indian villages in northern Venezuela in 1973 described by Chagnor & Melancon (1981) where the population is highly inbred the mortality rate was approximately 27%. As recorded in table 1 even the most devastating of the epidemics recorded in the historical literature, rarely have mortality rates in larger communities that would check a human population growth of 3% per annum in the absence of the disease. Resistance and innate immunity in some fraction of the population as a consequence of genetic variability has undoubtably been a very important factor in the ability of human communities to survive repeated epidemics.

(*b*) *Sexually transmitted infections that influence fertility*

Population growth can also be influenced by infectious agents that affect human fertility. Many of these are sexually transmitted and two of the most important examples are gonorrhoea (*Neisseria*

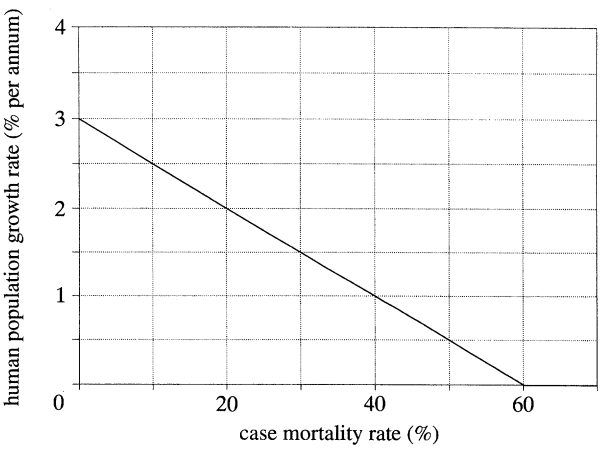

Figure 32. The impact of the case mortality rate (% of those infected who die from infection) on human population growth (rate (%) per annum) in a community in which the growth rate is 3% year^{-1} in the absence of infection. The predictions are generated by equation (64) in the main text with $1/\mu = 50$ years, $1/v = 14$ days, $r = 0.03$ year^{-1}.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table 1. *Case mortality rates in various epidemics*

infection/disease	case mortality rate	place	time	reference
influenza	27%	northern Venezuela	1973	Chagnor *et al.* (1981)
influenza	8%–30%	world pandemic	1918	Sinnecker (1976)
typhoid	10%	U.S.S.R.	1919	Sinnecker (1976)
dysentery (*Shigella dysenterae*)	12%–33%	Sweden	1803–1813	Henschen (1966)
dysentery	2%	U.S.A.	1861–1865	Henschen (1966)
cholera	34%	China	1932	Henschen (1966)
plague	33%–50%	world pandemic	14th century	Sinnecker (1976)

gonorrhoea) and *Chlamydia trachomatis* (Cates *et al.* 1990; Sherris & Fox 1983). Both infections cause salpingitis which is a major risk factor for infertility in women, particularly in Africa where both infections are prevalent (Cates *et al.* 1985). Seroepidemiological studies comparing the prevalence of both chlamydia and gonococcal antibodies among women with tubal infertility suggest that both agents are independent causes of infertility (Arya *et al.* 1973; Cates & Wasserheit 1991). In areas of the world where contraception is infrequently used, such as Sub-Saharan Africa, tubal infertility induced by sexually transmitted infections is thought to be an important determinant of the general fertility rate and, concomitantly, the net population growth rate (Brunham *et al.* 1993).

The impact of both infections on human population growth rates in Sub-Saharan Africa have recently been analysed using mathematical models that meld demographic and epidemiological processes (Brunham *et al.* 1992; Garnett *et al.* 1992; Swinton *et al.* 1992). The stimulus for these studies lay in two well designed field surveys conducted in 15 districts in Uganda by Arya *et al.* (1973) and Griffith (1963) which concluded that net birth rates in the districts were largely determined by the prevalence of gonorrhoea (figure 33). Simple models of the transmission dynamics of gonococcal and chlamydial infections were constructed to take account of various epidemiological complications such as heterogeneity in sexual activity, mixing patterns between different sexual activity classes, and latent plus infectious

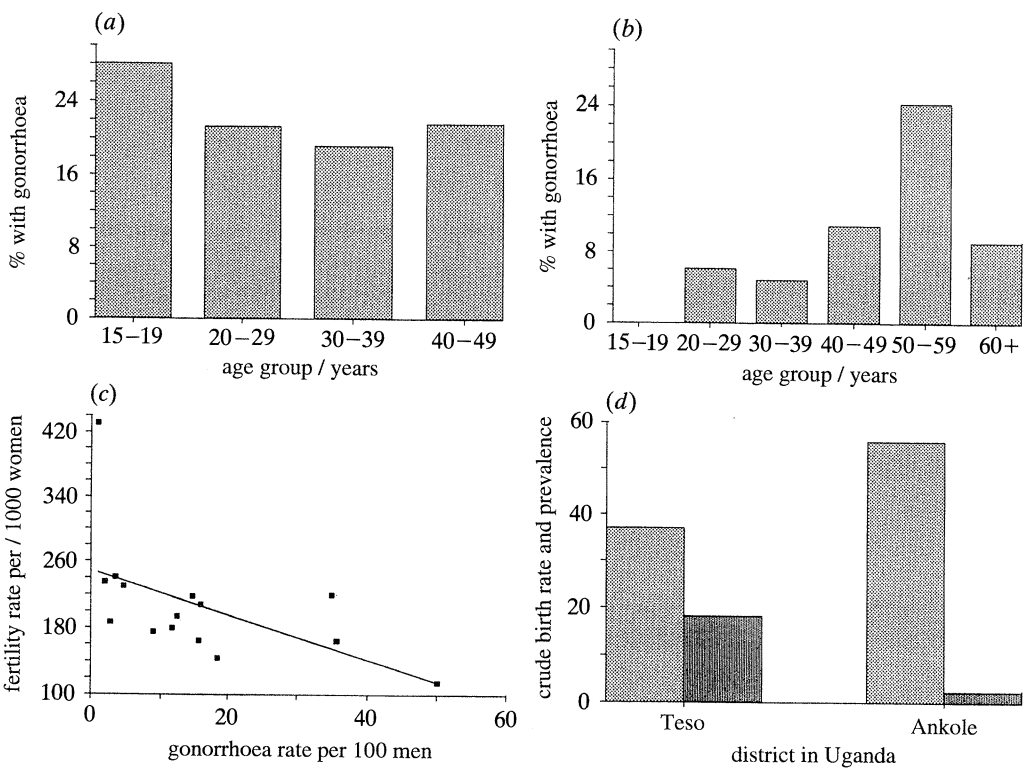

Figure 33. Epidemiology of gonorrhoea in Africa (*Neisseria gonorrhoea*). Age distribution of (*a*) women and (*b*) men infected with gonorrhoea in the Teso district of Uganda in 1971 (from Arya *et al.* 1973). (*c*) Observed relation between fertility rate of women (per 1000 head of women year^{-1}) in 15 districts in Uganda in 1959–1960 and the percentage of men having an annual attack of gonorrhoea (rate) as recorded by Griffith (1963). (*d*) The observed associations between the crude birth rate (per 1000 head year^{-1}) (light-shaded bars) and the prevalence of gonorrhoea in women in Teso and Ankole districts of Uganda in 1971 (dark-shaded bars) (from Ayra *et al.* 1973).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

periods during the course of infection. Demographic processes were included in a somewhat crude manner in which age structure was ignored but appropriate time delays were incorporated to mirror the delay between birth and reproductive maturity (Brunham *et al.* 1992). Parameter assignments were made on the basis of published data on the epidemiology of both infections and on demographic trends in Sub-Saharan Africa.

The results of the analysis were somewhat surprising. For example, in the case of gonorrhoea where untreated infection in women has a 12% likelihood of inducing infertility (Cates *et al.* 1990), in a population with a 4%/annum population growth rate in the absence of infection, an overall average prevalence of infection in women in the age classes 15 to 45 years of age of 10% was predicted to be able to reduce the net population growth by 30% (i.e. 4%/annum to 2.8%/annum) (figure 34). Chlamydia was predicted to have a less severe effect where a 10% prevalence was predicted to reduce the growth rate by approximately 10% (Brunham *et al.* 1993). In both sets of calculations, it was assumed that mixing between sexual activity classes was assortative (like with like) but with moderate degrees of contact between classes. The predictions are significantly changed if mixing is highly assortative. In these circumstances the impact of both infections at moderate to low prevalences will be limited. It should be stressed that these calculations are preliminary since more complex models with the details of age-specific fertility and mortality, plus those influencing mixing between different age classes of the two sexes are ideally required in future analyses.

Bearing in mind this caveat, the preliminary analyses suggest that sexually transmitted infections can have a significant impact on net population growth if the average prevalence of infection is high in the reproductive age classes of women. Because men in many Sub-Saharan African communities tend, on average to form sexual partnerships with women much younger than themselves (Hogsborg & Aaby 1992), observed epidemiological patterns reveal high prevalences in young women in many communities (see figure 33*a*,*b*). In poor regions with inadequate health case facilities (especially for women), the incidences of untreated sexually transmitted diseases (STDs) has been rising over the past few decades (Aral & Holmes 1990) in part due to the phenomena of migrant male labour where male spouses spend long periods in urban centres away from their families in the rural regions, for reasons associated with employment availability. One consequence of this is a greater degree of sexual contact with female prostitutes who act as an important reservoir (the so-called core group) (Hethcote & Yorke 1984) for STDs. Via such contact, infection is transmitted back from males to their regular partners in the rural areas.

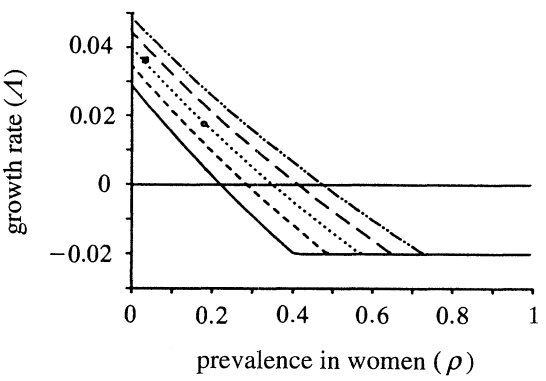

Figure 34. Predicted relation between the population growth rate (Λ) and the prevalence of gonorrhoea in women (p) based on a simple model of the transmission dynamics of the infection (from Brunham *et al.* 1992). The various lines record predictions for different assumptions concerning maximum potential fertility (top line 0.85, bottom line 0.45) for a basic reproductive rate value of $R_0 = 1.18$ (the circles denote observed values for Ankole and Teso in Uganda: see figure 33).

The currently observed population growth rates in Sub-Saharan Africa are probably influenced by STD-induced fertility but the effect is likely to be heterogeneous both within and between countries due to varying patterns of sexual behaviour and health care provision (Potts *et al.* 1991). Improved treatment of these infections is not only desirable in the context of improving women's health in poor regions but also in relation to their probable influence in enhancing (as cofactors) the likelihood of HIV transmission (Laga *et al.* 1991). Enhanced fertility as a result of STD treatment could act to offset (to some extent) the mortality induced by AIDS (Brunham *et al.* 1992).

(*c*) *The demographic impact of* AIDS

As described in the previous section, sexually transmitted diseases are typically a cause of infertility as opposed to mortality. In the past, the major exceptions were syphilis and Hepatitis B virus (a cause of carcinoma of the liver). The former is an old disease in human societies with evidence of its influence dating back at least to the 15th century (Henschen 1966). Some anthropologists place the origins of the disease to North and Central America (well before the arrival of Europeans) some tens of thousands of years ago but the topic is one surrounded by controversy and much speculation. Whatever the origin, syphilitic bone lesions have been identified dating back to 1493. Historical records speak of the explosive spread of the disease in 1493 in Europe, but its impact on human mortality and fertility were limited since it is typically a chronic infection where mortality, if it occurs, takes place well after the peak reproductive years. Vertical transmission to offspring of infected women occurs and can have an impact on the survival of the infant.

Today, however, the severity of syphilis as a cause of morbidity and mortality (in the long term) is dwarfed by the emergence of AIDS. The aetiological agent, HIV-1, has spread widely throughout the world and in the worst afflicted regions the prevalence of infection has reached epidemic proportions (Piot *et al.* 1991). Space does not permit a detailed account of the discovery and emergence of this infection and

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

attention is restricted to its potential demographic impact in the most badly affected areas of the world such as Sub-Saharan Africa and Thailand. In both regions the predominant route of transmission is via penal-vaginal intercourse in heterosexual relationships. In attempting to understand the likely demographic impact of AIDS in a given community a number of key epidemiological parameters play a central role but data on these is still limited at present. The key processes are the incubation and infectious periods and their distributions, the probability of transmission via heterosexual intercourse and how it is influenced by the time since infection and other factors such as the presence of other STDs, the fraction of those infected who will eventually develop AIDS and on what time scale, patterns of sexual behaviour and heterogeneity therein (both in rates of sexual partner change and sexual acts per unit of time), patterns of mixing between different strata of the population (e.g. sexual activity groups, age classes and spatial location), rates of vertical transmission and how they are influenced by factors such as breast feeding, birth practices and the duration of infection in the mother, and survival once AIDS is diagnosed both in adults and infants. Current information on these factors is reviewed in a series of recent papers (Anderson *et al.* 1991; Anderson *et al.* 1992; Anderson 1993; Garnett & Anderson 1993*a*).

In determining demographic impact, the most important factors are the prevalence of HIV-1 and its distribution by age and sex, the average incubation period of AIDS, the fraction who will develop the disease and the rate of vertical transmission. The prevalence of HIV-1 varies widely, within and between different countries, in the different at-risk groups. The extent of spread of the virus can be monitored via serological surveys of pregnant women attending antenatal clinics (representative of the general sexually active population) and of high risk groups such as female prostitutes (= commercial sex workers, CSWs) and attendees at STD clinics. Some examples are depicted in figures 35 (India), 36 (Thailand) and 37 (Sub-Saharan Africa). Three general patterns are apparent in these graphs. First, the prevalence of infection rises rapidly in high risk groups often attaining levels of 80% (e.g. figure 37*a*). These groups provide a source of infection for males who then transmit the virus to their regular partners; the rise in seroprevalence in the general population typically lags that in high risk groups by five years or more and the former tends to plateau at a much lower level than the latter. Second, there is evidence, particularly in Sub-Saharan Africa, of a plateau being reached where the epidemic is slowly moving to an endemic state (figure 37). Lastly, great heterogeneity is apparent in the level of this plateau in pregnant women. It ranges from around 6% in Kinshasa, Zaire to around 30% in Francis Town, Botswana. Before turning to the interpretation of these patterns and their likely significance to demographic trends it is important to note that certain epidemiological parameters appear to differ in different geographical locations. For example, recent evidence from Kenya (Plummer *et al.* 1993) suggests that the average incubation period of AIDS in adults is of the order of 4–5 years. This is to be compared with an average of 10 years in all adult risk groups in Europe and North America (Hendriks *et al.* 1993). The rate of vertical transmission is 12%–15% in developed countries (European Collaborative Study 1992) but 20%–40% in Sub-Saharan Africa (Hira *et al.* 1989; Lallemant *et al.* 1989; Ryder *et al.* 1989). The probability of transmission between heterosexual couples per act of penetrative penal-vaginal sex is of the order of 0.001–0.002 for female to male transfer in Europe and North America (European Study 1992; Saracco *et al.* 1993) but a recent study in Thailand suggests a figure of 0.025–0.04 which is an order of magnitude greater (Mastro *et al.* 1994). The reasons for these differences are unclear but may be related to the genetic diversity of the virus in different locations (relevant to all three processes) (Ou *et al.* 1993), nutritional status and rate of exposure to opportunistic infectious agents (the incubation period), frequency of breast feeding (the rate of vertical transmission: see Ziegler 1993) and host genetic background (relevant to all three).

The interpretation of observed pattern has been helped by theoretical studies of the transmission dynamics of the virus (Anderson & May 1988; Jacquez *et al.* 1988; Gupta *et al.* 1989; Anderson *et al.* 1991). Whether the prevalence of infection stabilizes at 5%, 10% or 40% in the general

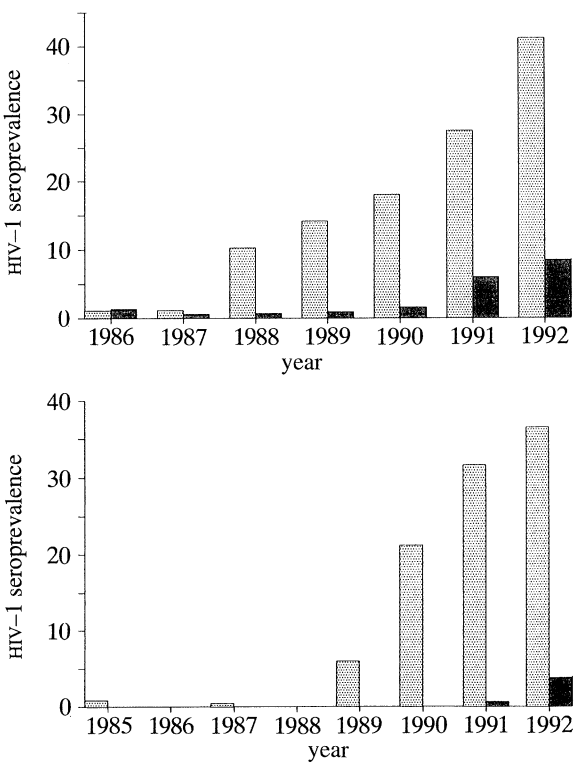

Figure 35. The spread of HIV-1 in India. Two illustrations of longitudinal trends in HIV-1 seroprevalence in (*a*) prostitutes in Bombay (stippled bars) and STD patients in Madras (shaded bars), and (*b*) prostitutes (stippled bars) and pregnant women (shaded bars) in Puna. (Data from US Bureau of the Census HIV/AIDS database, 1993.)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

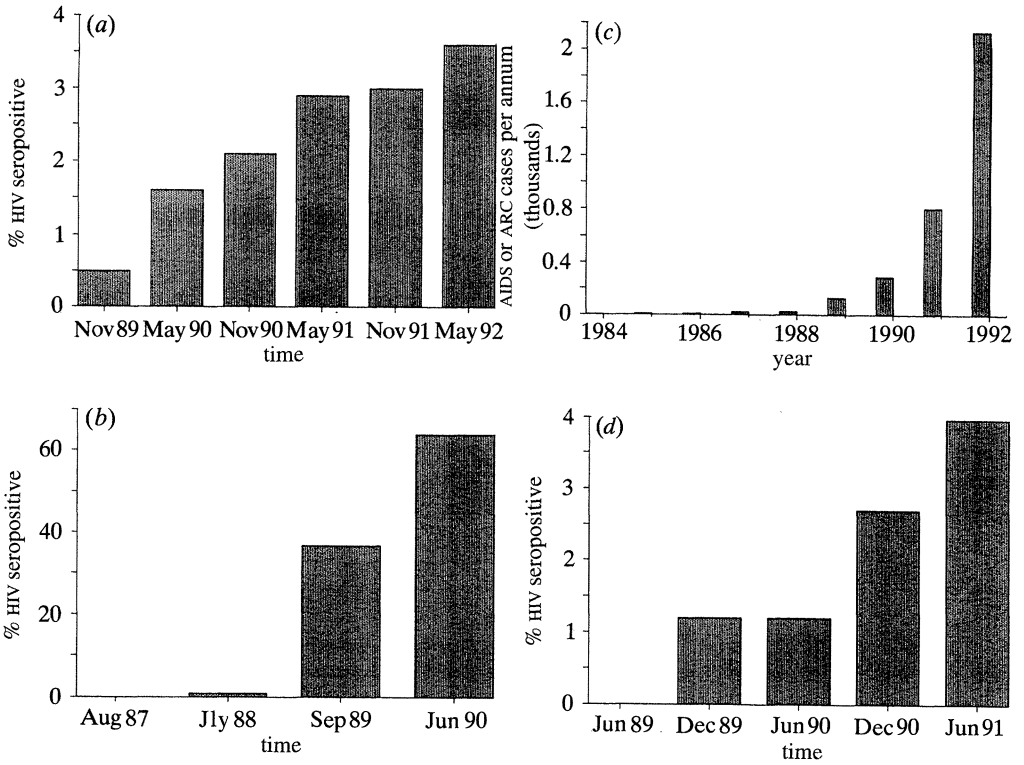

Figure 36. The spread of HIV-1 in Thailand. Examples of longitudinal trends in HIV-1 seroprevalence in (*a*) young men entering the Thai army, (*b*) female prostitutes in Chiang Rai, and (*d*) high class female sex workers in Bangkok. (*c*) Reported cases of AIDS and ARC (AIDS related complex) in Thailand. (Data from the US Bureau of the Census, HIV/AIDS database, 1993.)

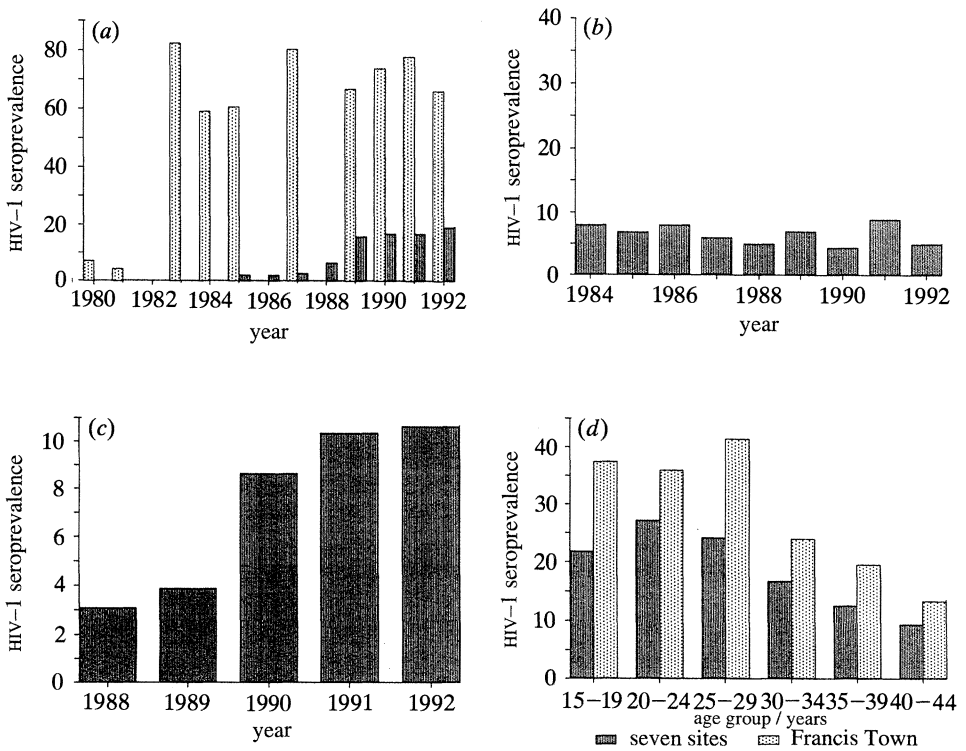

Figure 37. The spread of HIV-1 in sub-Saharan Africa. Examples of longitudinal trends in HIV-1 seroprevalence in (*a*) prostitutes (stippled bars) and pregnant women (shaded bars) in Nairobi, Kenya, (*b*) pregnant women in Kinshasa, Zaire, (*c*) pregnant women in Mbeya, Tanzania. (*d*) A cross-sectional survey in 1993 in Botswana (shaded bars, seven sites; stippled bars, Francis Town). (Data from the US Bureau of the Census, HIV/AIDS database, 1993.)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

populations is largely determined by rates of sexual partner acquisition and heterogeneity therein, and the prevailing pattern of mixing between groups with low, medium and high rates of partner acquisition (Anderson *et al.* 1991; Garnett & Anderson 1993*a*). Highly assortative mixing (like with like) implies an epidemic of limited magnitude as experienced in Kinshasa, Zaire (figure 37*b*) while moderate to high mixing implies a major epidemic as seen in Francis Town, Botswana (figure 37*d*). The current generation of mathematical models incorporate a fair degree of biological complexity including distributed incubation and infectious periods, heterogeneity in sexual activity and variable mixing beween age and sexual activity classes of the two sexes (Garnett & Anderson 1993*a*). They are certainly able to mimic the longitudinal trends recorded in figure 37, both for pregnant women and the high risk group. The time to the endemic state is much influenced by the average incubation period of AIDS (or more precisely the average time from infection to death). The recent information that hints at a 4–5 year average in Sub-Saharan Africa does suggest that, with the establishment of the virus in urban centres around the late 1970s to early 1980s, the HIV epidemic in the worst affected areas may be nearing its peak (the peak in the incidence of AIDS and disease related mortality will be later due to the long incubation period of the diseases).

Early attempts to meld understanding of the epidemiology of the virus and the demographic of human communities in the developing world suggested that AIDS could negate a 3% per annum population growth rate over timescales of a few decades (Anderson *et al.* 1988). Such predictions caused much controversy but recent work based on much more sophisticated models still supports this view for communities with high HIV prevalence across the sexually active age classes (Anderson *et al.* 1991; Garnett & Anderson 1993*a*). Additional support is provided by simple demographic calculations (requiring no knowledge of the prevailing patterns of sexual behaviour) which provide estimates of what prevalence is required to negate a defined population growth rate (Garnett & Anderson 1993*c*). An example of such a calculation is presented in figure 38*b*,*c*, for an average incubation period of 8 years, which suggests that an average prevalence in women of around 30%–35% will negate a 3% growth rate. In the worst affected urban centres in Africa (see figure 37*d*) seroprevalence is approaching this level.

More generally, however, it is to be hoped that HIV prevalence in the majority of African communities will saturate at much lower levels. Whether it does or does not will depend on patterns of mixing, rates of sexual partner change and the prevalence of condom use. In illustration of one of these points, figure 39 records theoretical predictions of the impact of different patterns of mixing on the course of the epidemic (all other parameters held constant in the three simulations plotted). When mixing is highly assortative a moderate epidemic (with HIV-1 prevalence attaining 20% in the 15–49 age group of women) halves the pristine population growth rate of

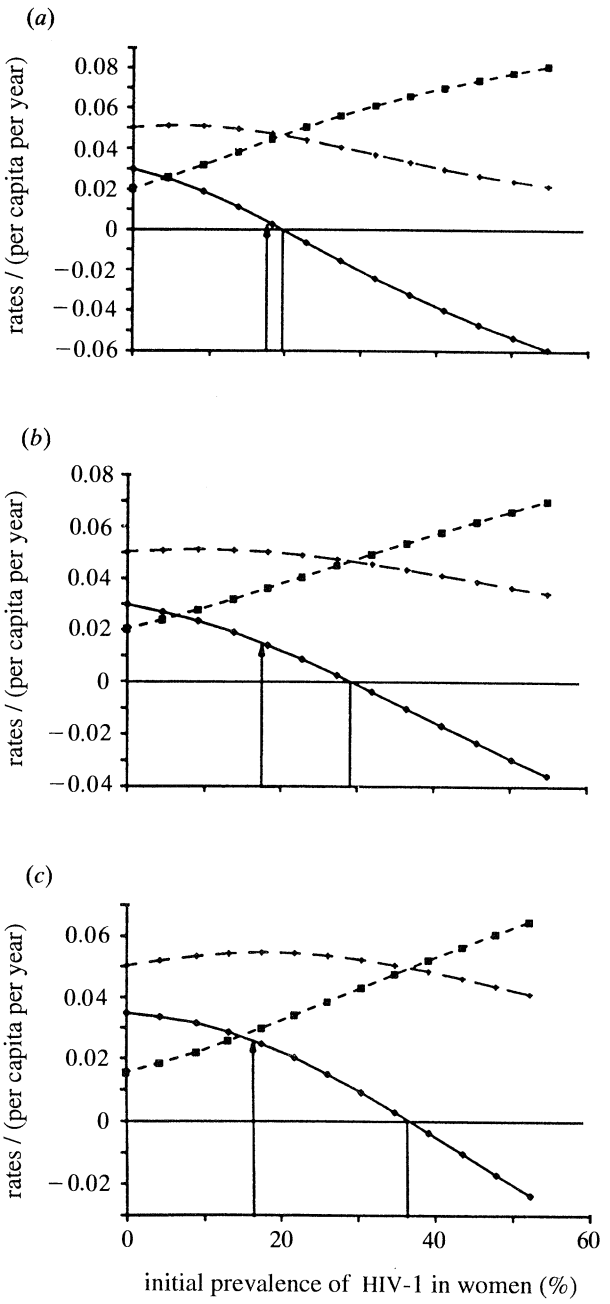

Figure 38. The predictions of a simple 'equilibrium' model of the potential demographic impact of AIDS in a community with a 3% per annum growth rate prior to the introduction of HIV-1 (from Garnett & Anderson 1993*c*). The three graphs record the predicted relationship between the per capita growth rate (solid line), death rate (lightly dashed line) and birth rate (heavy dashed line) and the prevalence of HIV-1 in women (%) for different assumptions concerning life expectancy prior to the AIDS epidemic ((*a*,*b*) 49 years, (*c*) 65 years) and the average incubation period of AIDS ((*a*) 5 years, (*b*,*c*) 8 years).

2.6%, moderately assortative mixing reduces the growth rate to around 0.5% whilst weak assortative mixing takes it below replacement level (the endemic prevalence reaches 45% in sexually active women). At present very little information is available on mixing patterns but the trends revealed in the longitudinal data plotted in figure 37 suggest that it

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

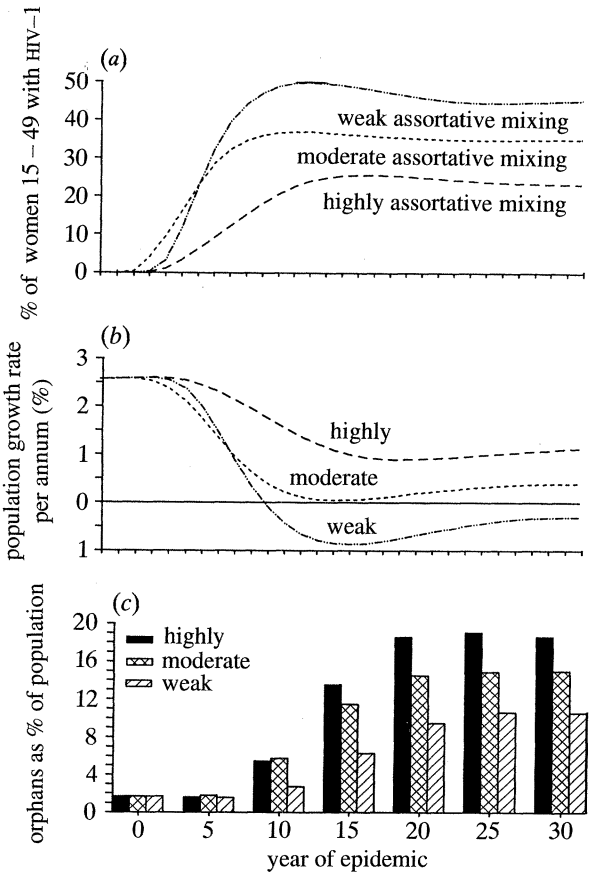

Figure 39. The predicted effects of (*a*) the pattern of mixing between different sexual activity classes (defined on the basis of rates of sexual partner change) – ranging from weak to highly assortative (like-with-like) – on temporal trends in the percentage of women 15–49 years of age with HIV-1, (*b*) the population growth rate per annum and (*c*) the number of orphans as a percentage of the total population (from Gregson *et al.* 1994*b*).

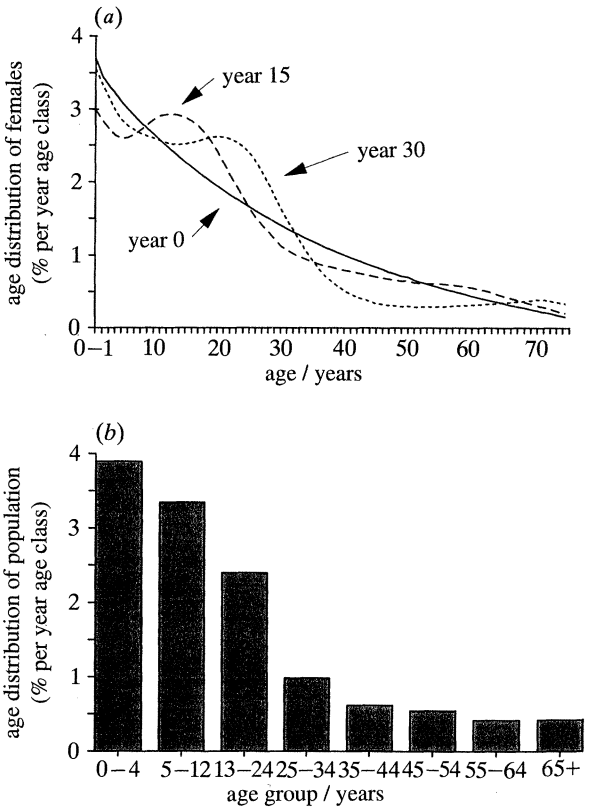

Figure 40. The impact of the AIDS epidemic on the age distribution of the population. (*a*) Predicted age distribution of females (%/yearly age class) at year 0, year 15 and year 30 of the epidemic (from Garnett & Anderson 1994; Gregson *et al.* 1994*b*). (*b*) Observed age distribution of the population of Masaka, Uganda in 1992 in which the AIDS epidemic is still in its early stages (from Mulder *et al.* 1993). Note that the age distribution is reported per age group, where the various groups record different intervals of age.

is moderate to highly assortative but variable across different regions.

Irrespective of the impact on the net population growth rate; the epidemic – whether moderate or severe – is predicted to have a major effect on population structure (Gregson *et al.* 1994*a*,*b*,*c*). An illustration is provided in figure 39*c* where the percentage of orphans predicted to arise as a consequence of HIV-related mortality is plotted over the course of simulated epidemics of minor, moderate and major size. Even in the best case, orphans are predicted to form 12% of the population by year 30 of the epidemic. Under the worst scenario they are predicted to form nearly 20% of the population. This observation has important implications for the provision of aid to help societies cope with the consequences of the AIDS epidemic.

A comparison between prediction and observation is provided in figure 40, where the age distribution of a population studied by Mulder *et al.* (1994) in Masaka, Uganda is plotted, in which an HIV-1 epidemic is in progress (8.2% average prevalence in adults in 1990, 4.8% over all age classes). Predicted changes in the age distribution are also plotted for various time points in simulated epidemics (year 0, 15 and 30) (Garnett & Anderson 1994). Note that the observed data shows evidence of a steep decline in life expectancy as the adult years are entered, as predicted in the simulations.

A more detailed comparison is possible for this particular study since Mulder *et al.* (1994) recorded the mortality rate by sex and age in uninfected and HIV-1 infected individuals (figure 41). Despite the moderate to low level of HIV-1 prevalence at present (i.e. 8.2% in 15+ individuals), over 80% of total mortality in the 25–34 year olds is HIV-1 related. This depressing figure is closely matched by theoretical predictions of the excess mortality induced by the epidemic as recorded for various stages after its induction in figure 42.

The main conclusion of recent field (De Cock *et al.* 1994; Mulder *et al.* 1994) and theoretical (Anderson *et al.* 1991; Garnett & Anderson 1993*a*,*c*; Gregson *et al.* 1994*a*,*b*,*c*) studies may be summarized as follows. Depending on the precise pattern of sexual behaviour, particularly the pattern of mixing between high and low risk groups, the spread of HIV-1 may result in major or minor epidemics with, concomitantly, major or less serious implications for net population growth. Irrespective of the magnitude of the epidemic,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

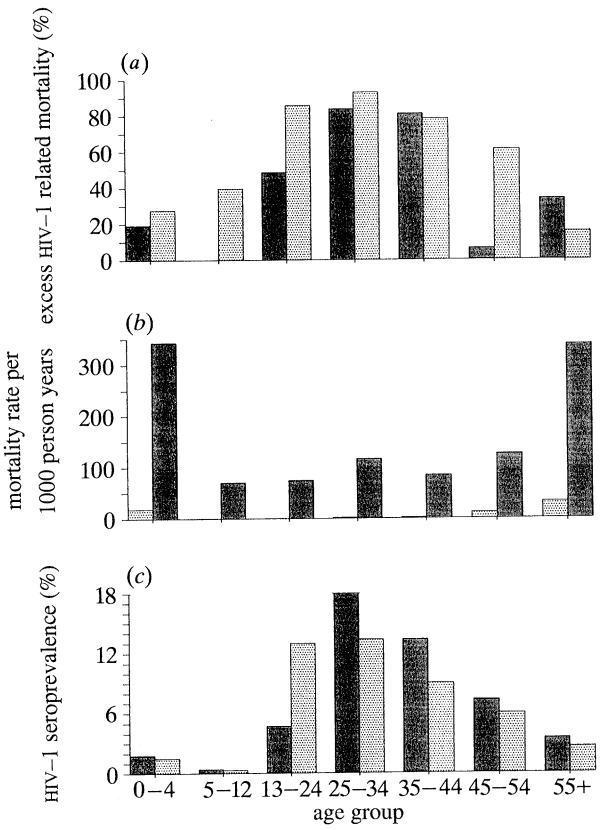

Figure 41. The epidemiology and demographic impact of AIDS in Masaka, Uganda in 1992 (from Mulder *et al.* 1993). (*a*) Excess mortality in HIV-1 infected males (shaded bars) and females (stippled bars) (i.e. a value of 40% denotes a 40% greater mortality rate in HIV-1 infected persons of that age, by comparison with HIV-1 uninfected persons), (*b*) mortality rate (/1000 person years) in infected (shaded bars) and uninfected (stippled bars) persons, and (*c*) HIV-1 seroprevalence by age and sex (males, shaded bars; females, stippled bars).

even in the case of low rates of spread (for Africa defined as 6–10% seroprevalence in pregnant women) the demographic implications are very serious. Both observation and theory predict that it will significantly change the age and sex composition of populations and that it will lead to a rapid rise in the number of orphans in most communities. These effects are just beginning to be documented in Africa (Mulder *et al.* 1994), but in the coming decade they will be observed also in India (Jain *et al.* 1994) and parts of South East Asia (e.g. Thailand) (Viravaidya *et al.* 1992; Editorial 1994).

A question that surfaced early on in the current pandemic was whether this new disease would ultimately dwarf all past major epidemics in its scope and impact? It is difficult to judge at present but a number of features of the aetiological agent and the disease it induces are a cause for concern. In particular, the lethality of the infection by comparison with others (see table 1) is striking. Current evidence from studies in Europe and North America suggests that 50% of those infected will have died 12–14 years post acquisition of infection. Mortality rates in Africa appear to be much higher and the average period to death is probably much shorter (Mulder *et al.* 1994; Anzala *et al.* 1993). Throughout this long period of chronic infection, people are infectious to others (with peaks early and late on in the incubation period). As such the transmission potential of the virus is high, despite the fact that the probability of transmission per sexual act is low by comparison with other STDs. The sexual mode of transmission is also a cause for concern since it implies that the disease will be able to persist endemically in low density communities (unlike many directly transmitted respiratory or gastro-intestinal pathogens (see Anderson & May 1979; May & Anderson 1979)). Lastly, the rapid evolution of the virus is of major significance both for the development of effective drugs and vaccines and for its ability to persist (for many years in the human host) in the face of strong immunological attack. As discussed earlier, whether evolutionary pressures will drive the virus to reduced or increased virulence is very uncertain at present. Perhaps the one ray of hope in this still unfolding tragedy is the preliminary evidence from studies of female prostitutes in Nairobi, Kenya, that some individuals appear not to acquire infection despite repeated exposure to infected persons (Anzala *et al.* 1993). Perhaps, yet again, the inherent genetic diversity of the human species in its ability to detect and eliminate disease-inducing agents will defeat the virus. However, it is to be hoped that education will limit its spread in the majority of regions of the world and that in the longer term effective drugs and immunotherapeutic agents (to alleviate the suffering of those infected) plus vaccines will be developed. Even greater efforts will be needed to promote educational messages (Carael *et al.* 1994) since in some areas or risk groups where public health programmes have focused on disseminating knowledge about AIDS, the incidence of new HIV infections often remains depressingly high (Melbye & Smith 1993; Wawer *et al.* 1994).

7. DISCUSSION

The focus of attention in the various topics discussed in the preceding sections, has ranged from the details of the immune system, via antigenic variation in pathogens, to human sexual behaviour and demography. In all cases the problems have been viewed from a population perspective, whether dealing with antibodies, lymphocytes, parasites or human hosts. A common theme has been the formulation of the interrelationships between variables in mathematical terms and the study of the dynamical properties of the resultant nonlinear models. In the majority of cases, attempts were made to estimate parameter values from experiments or observed epidemiological pattern, and in all cases predictions were compared with observation, either in a qualitative (i.e. similarity in pattern) or a quantitative manner. A recurring observation in the various problems was the dynamical complexity of even the simplest models of the interaction between infection and immunity, whether within the host or within the population of hosts.

The major conclusions arising from this collection of

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

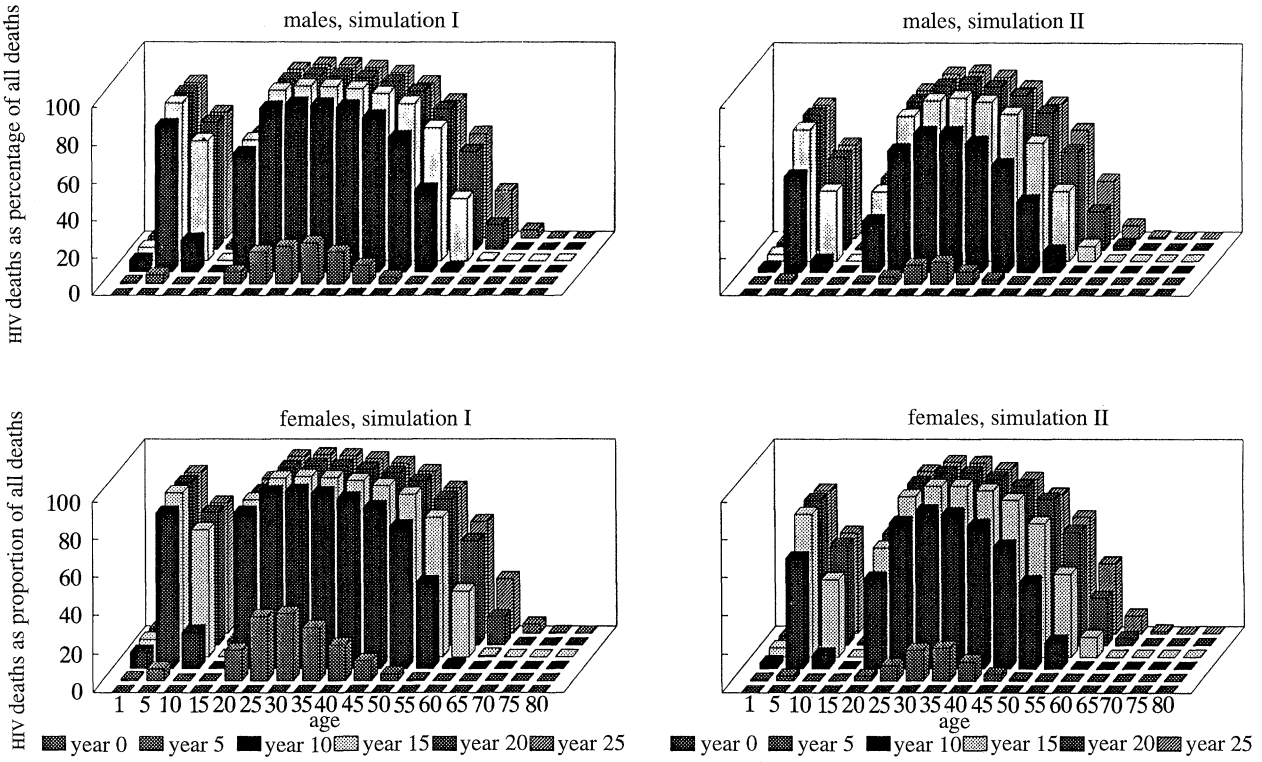

Figure 42. Predictions of a mathematical model of the demographic impact of AIDS. The graphs record predicted trends in HIV-1 related deaths (as a percentage of total deaths) by age, sex (top graphs, males; bottom graphs, females) and year of the epidemic. The graphs on the left hand side record simulations of a severe epidemic (mixing weakly assortative) while those on the right record a moderate epidemic (highly assortative mixing) (from Gregson *et al.* 1994*b*).

studies are as follows. For directly transmitted antigenically homogeneous viral infections of children, the insights generated by mathematical models can help both in the interpretation of observed epidemiological pattern and in the design of control policies. This is demonstrably an area of success in mathematical epidemiology, since model predictions can mirror in detail a variety of cross-sectional and longitudinal epidemiological patterns prior to, and post, the introduction of mass vaccination programmes. It is to be hoped that public health authorities, and the medical profession in general, make use of these methods to add rigour and precision to the decision making process. In the broader context of health in tropical regions, these models (McLean *et al.* 1991) and those of the transmission dynamics of the very prevalent helminth infections of children (Bundy & Guyatt 1992; Guyatt *et al.* 1994; Medley *et al.* 1994) can be employed to examine cost-effectiveness of different community-based policies, targeted at the control of either infection, disease or both.

As many childhood infections become rare in western societies due to vaccine-induced herd immunity, it is important to maintain vaccine coverage at high levels (despite the virtual absence of the infection) to prevent reintroduction from other regions of the world in which these infectious agents persist due to low levels of vaccine-induced herd immunity. Similarly, it will be important to keep track of vaccine failure rates over time in order to monitor any antigenic changes (perhaps driven by the selective pressure exerted by high levels of vaccine-induced immunity). In addition, as herd immunity due to vaccination rises, a high fraction of mothers will pass on transient immunity to their newborn offspring (maternally derived antibodies), derived not from recovery from infection but from vaccination. The latter may not be as protective as the former, hence surveillance should focus on longitudinal trends in the incidence of infection in infants during the first year of life. Lastly, in the context of mass vaccination, as coverage rises spatial heterogeneity in uptake will become of increasing importance and care must be taken to prevent reservoirs of infection persisting in poor communities in urban centres. Theoretical studies of the spatial dynamics of such infections, with heterogeneity in vaccine uptake and age of delivery, could help in assessing how best to tackle this problem.

Heterogeneities of all sorts influence transmission and these are most apparent in those infections transmitted via sexual contact. In most cases the heterogeneities act to enhance the likelihood of the persistence of infection in a defined community, but they tend to suppress the endemic prevalence of infection. Theoretical studies have delineated these influences and they have helped to focus attention on the quantitative study of patterns of human sexual behaviour, despite much opposition from certain quarters. Understanding average behaviour is insufficient; the variability around the average is often of much greater significance.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

One of the most exciting 'nonlinear worlds' in the field of infection and immunity, is the dynamical behaviour of the immune system in response to invasion by replicating infectious agents. Theoretical immunology has been very distant from the problems and hypotheses that excite experimentalists. However, yet again there are encouraging signs of the two fields moving closer together. It is the mathematical studies that have to change most, but there is a concomitant need for experimentalists to appreciate the precision provided by mathematics in testing hypothesis in complex interactive systems. Particular progress has been made recently in the study of antigenic variation in parasite populations, both within and between hosts. The large systems of nonlinear equations needed to characterize multi-parasite strain interactions either with the immune system or within the host population are amenable to study by both analytic and numerical methods. Such studies reveal very complex patterns of dynamical behaviour that would be difficult to understand in the absence of an appropriate mathematical template (Nowak *et al.* 1990). Similarly, new molecular methods facilitate the study of the temporal dynamics and evolution, both within and between hosts, of many antigenically distinct parasite strains. The mathematical framework of these systems is beginning to provide insights into evolutionary questions such as that of the persistence of virulence (Anderson & May 1982; Nowak & May 1993) and applied problems such as vaccine design and use in the community (Gupta *et al.* 1994*c*). There is much scope for further research that combines theoretical and experimental approaches.

The AIDS pandemic has raised awareness in the developed world that infectious agents are not simply a cause of morbidity. They can also be a major cause of mortality. This has always been the case in tropical regions of the developing world, but even the severest of infections such as falciparum malaria and tuberculosis have not checked the exponential rise in human population size in these regions. Theory that melds epidemiological and demographic processes helps us examine whether or not the demographic impact of AIDS is likely to be greater or less than the other major infections that afflict human communities. The answer seems to be that it has the potential to have a much greater impact but its precise magnitude will vary widely from location to location within and between countries, depending on the prevailing patterns of sexual behaviour. In some of the worst afflicted regions in Sub-Saharan Africa, data is emerging to test various predictions of trends in infection and mortality, both cross-sectionally and longitudinally. There is reasonable qualitative agreement between prediction and observation in such areas. Some believe that pessimistic statements about the potential demographic impact of AIDS in parts of the developing world are harmful, since they encourage inaction both by governments in the regions and by international bodies. Whether examined from a humanitarian viewpoint or from the international public health stance of the control of infection in a world increasingly influenced by mixing and travel between countries, or in the context of preserving political, social and economic stability, the need to allocate very significant resources to combat the spread of HIV is clear to the vast majority of informed observers of the pandemic. A scientific understanding of future trends must be used to inform policy makers and to guide efforts to alleviate the impact of the disease.

Finally, the last decade in particular has seen the biological sciences enter a phase of very rapid growth in information about structure and function, largely due to advances in molecular biology. The details of new genes, and of the new nucleotide sequences associated with them, appear with extraordinary frequency. Emphasis is on description and on whether or not the techniques that allow description are correct. It is rarely on quantification or on the dynamics of systems with many interacting variables. The pace of descriptive studies is fast since many are able to make the same observations (Maddox 1993). Hopefully some will find time to reflect on the accumulating data to seek similarities and differences and to construct theories and hypotheses to explain pattern. This process is urgently needed in the areas of infection and immunity of greatest practical relevance, such as the development of vaccines and immunotherapies for antigenically variable pathogens. It is hoped that the problems discussed in this paper convince some experimental biologists that mathematics, if sensibly used in combination with observation, can both enhance precision in interpretation and facilitate the formulation plus testing of hypothesis to explain observed pattern.

R. M. A. thanks the MRC, Overseas Development Agency, the Department of Health and the Wellcome Trust for research grant support. Very warm thanks are due to a host of scientific colleagues who have helped in formulating and testing ideas, especially Robert May. A particular thank you goes to past (at Imperial College, London) and present (at Oxford) members of the Parasite Epidemiology Research Group.

REFERENCES

Aaby, P., Andersen, M. & Knudsen, K. 1993 Excess mortality after early exposure to measles. *Int. J. Epidemiol.* **22**, 156–162.

Aaby, P., Pedersen, I.R., Knudsen, K. *et al.* 1989 Child mortality related to seroconversion or lack of seroconversion after measles vaccination. *Ped. Infect. Dis.* **8**, 197–200.

ACSF 1992 AIDS and sexual behaviour in France. *Nature, Lond.* **360**, 407–409.

Agur, Z., Abiri, D. & Vanderploeg, L.H.T. 1989 Ordered appearance of antigenic variants of African trypanosomes explained in a mathematical-model based on a stochastic switch process and immune-selection against putative switch intermediates. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9626–9630.

Agur, Z., Cojocaru, L., Mazor, G., Anderson, R.M. & Danon, Y.L. 1993*a* Pulse mass measles vaccination across age cohorts. *Proc. natn. Acad. Sci. U.S.A.* **90**, 11698–11702.

Agur, Z., Danon Y.L., Anderson, R.M. & Coocaru, L.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

1993*b* Measles immunization strategies for an epidemiologically heterogeneous population: the Israeli case study. *Proc. R. Soc. Lond. B* **252**, 81–84.

Ancelle, R., Bietry, O., Baglin, A.C., Brun-Vezinet, F., Rey, M.A. & Godeau, P. 1987 Long incubation period for HIV-2 infection. *Lancet* i, 234–238.

Anderson, R.M. 1981 Vertebrate populations, pathogens and the immune system. In *Population and biology* (ed. N. Keyfitz), pp. 249–268. Liege: Ordina Editions.

Anderson, R.M. 1988 Epidemiology of HIV infection: variable incubation plus infectious periods and heterogeneity in sexual activity. *Jl R. statist. Soc. A* **151**, 66–93.

Anderson, R.M. 1991*a* Populations and infectious diseases: ecology or epidemiology? *J. Anim. Ecol.* **60**, 1–50.

Anderson, R.M. 1991*b* The transmission dynamics of sexually transmitted diseases: the behavioural component. In *Research issues in human behaviour and sexually transmitted diseases in the AIDS era* (ed. N. Wasserheit, S. O. Aral & K. K. Holmes), pp. 38–60). Washington DC: American Society for Microbiology.

Anderson, R.M. 1993 AIDS in developing countries: heterogeneity in potential demographic impact. In *AIDS, health and human rights* (ed. J. Mann & C. Dupy), pp. 21–46. Veyrier du Lac, France: Foundation Marciel Merieux.

Anderson, R.M. 1994 Mathematical studies of parasitic infection and immunity. *Science, Wash.* **264**, 1884–1886.

Anderson, R. M., Crombie, J.A. & Grenfell, B.T. 1987 The epidemiology of mumps in the U.K.: a preliminary study of virus transmission, herd immunity and the potential impact of immunization. *Epidemiol. & Infect.* **99**, 65–84.

Anderson, R.M., Gupta, S. & Ng, W. 1990 The significance of sexual partner contact networks for the transmission dynamics of HIV-1. *J. Aquired Immune Defic. Synd.* **3**, 417–429.

Anderson, R.M. & May, R.M. 1979 Population biology of infectious diseases. Part I. *Nature, Lond.* **280**, 361–367.

Anderson, R.M. & May, R.M. 1981 The population dynamics of microparasites and their invertebrate hosts. *Phil. Trans. R. Soc. Lond. B* **291**, 451–524.

Anderson, R.M. & May, R.M. 1982*a* Coevolution of hosts and parasites. *Parasitology* **85**, 411–426.

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

Anderson, R.M. & May, R.M. 1983 Vaccination against rubella and measles: quantitative investigations of different policies. *J. Hyg., Camb* **90**, 259–325.

Anderson, R.M. & May, R.M. 1986 The invasion, persistence and spread of infectious diseases within animal and plant communities. *Proc. R. Soc. Lond. B* **314**, 31–66.

Anderson, R.M. & May, R.M. 1988 Epidemiological parameters of HIV transmission. *Nature, Lond.* **333**, 414–519.

Anderson, R.M. & May, R.M. 1990 Immunisation and herd immunity. *Lancet* **335**, 641–645.

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

Anderson, R.M., May, R.M., Boily, M.C., Garnett, G.P. & Rowley J.T. 1991 The spread of HIV-1 in Africa: sexual contact patterns and the predicted demographic impact of AIDS. *Nature, Lond.* **352**, 581–589.

Anderson, R.M., May, R.M. & Gupta, S. 1989 Non-linear phenomena in host-parasite interactions. *Parasitology* **90**(Suppl.), S59–S80.

Anderson, R.M., May, R.M. & McLean, A.R. 1988 Possible demographic consequences of AIDS in developing countries. *Nature, Lond.* **332**, 191–195.

Anderson, R.M., May, R.M., Medley, G.F. & Johnson, A.M. 1986 A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS. *IMA J. Maths Appl. Med. Biol.* **3**, 229–263.

Anderson, R.M., May, R.M., Ng, T.W. & Rowley, J.T. 1992 Age dependent choice of sexual partners and the transmission dynamics of HIV in sub-Saharan Africa. *Phil. Trans. R. Soc. Lond. B* **336**, 135–155.

Antia, R., Koella, J.C., Levin, B.R., Garnett, G.P. & Anderson, R.M. 1993 Parasite evolution in response to immunological defenses. In *Oxford surveys in evolutionary biology* (ed. D. Futuyma & J. Antonovics), pp. 383–405). Oxford: Oxford University Press.

Antia, R., Levin, B.R. & May, R.M. 1994*a* Within-host population dynamics and the evolution and maintenance of microparasite virulence. *Am. Nat.* **144**. (In the press.)

Antia R., Nowak, M. & Anderson R.M. 1994*b* Antigenic variation and the within host dynamics of microparasites. *Parasitology* (Submitted.)

Anzala, A.O., Ndiny-Achola, J.O., Ngugi, E.N., Simonsen, J.N., Kreiss, J.K. & Plummer, F.A. 1993 Rapid progression to disease in African prostitutes with HIV-1 infection. *Poster 149, The first National Conference on Human Retroviruses and Related Infections, Washington DC, December 1993.*

Aral, S.O. & Holmes, K.K. 1990 Epidemiology of sexual behaviour and sexually transmitted diseases. In *Sexually transmitted diseases* (ed. K. K. Holmes, P.-A. Mardh, P. F. Sparling, P. J. Wiesner, J. Cates, W. S. M. Lemon & W. E. Stamm). New York: McGraw-Hill Book Company.

Aron, J.L. & Schwartz, I.B. 1984 Seasonality and period-doubling bifurcations in an epidemic model. *J. theor. Biol.* **110**, 665–679.

Arya, O.P., Nsanzumuhire, H. & Taber, S.R. 1973 Clinical, cultural and demographic aspects of gonorrhoea in a rural community in Uganda. *Bull. Wld Hlth Org.* **49**, 587-595.

Babad, H.R., Nokes, D.J., Gay, N., Miller, E., Morgan-Capner, P. & Anderson, R.M. 1994 Predicting the impact of measles vaccination in England and Wales: model validation and policy projections. *Epidemiol. Infect.* (In the press.)

Bailey, N.J.T. 1975 *The mathematical theory of infectious diseases and its application*. London: Griffin.

Barbour, A.G. & Fish, D. 1993 The biological and social phenomenon of Lyme disease. *Science, Wash.* **260**, 1610–1616.

Barry, J.D. & Turner, C.M.R. 1991 The dynamics of antigenic variation and growth of African trypanosomes. *Parasitol. Today* **7**, 207–211.

Bartlett, M.S. 1957 Measles periodicity and community size. *Jl R. Statist. Soc. A* **120**, 48–70.

Beck, K. 1984 Coevolution: mathematical analysis of host-parasite interactions. *J. Math. Biol.* **19**, 63–78.

Black, F.L. 1993 Why did they die? *Science, Wash.* **258**, 1739–1740.

Bloom, B.R. & Murray, C.J.L. 1992 Tuberculosis – commentary on a reemergent killer. *Science, Wash.* **257**, 1055–1064.

Blower, S.M., Hartel, D., Dowlatabadi, H., Anderson, R.M. & May, R.M. 1991 Drugs, sex, and HIV: a mathematical model for New York city. *Phil. Trans. R. Soc. Lond. B* **321**, 171–187.

Blower, S.M. & McLean, A.R. 1991 Mixing ecology and epidemiology. *Proc. R. Soc. Lond. B* **245**, 187–192.

Bolker, B. 1993 Chaos and complexity in measles models: a

comparative study. *IMA J. Maths Appl. Med. Biol.* **10**, 83–95.
Bolker, B. & Grenfell, B.T. 1993 Chaos and complexity in measles dynamics. *Proc. R. Soc. Lond. B* **251**, 75–81.
Bongaarts, J. 1994 Population policy options in the developing world. *Science, Wash.* **263**, 771–776.
Bremermann, H.J. & Thieme, H.R. 1989 A competitive exclusion principle for pathogen virulence. *J. Math. Biol.* **27**, 179–190.
Brown, K.N. & Brown, I.N. 1965 *Nature, Lond.* **208**, 1286–1288.
Brownlee, J. 1906 Statistical studies in immunity: the theory of epidemics. *Proc. R. Soc. Edinb.* **26**, 484–521.
Bruce-Chwatt, L.J. 1952 Malaria in African infants and children in southern Nigeria. Annls Trop. Med. & Parasitol. **46**, 173–200.
Brunham, R.C., Cheang, M., McMaster, J., Garnett, G.P. & Anderson, R.M. 1993 *Chlamydia trachomatis*, infertility, and population growth in sub-Saharan Africa. *Sex. Trans. Dis.* **20**(**3**), 168–173.
Brunham, R.C., Garnett, G.P., Swinton,J. & Anderson R.M. 1992 Gonococcal infection and human infertility in sub-Saharan Africa. *Proc. R. Soc. Lond. B* **246**, 173–177.
Bundy, D.A.P. & Guyatt, H.L. 1992 Cost-analysis of schistosomiasis. *Trans. R. Soc. trop. Med. Hyg.* **86**, 646–648.
Busenberg, S. & Cooke, K. 1993 *Vertically transmitted diseases*. Berlin: Springer Verlag.
Carael, M., Cleland, J. & Coll, G. 1994 Reported extra-marital sex: a cross-national study. Implications for STD/HIV transmission. In *AIDS Prevention and Impact in the Developing World: the Contribution of Demography and Social Science*. Liege: Ordina Publications. (In the press.)
Cates, W., Farley, I.M.M. & Rowe, P.J. 1985 Worldwide patterns of infertility: is Africa different? *Lancet ii*, 596–598.
Cates, W., Rolfs, R. & Aral, S. 1990 Sexually transmitted diseases, pelvic inflammatory disease and infertility. An epidemiologic update. *Epidemiol. Rev.* **12**, 199–220.
Cates, W. & Wasserheit, J.N. 1991 Genital chlamydial infections: epidemiology and reproductive sequalae. *Am. J. Obstet. Gynec.* **164**, 1771–1781.
Cavacini, L.A., Long, C.A. & Weidanz, W.P. 1986 T cell immunity in murine malara: adoptive transfer of resistance to *Plasmodium chabaudi adami* in nude mice with splenic T cells. *Infect. Immun.* **52**, 637–643.
CDC 1993 *Morbidity and Mortality Weekly Record* **42**, 601–603.
Cerottini, J.-C. & MacDonald, H.R. 1989 The cellular basis of T cell memory. *A. Rev. Immunol.* **7**, 77–89.
Chagnon, N.A. & Melancon, T.F. 1981 Reproduction, numbers of kin and epidemics in tribal populations: a case study. In *Population and biology* (ed. N. Keyfitz), pp. 147–168). Liege: Ordina Editions.
Christenson, B., Heller, L. & Bottiger M. 1983 The immunizing effect and reactogenicity of two live attenuated mumps vaccines in Swedish schoolchildren. *J. biol. Standard.* **11**, 323–332.
Clark, E.A. & Ledbetter, J.A. 1994 How B and T cells talk to each other. *Nature, Lond.* **367**, 425–428.
Connor, R.I., Mohri, H., Cao, Y. & Ho, D.D. 1993 Increased viral burden and cytopathicity correlate temporally with CD4+ T-lymphocyte decline and clinical progression in human immunodefeciency virus type 1-infected individuals. *J. Virol.* **67**, 1772–1777.
Day, K.P., Koella, J.C., Nee, S., Gupta, S. & Read, A.F. 1992 Population genetics and dynamics of *Plasmodium falciparum*: an ecological view. *Parasitology* **104**, 535–552.
Day, K.P. & Marsh, K. 1991 Naturally aquired immunity to *Plasmodium falciparum*. *Immunol. Today* **12**, A68–A71.
De Cock, K.M. & Brun-Vezinet, F. 1989 Epidemiology of HIV-2 infection. *AIDS* **3**, S89–S95.
De Cock, K.M., Zadi, F., Adorlolo, G., Diallo, M.O., Sassan-Morokro, M., Ekpini, E., Sibailly, T., Doorly, R., Batter, V., Brattegaard, K. & Gayle, H. 1994 Retrospective study of maternal HIV-1 and HIV-2 infections and child survival in Abidan, Cote d'Ivoire. *Br. med. J.* **308**, 441–443.
Dearsley, A.L., Sinden, R.E. & Self, I.A. 1990 Sexual development in malarial parasites: gametocyte production, fertility and infectivity to the mosquito vector. *Parasitology* **100**, 359-368.
Defoe, D. 1722 *A journal of the plague year*. London: Thomas Tegg. (Reprinted 1992 (ed. P. R. Backscheider). London: Norton.)
Delwart, E.L., Shaper, E.G., Louwagie, J., McMutchan, F.E., Grez, M., Rubsamen-Waigmann, H. & Mullins, J.I. 1993 Genetic relationships determined by a DNA heteroduplex mobility assay: analysis of HIV-1 *env* genes. *Science, Wash.* **262**, 1257–1261.
Diekmann, O., Heesterbee, J.A. & Metz, J.A.J. 1990 On the definition and computation of the basic reproductive ratio R_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382.
Dietz, K. 1976 The incidence of infectious diseases under the influence of seasonal fluctuations. *Lecture notes in biomathematics* **11**, 1–15.
Dietz, K. 1980 Models for vector-borne parasitic diseases. *Lecture notes in biomathematics* **39**, 264–277.
Dietz, K. & Hadler, K.P. 1988 Epidemiological models for sexually transmitted diseases. *J. Math. Biol.* **26**, 1–25.
Dietz, K. & Schenzle, D. 1985 Mathematical models for infectious disease statistics. In *A celebration of statistics* (ed. A. C. Atkinson & S. E. Fienberg), pp. 167–204. New York: Spinger Verlag.
Dougherty, J.P. & Trenin, H.M. 1988 Determination of the rate of base-pair substitution and insertion mutations in retrovirus replication. *J. Virol.* **62**, 2817–2822.
Dye, C.M. & Hasibeder, G. 1986 Population dynamics of mosquito-borne disease: effects of flies which bite some people more frequently than others. *Trans. R. Soc. trop. Med. Hyg.* **83**, 69–77.
Editorial 1994 AIDS: the third wave. *Lancet* **343**, 186188.
Eling, W.C.M., Zon, A.A.J.C. & Jerusalem, C. 1977 The course of a *Plasmodium berghei* infection in six different mouse strains. *Parasitol. Res.* **54**, 29–45.
Embretson, J., Zupancic, M., Ribas, J.L., Burke, A., Racz, P., Tenner-Racz, K. & Hasse, A.T. 1993 Massive covert infection of helper T lymphocytes and macrophages by HIV during the incubation period of AIDS. *Nature, Lond.* **362**, 359–362.
European Collaborative Study 1992 Risk factors for mother-to-child transmission. *Lancet* **339**, 1007–1012.
Ewald, P. 1993 The evolution of virulence. *Scient. Am.*(April), 56–62.
Farr, W. 1840 *Progress of epidemics* Second Report of the Registrar General of England pp. 91–98.
Fauci, A.S. 1993*a* Immunopathogenesis of HIV infection. *J. Aquired Immune Defic. Synd.* **6**, 655–662.
Fauci, A.S. 1993*b* Multifactorial nature of human immunodeficiency virus disease: implications for therapy. *Science, Wash.* **262**, 1011–1018.
Fenner, F. & Ratcliffe, F.N. 1965 *Myxomatosis*. Cambridge: Cambridge University Press.
Fescharek, R., Quast, U., Maass, G., Merkle, W. & Schwarz, S. 1990 Measles-mumps vaccination in the FRG: an empirical analysis after 14 years of use. II. Tolerability and analysis of spontaneously reported side effects. *Vaccine* **8**, 446–456.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Forsyth, K.P., Anders, R.F., Kemp, D.J. & Alpers, M.P. 1988 New approaches to the serotypic analysis of the epidemiology of *Plasmodium falciparum*. *Proc. R. Soc. Lond. B* **321**, 159–164.

Frank, S.A. 1992 A kin selection model for the evolution of virulence. *Proc. R. Soc. Lond. B* **250**, 195–197.

Freeman, G.J., Gribben, J.G., Boussiotis, V.A., Ng, J.W., Restivo, V.A., Lombard, L.A., Gray, G.S. & Nadler, L.M. 1993 Cloning of B7-2 – a CTLA-4 counter-receptor that costimulates human T-cell proliferation. *Science, Wash.* **262**, 909–911.

Freitas, A.A. & Rocha, B.B. 1993 Lymphocyte lifespans: homeostasis, selection and competition. *Immunol. Today* **14(1)**, 25–29.

Frost, S.D.W. & McLean, A.R. 1994 Germinal centre destruction as a major pathway of HIV pathogenesis. *J. Aquired Immune Defic. Synd.* (In the press.)

Furez, J. & Contreras, G. 1990 Vaccine-related mumps meningitis – Canada. *Can. Dis. Weekly Record* **16**, 253–254.

Garnett, G.P. & Anderson, R.M. 1993*a* Factors controlling the spread of HIV in heterosexual communities in developing countries: patterns of mixing between different age and sexual activity classes. *Phil. Trans. R. Soc. Lond. B* **342**, 137–159.

Garnett, G.P. & Anderson, R.M. 1993*b* Contact tracing and the estimation of sexual mixing patterns: the epidemioloy of gonococcal infections. *Sex. Trans. Dis.* **20**, 181–191.

Garnett, G.P. & Anderson, R.M. 1993*c* No reason for complacency about the potential demographic impact of AIDS in Africa. *Trans. R. Soc. trop. Med. Hyg.* **87** (Suppl. 1), 19–22.

Garnett, G.P. & Anderson, R.M. 1994 Balancing sexual partnerships in an age and activity stratified models of HIV transmission in heterosexual populations. *IMA J. Maths. appl. med. Biol.* (In the press.)

Garnett, G.P., Swinton, J., Brunham, C. & Anderson, R.M. 1992 Gonococcal infection, infertility, and population growth: the influence of heterogeneity in sexual behaviour. *IMA J. Maths Applied Med. Biol.* **9**, 127–144.

Greenwood, B.M., Marsh, K. & Snow, R.W. 1992 Why do some African children develop severe malaria? *Parasitol. Today* **7**, 277–281.

Gregson, S., Garnett, G.P. & Anderson, R.M. 1994*a* Assessing the potential impact of the HIV-1 epidemic on orphanhood and the demographic structure of populations in sub-Saharan Africa. *J. Pop. Studies.* (In the press.)

Gregson, S., Garnett, G.P. & Anderson, R.M. 1994*b* Is HIV-1 likely to become a leading cause of adult mortality in sub-Saharan Africa? *J. Aquired Immune Defic. Synd.* **7**, 839–852.

Gregson, S., Garnett, G.P., Shakespeare, R., Foster, G. & Anderson, R.M. 1994*c* Mathematical model simulations and empirical estimates of the impact of HIV-1 on orphanhood in sub-Saharan Africa. *Health Transit. Rev.* (In the press.)

Grenfell, B.T. & Anderson, R.M. 1985 The estimation of age related rates of infection from case notifications and serological data. *J. Hyg.* **95**, 419–436.

Griffith, H.R. 1963 Gonorrohoea and fertility in Uganda. *Eugen. Rev.* **55**, 103–108.

Griffiths, D.A. 1974 A catalytic model of infection for measles. *Appl. Statist.* **23**, 330–339.

Gupta, S., Anderson, R.M. & May, R.M. 1989 Networks of sexual contacts: implications for the transmission dynamics of HIV. *AIDS* **3**, 807–817.

Gupta, S. & Day, K.P. 1994 A theoretical framework for the immunoepidemiology of *Plasmodium falciparum* malaria. *Parasite Immunol.* (In the press.)

Gupta, S., Hill, A.V.S., Kwiatkowski, D., Greenwood, A.M., Greenwood, B.M., & Day, K.P. 1994*a* Parasite virulence and disease patterns in *Plasmodium falciparum* malaria. *Proc. natn. Acad. Sci. U.S.A.* (In the press.)

Gupta, S., Swinton, J. & Anderson, R.M. 1994*b* Theoretical studies of the effects of heterogeneity in the parasite population on the transmission dynamics of malaria. *Proc. R. Soc. Lond.* B **256**, 231–238.

Gupta, S., Trenholme, K., Anderson, R.M. & Day, K.P. 1994*c* Antigenic diversity and the transmission dynamics of *Plasmodium falciparum*. *Science, Wash.* **263**, 961–963.

Guyatt, H.L., Bundy, D.A.P. & Evans, D. 1994 A population dynamic approach to the cost-effectiveness analysis of mass treatment: effects of treatment frequency on *Ascaris lumbricoides* infection. *Trans. R. Soc. trop. Med. Hyg.* (In the press.)

Gyllensten, U.B., Lashkari, D. & Erlich, H.A. 1990 Allelic diversification at the class-II DQB locus of the mammalian major histocompatibility complex. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1835–1839.

Haldane, J.B.S. 1949 Disease and evolution. *La Ricerca Scientifica* **19** (Suppl.), 68–76.

Haraldsdottir, S., Gupta, S. & Anderson, R.M. 1992 Preliminary studies of sexual networks in a male-homosexual community in iceland. *J. Acquired Immune Defic. Synd.* **5**, 374–381.

Harvey, P.H., Holmes, E.C., Mooers, A.O. & Nee, S. 1994 Inferring evolutionary processes from molecular phylogenies. In *Models in phylogeny reconstruction* (Systematics Association vol. 52) (ed. R. W. Scotland, D. G. Sibert & D. M. Williams), pp. 313–333. Oxford University Press.

Hendriks, J.C.M., Medley, G.F., van Grievsen, G.J.P., Coutinho, R.A., Heisterkamp, S.H. & van Druten, H.A.M. 1993 The treatment-free incubation period of AIDS in a cohort of homosexual men. *AIDS* **7**, 231–239.

Henschen, F. 1966 *The History of Diseases*. London: Longmans.

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

Hetzel, C. 1993 The dynamics of the immune response to infection: mathematical models and experimental investigations. Ph.D. thesis, London University.

Hetzel, C. & Anderson, R.M. 1994 Dynamics of merozoite invasion of erythrocytes. *Parasitology* (In the press.)

Hill, A.V.S., Allsopp, C.E.M., Kwiatkowski, D., Anstey, N.M., Twumasi, P., Rowe, P.A., Bennett, S., Brewster, D., Mcmichael, A.J. & Greenwood, B.M. 1991 Common West African HLA antigens are associated with protection from severe malaria. *Nature, Lond.* **352**, 595–600.

Hira, S.K., Kamanga, J., Bhat, G.J., Muvale, C., Tembo, G., Luo, N. & Perine, D.L. 1989 Perinatal transmission of HIV-1 in Zambia. *Br. med. J* **299**, 1250–1252.

Hogsborg, M. & Aaby, P. 1992 Sexaul relations, use of condoms and perceptions of AIDS in an urban area of Guinea-Bissau with a high prevalence of HIV-2. In *Anthropological and socio-cultural studies in the transmission* of HIV (ed. T. Dyson), pp. 29–36. Liege: Ordina Publications.

Holmes, E.C., Zhang, L.Q., Simmonds, P., Ludlam, C.A. & Leigh Brown, A.J. 1992 Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single patient. *Proc. natn. Acad. Sci. U.S.A.* **89**, 4835–4839.

Holzer, T.J., Heynen, C.A., Novak, R.M., Pitrak, D.L. & Dawson, G.J. 1993 Frequency of cells positive for

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

HIV-1 p24 antigen by flow cytometry. *AIDS* **7** (Suppl. 2), S3–S6.

Hsu, H.-Y., Chang, M.-H., Chen, D.-S., Lee, C.-Y. & Sung, J.-L. 1986 Baseline seroepidemiology of hepatitis B virus in children in Tapei, 1984: a study just before mass hepatitis B vaccination. *J. Med. Virol.* **18**, 301–307.

Hyman, J.M. & Stanley, E.A. 1994 A risk-based heterosexual model for the AIDS epidemic with sexual partner selection. In *Modeling the AIDS epidemic* (ed. E. Kaplan & M. L. Brandeau) pp. 331–364. New York: Raven Press.

Isaacs, D. & Menser, M. 1990 Measles, mumps, rubella and varicella. In *Modern vaccines* (ed. E. R. Moxon), pp. 50–58. London: Edward Arnold.

Jain, M.K., John, T.J. & Keusch, G.T. 1994 A review of human immunodeficiency virus infection in India. *J. Acquired Immune Defic. Synd.* **7**, 1185–1194.

Jacquez, J.A., Simon, C.P., Koopman, J., Sattenspiel, J. & Perry, T. 1988 Modelling and analysing HIV transmission: the effect of contact patterns. *Math. Biosci.* **92**, 119–199.

Jenkins, M.K. 1992 The role of cell division in the induction of clonal anergy. *Immunol. Today* **13**(**2**), 69–73.

Johnson, A.M., Wadsworth, J., Wellings, K., Bradshaw, S. & Field, J. 1992 Sexual lifestyles and HIV risk. *Nature, Lond.* **360**, 410–412.

Johnson, A.M., Wadsworth, J., Wellings, K. & Field, J. 1994 *Sexual attitudes and lifestyles*. Oxford: Blackwell Scientific Publications.

Jones, K.R., Hickling, J.K., Targett, G.A.T. & Playfair, J.H.L. 1990 Polyclonal *in vitro* proliferative responses from non-immune donors to *Plasmodium falciparum* malaria antigens require UCHL1 + (memory) T cells. *Eur. J. Immun.* **20**, 307–315.

Kaplan, E.H. 1990 Modeling HIV infectivity: must sex acts be counted? *Journal of Acquired Immune Deficiency Syndromes* **3**, 55–61.

Kermack, W.O. & McKendrick A.G. 1927 A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A* **115**, 700–721.

Kevrekedis, I.G., Zecha, A.D. & Perelson, A.S. 1988 Modelling dynamic aspects of the immune response: I. T cell proliferation and the effect of IL-2. In *Theoretical immunology* (ed. A. S. Perelson), pp. 167–197. New York: Addison-Wesley.

Klein, J., Takahata, N. & Ayala, F.J. 1993 MHC polymorphism and human origins. *Scient. Amer.* **269**, 78–83.

Knell, A.J. 1991 *Malaria*. Oxford: Oxford University Press.

Knight, S.C. 1987 Lymphocyte proliferation assays. In *Lymphocytes: a practical approach* (ed. G. G. B. Klaus), pp. 189–207). Oxford: IRL Press.

Krause, R.M. 1992 The origin of plagues: old and new. *Science, Wash.* **257**, 1073–1078.

Kretzschmar, M. 1989 Persistent solutions in a model for parasitic infections. J. math. Biol. **27**, 549–573.

Laga, M., Nzila, N. & Goeman, J. 1991 The interrelationship of sexually-transmitted diseases and HIV-infection – implications for the control of both epidemics in africa. *Aids* **5**, 55–63.

Lallemant, M., Lallemantlecoeur, S., Cheynier, D., Nzingoula, S., Jourdain, G., Sinet, M., Dazza, M.C., Blanche, S., Griscelli, C. & Larouze, B. 1989 Mother-child transmission of HIV-1 and infant survival in Brazzaville, Congo. *Aids* **3**, 643–646.

Langhorne, J., Simon-Haarhaus, B. & Meding, S.J. 1990 The role of CD4+ T cells in the protective response to *Plasmodium chabaudi* in vitro. *Immunol. Let.* **25**, 101–108.

Lederberg, J., Shope, R.E. & Oaks, S.C. Jr (eds) 1992 *Emerging infections* Washington D.C.: National Academy Press.

Lenski, R.E. & May, R.M. 1993 The evolution of virulence in parasite and pathogens: reconciliation between two competing hypotheses. *J. theor. Biol.* **169**, 253–265.

Levin, S.A. 1983 Some approaches to modelling of coevolutionary interactions. In *Coevolution* (ed. M. Ntecki), pp. 21–65. Chicago: University of Chicago Press.

Levin, S. & Pimentel, D. 1981 Selection of intermediate rates of increase in parasitehost systems. *Am. Nat.* **117**, 308–315.

Levy, J.A. 1993 Pathogenesis of human immunodeficiency virus infection. *Microbiol. Rev.* **57**, 183–289.

Lipsitch, M., Nowak, M., Anderson, R.M. & May, R.M. 1994 The evolution of virulence in sexually transmitted HIV/AIDS. (In the press.)

Lu, W., Shih, W., Tourani, J., Eme, D., Alter H.J. & Andrieu, J. 1993 Lack of isolate-specific neutralizing activity is correlated with an increased viral burden in rapidly progressing HIV-1-infected patients. *AIDS* **7** (Suppl. 2), S91–S100.

MacDonald, G. 1957 *The epidemiology and control of malaria*. Oxford: Oxford University Press.

Maddox, J. 1993 The dark side of molecular biology. *Nature, Lond.* **363**, 13.

Mastro, T.D., Satten, A., Nopkesorn, T., Sangkharomya, S. & Longini, I.M. Jr 1994 Probability of female-to-male transmission of HIV-1 in Thailand. *Lancet* **343**, 204–207.

Matis, L.A., Glimcher, L.H., Paul, W.E. & Schwartz, R.H. 1983 Magnitude of the response of histocompatibility-restricted T cell clones is a function of the product of the concentration of antigen and Ia molecules. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6019–6023.

May, R.M. & Anderson, R.M. 1979 Population biology of infectious diseases: part II. *Nature, Lond.* **280**, 455–461.

May, R.M. & Anderson, R.M. 1983 Epidemiology and genetics in the coevolution of parasites and hosts. *Proc. R. Soc. Lond. B* **219**, 281–313.

May, R.M. & Anderson, R.M. 1984 Spatial heterogeneity and the design of immunisation programmes. *Math. Biosci.* **72**, 83–111.

May, R.M. & Anderson, R.M. 1987 The transmission dynamics of HIV infection. *Nature, Lond.* **326**, 137–142.

May, R.M. & Anderson, R.M. 1989 Heterogeneities, cofactors, and other aspects of the transmission of HIV/AIDS. In *Current topics in AIDS*, pp. 33–67. New York: John Wiley.

May, R.M. & Anderson, R.M. 1994 Heterogeneity in the transmission of sexually transmitted diseases: sex acts and sex partners. (In preparation.)

McLean, A.M. & Anderson, R.M. 1988 Measles in developing countries. Part I. Epidemiological parameters and patterns. *Epidemiol. & Infect.* **100**, 111–133.

McLean, A.R. & Kirkwood, T.B.L. 1990 A model of human immunodeficiency virus infection in T helper clones. *J. theor. Biol.* **147**, 177–203.

McLean, A.R., Nokes, D.J. & Anderson, R.M. 1991 Model based comparisons of measles immunization strategies using high dose EdmonstonZagreb type vaccines. *Int. J. Epidemiol.* **20**, 1107–1117.

McNeill, W.H. 1977 *Plagues and peoples*. Oxford: Basil Blackwell.

McQullan, G.P., Townsend, T.T., Fields, H.A., Carroll, M., Leahy, M. & Polk, B.F. 1989 Seroepidemiology of hepatitis B virus infection in the United States. *Am. J. Med.* **87**, 7–10.

Medley, G.F., Guyatt, H.L. & Bundy, D.A.P. 1994 A quantitative framework for evaluating the effect of community treatment on the morbidity due to ascariasis. *Parasitology*. **106**, 211–221.

Melbye, M. & Smith, E. 1993 Preventing HIV spread in

homo/bisexual men: how effective is it? Experience from the national mandatory HIV registry in Denmark. *J. Aquired Immune Defic. Synd.* **6**, 5536–5537.

Michie, C.A., McLean, A.R., Alcock, C. & Beverley, P.C.L. 1992 Life-span of human lymphocyte subsets defined by CD45 isoforms. *Nature, Lond.* **360**, 264–265.

Miller, D., Madge, N., Diamond, J., Wadsworth, J. & Ross, E. 1993 Pertussis immunisation and serious acute neurological illness in children. *Br. med. J.* **307**, 1171–1176.

Molineaux, L. & Gramiccia, G. 1980 *The Garki Project: research on the epidemiology and control of Malaria in the Sudan Savanna of West Africa.* Geneva: World Health Organisation.

Morgan-Capner, P., Wright, J., Miller, C.L. & Miller, E. 1988 Surveillance of measles, mumps and rubella by age. *Br. Med. J.* **297**, 770–772.

Morse, M.M. 1993 *Emerging viruses.* Oxford: Oxford University Press.

Morse, S.S. & Schluederberg, A. 1990 Emerging viruses: the evolution of viruses and viral diseases. *J. infect. Dis.* **162**, 358–360.

Moxon, E.R., Rainey, P.B., Nowak, M. & Lenski, R.E. 1994 Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* **4**, 24–33.

Mueller, D.L., Jenkins, M.K. & Schwartz, R.H. 1989 Clonal expansion versus functional clonal inactivation: a costimulatory signalling pathway determines the outcome of T cell antigen receptor occupancy. *A. Rev. Immunol.* **7**, 445–480.

Muench, H. 1959 *Catalytic models in epidemiology.* Cambridge, Massachusetts: Harvard University Press.

Mulder, D.W., Kamali, A., Nakyinge, J. & Nunn, A.J. 1994 HIV-1 incidence and HIV-1 associated mortality in a rural Ugandan population. *Lancet* **343**, 1021–1023.

Murray, G.D. & Cliff, A.D. 1975 A stochastic model of measles epidemics in a multi-region setting. *Inst. Br. Geog.* **2**, 158–174.

Nee, S., May, R.M. & Harvey, P.H. 1994 The reconstructed evolutionary process. *Phil. Trans. R. Soc. Lond. B* **344**, 305–311.

Nokes, D.J. & Anderson, R.M. 1988 The use of mathematical models in the epidemiological study of infectious diseases and in the design of mass immunization programmes. *Infect. Immun.* **101**, 120.

Nokes, D.J. & Anderson, R.M. 1991 Vaccine safety versus vaccine efficacy in mass immunisation programmes. *Lancet* **338**, 1309–1312.

Nokes, D.J., Anderson, R.M. & Anderson, M.J. 1986 Rubella epidemiology in South East England. *J. Hyg.* **86**, 291–304.

Nokes, D.J. & Cutts, F.T. 1993 Immunization in the developing world: strategic challenges. *Trans. R. Soc. trop. Med. Hyg.* **87**, 353–354.

Nokes, D.J. & Swinton, J. 1994 The control of childhood infection by pulse vaccination: an epidemiological approach. *IMA J. math. appl. Med. Biol.* (In the Press.)

Nokes, D.J., Wright, J., Morgan-Capner, P. & Anderson, R.M. 1990 Serological study of the epidemiology of mumps-virus infection in north-west england. *Epidemiol. Infect.* **105**, 175–195.

Nold, A. 1980 Heterogeneity in disease transmission modelling. *Math. Biosci.* **52**, 227–240.

Nowak, M. 1991 The evolution of viruses: competition between horizontal and vertical transmission of mobile genes. *J. theor. Biol.* **150**, 339–347.

Nowak, M.A., Anderson, R.M., McLean, A.R., Wolfs, T.F.W., Goudsmit, J. & May, R.M. 1991 Antigenic diversity thresholds and the development of AIDS. *Science, Wash.* **254**, 963–969.

Nowak, M. & May, R.M. 1993 AIDS pathogenesis: mathematical models of HIV and SIV infections. *AIDS* **7** (Suppl. 1), S3–S18.

Nowak, M. & May, R.M. 1994*a* Super infection, metapopulation dynamics and the evolution of diversity. *J. theor. Biol.* (In the press.)

Nowak, M.A. & May, R.M. 1994*b* Superinfection and the evolution of virulence. *Proc. R. Soc. Lond. B* **255**, 81–89.

Nowak, M.A., May R.M. & Anderson, R.M. 1990 The evolutionary dynamics of HIV-1 quasispecies and the development of immunodeficiency disease. *AIDS* **4**, 1095–1105.

O'Hehir, R.E. & Lamb, J.R. 1990 Induction of specific clonal anergy in human T lymphocytes by *Staphylococcus aureus* entrotoxins. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8884–8888.

Olsen, L.F. & Schaffer, W.M. 1990 Chaos versus noisy periodicity – alternative hypotheses for childhood epidemics. *Science, Wash.* **249**, 499–504.

Ou, C.-C., Takebe, Y., Weniger, B.G., Luo, C.-C., Kalish, M.L., Auwanit, W., Yamazaki, S., Gayle, H.D., Young, N.L. & Schochetman, G. 1993 Independent introduction of two major HIV-1 genotypes into distinct high-risk populations in Thailand. *Lancet* **341**, 1171–1174.

Pantaleo, G., Graziosi, C., Demarest, J.F., Butini, L., Montroni, M., Fox, C.H., Orenstein, J.M., Kotler, D.P. & Fauci, A.S. 1993 HIV-infection is active and progressive in lymphoid-tissue during the clinically latent stage of disease. *Nature, Lond.* **362**, 355–358.

Patriarca, P.A., Foege, W.H. & Swartz, T.A. 1993 Progress in polio eradication. *Lancet* **342**, 1461–1464.

Pepin, J., Morgan, G., Dunn, D., Gavao, S., Mendy, M., Gaye, I., Scollen, N., Tedder, R. & Whittle, H. 1991 HIV-2 induced immunosuppression among asymptomatic West African prostitutes: evidence that HIV-2 is pathogenic, but less so than HIV-1. *AIDS* **5**, 1165–1172.

Perelson, A.S. 1990 Theoretical immunology. In *1989 Lecture in Complex Systems, SFI Studies in the Sciences of Complexity* (ed. E. Jen), pp. 465–499. New York: AddisonWesley.

Perry, K.R., Brown, D.W.G., Parry, J.V., Panday, S., Pipkin, C. & Richards, A. 1993 Detection of measles, mumps, and rubella antibodies in saliva using antibody capture radioimmunoassay. *J. Med. Virol.* **40**, 235–240.

Piot, P., Kapita, B.M. & Were, J.B.O. (eds) 1991 *AIDS in Africa.* London: Current Science.

Popow–Kraupp, T., Kundi, M., Ambrosch, F., Vanura H. & Kunz, C.A. 1986 A controlled trail for evaluating two live attenuated vaccines (Urabe Am 9-Schwarz and Jerl Lynn-Moraten) in young children. *J. Med. Virol.* **18**, 69–79.

Potterat, J.J., Rothenberg, R.B., Woodhouse, D.E., Muth, J.B., Pratts, C.S., & Fogle, J.S. 1985 Gonorrhea. *Sex. Trans. Dis.* **12**, 25–32.

Potts, M., Anderson, R.M. & Boily, M.-C. 1991 Slowing the spread of HIV in developing countries: policy issues. *Lancet ii*, 608–613.

Poulsen, A.G., Kvinesdal, B., Aaby, P., Molbak, K., Fredriksen, K., Dias, F. & Lauritzen, E. 1989 Prevalence of mortality from human immunodeficiency virus type 2 in Bissau, West Africa. *Lancet ii*, 827–830.

RCGP 1974 A retrospective study of the complications of mumps. *J. RCGP* **24**, 552–555.

Reibnegger, G., Fuchs, D., Hausen, A., Werner, E.R., Wernerfelmayer, G., Dierich, M.P. & Wachter, H. 1989 Stability analysis of simple models for immune cells

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

interacting with normal pathogens and immune system retroviruses. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2026–2030.

Roberts, D.J., Biggs, B.-A., Brown, G. & Newbold, C.I. 1993 Protection, pathogenesis and phenotypic plasticity in *Plasmodium falciparum* malaria. *Parasitol. Today* **9**(**8**), 281–286.

Roberts, D.J., Craig, A.G., Berendt, A.R., Pinches, R., Nash, G., Marsh, K. & Newbold, C.I. 1992 Rapid switching to multiple antigenic and adhesive phenotypes in malaria. *Nature, Lond.* **357**, 689–692.

Roth, R.L. & Herman, R. 1979 *Plasmodium berghei*: correlation of *in vitro* erythrophagocytosis with the dynamics of early onset anaemia and reticulocytosis. *Expl Parasitol.* **47**, 169–179.

Rothenberg, R.B. & Potterat, J.J. 1988 Temporal and social aspects of gonorrhea transmission: the force of infectivity. *Sex. Trans. Dis.* **15**, 88–92.

Ryder, R.W., Ng, W., Hassig, S.E., Behets, F., Rayfield, M., Ekingola, B. *et al.* 1989 Perinatal transmission of HIV-1 to infants of seropositive women in Zaire. *New Engl. J. Med.* **320**, 1638–1642.

Ryder, R.W. & Timmerman, M. 1991 The effect of HIV-1 infection during pregnancy and the perinatal period on maternal and child health in Africa. *AIDS* **5** (Suppl. 1), S75–S86.

Saracco, A., Musicco, M., Nicholosi, A., Angarano, G. et al. 1993 Man-to-woman sexual transmission of HIV: longitudinal study of 343 steady partners of infected men. *J. Aquired Immune Defic. Synd.* **6**, 497–502.

Schaffer, W.M. & Knot, M. 1985 Nearly one dimensional dynamics in an epidemic. *J. theor. Biol.* **112**, 403–427.

Schenzle, D. 1984 An age-structured model of pre- and post-vaccination measles transmission. *IMA J. Maths Applied Med. Biol.* **1**, 169–191.

Schwartz, I.B. 1992 Small amplitude, long period outbreaks in seasonally driven epidemics. *J. Math. Biol.* 30, 473—491.

Schwartz, R.H. 1994 Immunologic tolerance. In *Fundamental immunology* (ed. W. E. Paul), pp. 48–54. New York: Raven Press.

Schwartz, I.B. & Smith, H.L. 1983 Infinite subharmonic bifurcation in an SEIR epidemic model. *J. math. Biol.* 18, 233–253.

Schweitzer, A.N. & Anderson, R.M. 1992 Dynamic interaction between CD4+ T cells and parasitic helminths: mathematical models of heterogeneity in outcome. *Parasitology* **105**, 513–522.

Schweitzer, A.N., Swinton, J. & Anderson, R.M. 1993 Dynamic interaction between *Leishmania* infection in mice and Th1-type CD4+ T-cells: complexity in outcome without a requirement fot Th2-type responses. *Parasite Immunol.* **15**, 85–99.

Seger, J. 1988 Dynamics of simple host parasite models with more than two genotypes in each species. *Phil. Trans. R. Soc. Lond. B* **319**, 541–555.

Seger, J. & Hamilton, W.D. 1988 Parasite and sex. In *The evolution of sex* (ed. R. E. Michod & B. R. Levin), pp. 176–193). Sunderland, Massachusetts: Sinauer.

Sherris, J.D. & Fox, G. 1983 Infertility and sexually transmitted disease: a public health challenge. *Pop. Report* **9** (L-4), L114–L151.

Sinnecker, H. 1976 *General epidemiology*. London: John Wiley & Sons.

Smith, H.L. 1982 On the basin of attraction of a perturbed attractor. *Nonlinear Anal.* **6**, 911-917.

Snow, R.W., Armstrong Schellenberg, J.R.M., Peshu, N., Forster, D., Newton, C.R.J.C., Winstanley, P.A., Mwangi, I., Waruiru, C., Warn, P.A., Newbold, C. & Marsh, K. 1993 Periodicity and space-time clustering of severe childhood malaria on the coast of Kenya. *Trans. R. Soc. trop. Med. Hyg.* **87**, 386–390.

Soper, M.A. 1929 The interpretation of periodicity in disease prevalence. *Jl. R. statist. Soc. A* **92**, 34–61.

Stuart-Harris, C.H. 1982 The epidemiology of influenza: key facts and remaining problems. In *Influenza models* (ed. P. Selby), pp. 87–103. Boston: MIT Press.

Sugihara, G., Grenfell, B. & May, R.M. 1990 Distinguishing error from chaos in ecological time-series. *Phil. Trans. R. Soc. Lond.* B **330**, 235–251.

Suss, G., Eichmann, K., Kury, E., Lincke, A. & Langhorne, J. 1988 The role of CD4- and CD8-bearing T lymphocytes in the immune response to the erythrocytic stage of *Plasmodum chabaudi. Infect. Immun.* **56**, 3081–3088.

Suzuki, G., Kawase, Y., Yahara, I., Koyasu, S., Kobayashi, Y. & Schwartz, R.H. 1988 Antigen-induced suppression of the proliferative response of T cell clones. *J. Immun.* **140**, 1359–1365.

Swinton, J., Garnett, G.P., Brunham, R.C. & Anderson, R.M. 1992 Gonococcal infection, infertility, and population growth: endemic states in behaviourally homogeneous growing populations. *IMA J Maths Applied Med. Biol.* **9**, 107–126.

Swinton, J., Schweitzer, A.N. & Anderson, R.M. 1994 Two signal activation as an explanation of high zone tolerance: a mathematical exploration of the nature of the second signal. *J. theor. Biol.* **169**, 23–30.

Tidd, C.W., Olsen L.F. & Schaffer W.M. 1993 The case for chaos in childhood epidemics. II. Predicting historical epidemics from mathematical models. *Proc. R. Soc. Lond. B* **234**, 257–273.

Torsella, J.A., Pruitt, K.M. & Lam, C.F. 1992 A method for the analysis of biological transduction phenomena. Math. Biosci. **110**, 191–200.

Ucker, D.S., Meyers, J. & Obermiller, P.S. 1992 Activation-driven T-cell death. 2. Quantitative differences alone distinguish stimuli triggering nontransformed T-cell proliferation or death. *J. Immunol.* **149**, 1583–1592.

Ukkonen, P. & Bronsdorff, C.V. 1988 Rubella immunity and morbidity: effects of vaccination in Finland. *Scand. J. Infect. Dis.* **20**, 255–259.

US Bureau of the Census 1993 *HIV/AIDS surveillance data base*. Center for International Research, Washington.

Vesikari, T., Andre, F.E., Simon, F. *et al.* 1983 Evaluation in young children of the Urabe Am 9 strain of live attenuated mumps vaccine in comparison with the Jerl Lynn strain. *Acta Pediat. Scand.* **72**, 32–40.

Vickerman, K. 1986 Trypanosome sociology and antigenic variation. *Parasitology* **99**, S37–S47.

Viravaidya, M., Obremsky, S.A. & Myers, C. 1992 *The economic impact of AIDS on Thailand*. Working paper No. 4, March 1992. Harvard School of Public Health.

Wawer, M.J., Sewankambo, N.K., Berkley, S., Serwadda, D., Musgave, S.D., Gray, R.H., Musagara, M., Stallings, R.Y. & Konde-Lule, J.K. 1994 Incidence of HIV-1 infection in a rural region of Uganda. *Lancet* **203**, 123–126.

Weid, T. von der & Langhorne, J. 1993 Altered response of CD4+ T cell subsets to *Plasmodium chabaudi chabaudi* in B cell-deficient mice. *Int. Immun.* **5**(**10**), 1343–1348.

Wellings, K., Feild, J., Johnson, A.M. & Wadsworth, J. 1994 *Sexual behaviour in Britain*. London: Penguin Books.

Wellings, K., Feild, J., Wadsworth, J., Johnson, A.M., Anderson, R.M. & Bradshaw, S.A. 1990 Sexual lifestyles under scrutiny. *Nature, Lond.* **348**, 276–278.

Wills, C. & Green, D.R. 1994 A genetic herd-immunity model for the maintenance of MHC polymorphisms. Preprint.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

World Bank 1993 *Disease control priorities in developing countries*. Oxford: Oxford University Press.

Yeager, A.S., Davis, J.H., Ross L.A. & Harvey, B. 1977 Measles immunisation: successes and failures. *J. Am. med. Ass.* **237**, 347–351.

Yorke, J.A., Hethcote, H.W. & Nold, A. 1978 Dynamics and control of the transmission of gonorrhea. *Sex. Trans. Dis.* **5**, 51–56.

Zeigler, J. 1993 Breast feeding and HIV. *Lancet* **342**, 1437–1438.

Lecture delivered 10 March 1994

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES