Infectious Diseases and Human Population History

Throughout history the establishment of disease has been a side effect of the growth of civilization

Andrew P. Dobson and E. Robin Carper

From the plagues of biblical times to the HIV pandemic of today, infectious diseases have played an indisputably major role in human history. The continual expansion of human populations since prehistoric times has led to successive invasions of the human population by increasing numbers of different pathogens. Today many people's worries about emerging pathogens have been sharply focused by the Ebola virus outbreak in Kikwit, Zaire, and by Lyme disease and hantavirus outbreaks throughout the United States (Garrett 1994, Levins et al. 1994). In this article, we examine the infectious diseases of humans from an ecological perspective.

Understanding pathogens at the population level is as important in disease prevention and control as understanding pathogens at the microscopic or molecular level. Three ecological processes are crucial in determining the impact, persistence, and spread of pathogens and parasites: the size and spatial distribution of the host population, the movement of infected and susceptible hosts and vectors, and the nutritional status of the human host population. Although medical advances continue to reduce the impact of degenerative and self-inflicted diseases on people who can afford to pay for treatment, the control and prevention of infectious diseases is likely to be increasingly dependent on a solid understanding of the ecology of pathogen transmission and persistence.

It is likely that many of the plagues mentioned in the Old Testament arose when human populations became sufficiently aggregated to sustain epidemics of new pathogens

Parasite life-history strategies

The enormous array of pathogens that infect humans and other animals may be conveniently divided on epidemiological grounds into microparasites and macroparasites (Anderson and May 1979, May and Anderson 1979). Microparasites include the viruses, bacteria, protozoa, and fungi; they are characterized by their ability to reproduce directly within individual hosts, their small size, their relatively short duration of infection, and the production of an immune response in infected and recovered individuals. Mathematical models examining the dynamics of these pathogens divide the host population into susceptible, infected, and recovered classes. In contrast, macroparasites (the parasitic worms, ticks, and fleas) do not multiply directly within an infected individual but instead produce infective stages that usually pass out of the host before transmission to another host. Macroparasites tend to produce a limited immune response in infected hosts, are relatively long-lived, and are usually visible to the naked eye. Mathematical models of the population dynamics of macroparasites have to consider the statistical distribution of parasites within the host population.

The complexities of parasite-host population dynamics may be reduced by the derivation of expressions that describe the most important epidemiological features of a parasite's life cycle (Anderson and May 1979, 1991, Dobson 1988, May and Anderson 1979). Three parameters are important in describing the dynamics of a pathogen: the basic reproductive ratio (which determines the rate at which the pathogen is likely to spread in the population), the threshold number of hosts required for the parasite to become established, and the mean levels of infection of the parasite in the host population.

Basic reproductive ratio, R_0

The number of individuals that each infected individual infects at the

Andrew P. Dobson is an assistant professor in the Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544. E. Robin Carper is on leave of absence from Johns Hopkins University, School of Hygiene and Public Health, 615 North Wolfe St., Baltimore, MD 21205.

beginning of an epidemic is formally termed R_0, the basic reproductive ratio of the disease. The first derivation of an expression for R_0 was formulated by epidemiologist George MacDonald in his seminal studies of malaria (MacDonald 1956). This work provided the key insight that if an infection is to persist in a host population, then each infected individual must on average transmit the infection to at least one other individual. If this level of transmission does not occur, the infection is likely to progressively disappear from the population.

The basic reproductive ratio for a macroparasite is defined as the number of daughter worms (or ticks) established in a host population following the introduction of a solitary fertilized female worm (or tick). In both the microparasite and macroparasite cases, the resultant expression for R_0 usually consists of a term for the rates of parasite transmission, divided by an expression for the rate of mortality of the parasite in each stage in the life cycle. Increases in host population size or rates of transmission tend to increase R_0, whereas increases in sources of parasite mortality or decreases in transmission rate tend to reduce the spread of the pathogen through the population.

It is also possible to derive expressions for the levels of prevalence (proportion of the hosts infected) and incidence (mean parasite burden) of parasites in the host populations (Anderson and May 1991). In general, changes in the parameter values that tend to increase R_0 tend also to produce increases in the proportion of hosts infected by a microparasite, as well as increases in the prevalence and mean burden of worm macroparasites (Anderson and May 1979, 1991, Dobson 1988, May and Anderson 1979). In particular, increases in the size of the host population usually lead to increases in the prevalence and incidence of the disease in the host population.

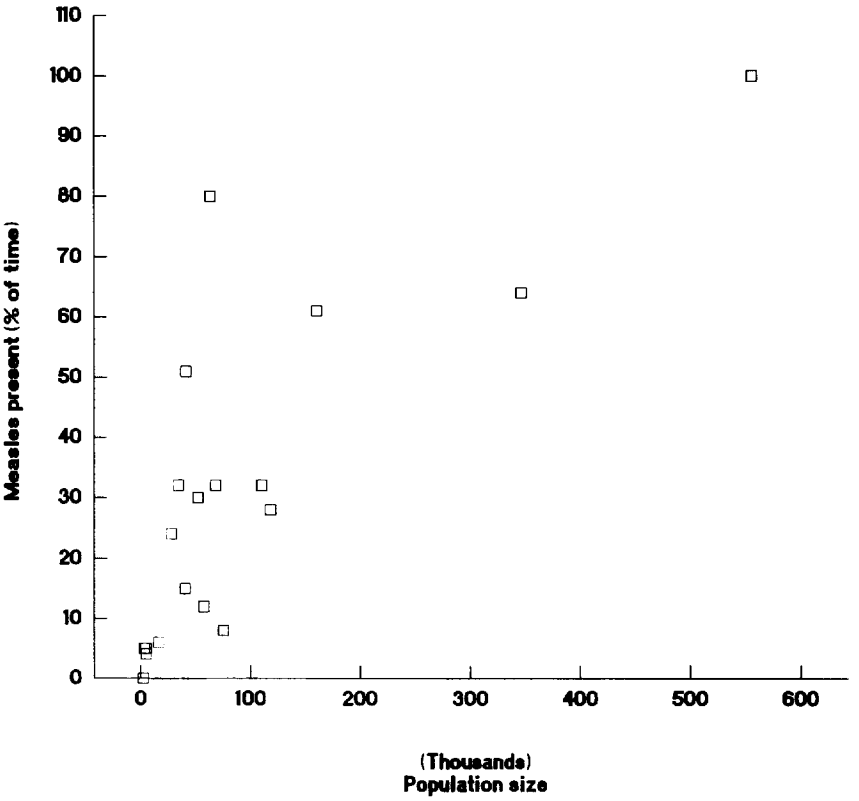

Figure 1. The relationship between human population size on oceanic islands and the percentage of months when measles was recorded on the islands (after Black 1966).

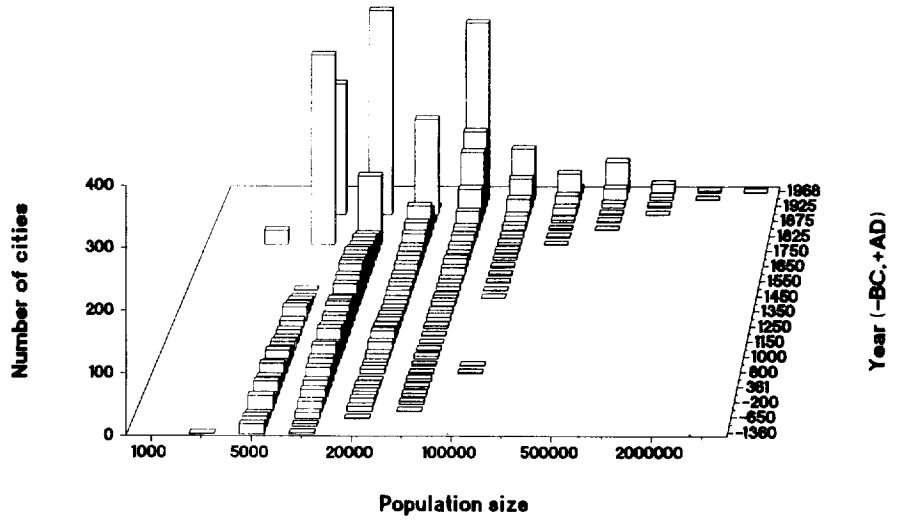

Figure 2. Distribution of population size of the world's cities from 1360 B.C. until A.D. 1968. The data are divided into frequency classes that increase on a logarithmic scale (e.g., 1000, 2000, 5000, 10,000). Unfortunately, it is logistically impossible to obtain information on historic aggregations of less than 2000–5000 people, or more recent data for towns of less than 20,000 people (data from Chandler and Fox 1974).

Thresholds for establishment, H_T

Epidemiological theory suggests that host population density is critical in determining whether a pathogen can become established and remain endemic in a population (Anderson and May 1991), a notion that was articulated as long ago as 1927 (Kermack and McKendrick 1927, 1932). An expression for the threshold for disease establishment, H_T, may be obtained by rearranging the expression for R_0 to find the population density where R_0 equals unity. The size of the host population needed to continuously sustain new

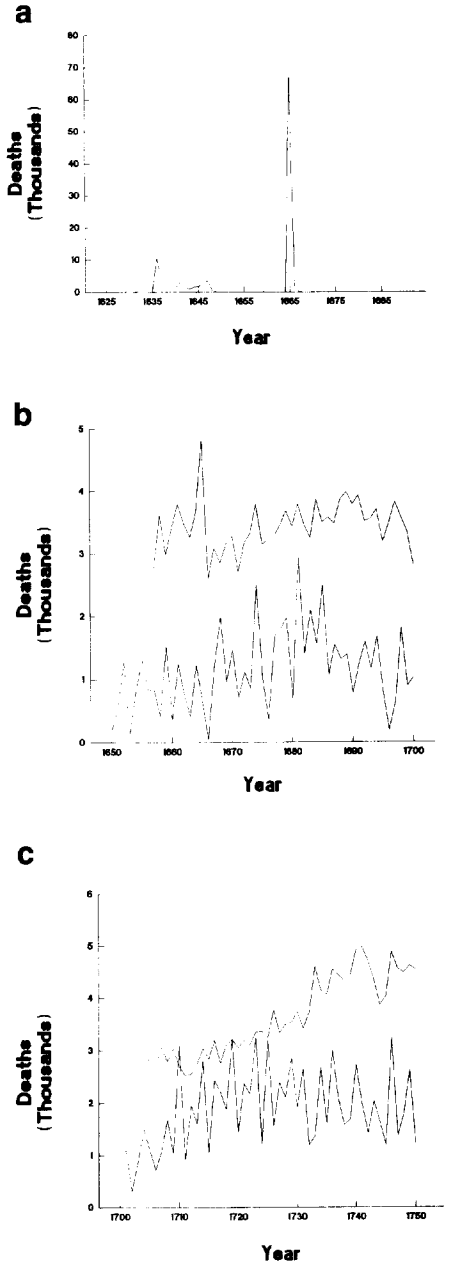

Figure 3. Recorded deaths from different diseases in the city of London 1625–1750. (a) During the period 1625–1700, plague had a dramatic impact on the population. (b) and (c) During the period 1650–1750, there were cyclic outbreaks of smallpox (1000–3000 cases per year; lower lines) and a steady rise in deaths from consumption (tuberculosis [TB]; upper lines). Redrawn from data in Appleby 1975.

infections usually varies inversely with the transmission efficiency of the pathogen but directly with its virulence (case mortality rate). Mathematical expressions may be derived for micro- and macroparasites with either simple or complex life cycles (Diekmann et al. 1990, Dietz 1993). The resultant expressions suggest that changes in the parameters that tend to increase R_0 tend to reduce H_T, and vice versa. Therefore, more virulent species require larger populations to sustain them, while reductions in the mortality rate of transmission stages may allow parasites to maintain infections in populations previously too small to sustain them. Some modern estimates suggest that a population of approximately 500,000 people is needed to produce annually the 7000 susceptible individuals needed to sustain an endemic infection of measles (Figure 1).

An important parallel can be drawn with the concept of population viability used in conservation biology to determine whether a population of an endangered species is likely to persist and not go extinct; essentially 500,000 people can be considered as the minimal viable host population required to sustain a continuous infection of measles. Obtaining similar empirical estimates for other infectious diseases would yield crucial insights into the role that urban aggregations play in maintaining infectious diseases endemic in human populations. Preliminary analysis of historical data for human diseases in Australia suggests that populations of approximately 200,000 people are required to sustain pertussis (also known as whooping cough) and scarlet fever, while only 50,000 people are required to sustain diphtheria (also known as croup; Cumpston 1927).

The most parsimonious way of examining when infectious diseases became established in human populations is to examine human epidemiological history from the perspective of the size and spatial distribution of human populations as they began to expand in numbers and aggregate into the first villages, towns, and cities in different parts of the world at different historical times.

Parasites and pathogens of early human populations

The nomadic hunter-gatherer bands that were typical of human social structure for most of humankind's 2-million-year history probably supported a parasite fauna closely resembling that of the higher apes, from whom humans inherited their original collection of pathogens. Parasites with high transmission rates that induced little or no immunity were probably the only pathogens able to establish in the groups of around 50–100 individuals that characterized these early societies. Pathogens such as hepatitis B, herpes, Epstein-Barr virus, and cytomegalovirus are still found in primitive tribes in the few remaining isolated parts of the world (Black 1975). Direct life cycle macroparasites (those that do not require vectors for transmission), such as pinworms, *Ascaris*, lice, and ticks, were probably also common in hunter-gatherer societies, as were sexually transmitted diseases, which can be sustained in low-density host populations. Similarly, vector-transmitted pathogens, such as malaria and yellow fever, may have become adapted to humans at this early stage in human history. The long latent periods associated with the development of *Plasmodium vivax*, probably the oldest species of human malaria, suggest a mechanism for parasite survival under low transmission conditions.

The more sedentary habits of the early agriculturists probably increased the incidence of direct life cycle macroparasites, such as the roundworm *Ascaris*, mainly because of the increasingly successful transmission of the long-lived free-living stages, which increased in numbers around more permanent dwellings. Water supplies became contaminated with toxic bacteria and protozoa as humans became more sedentary. Archeological studies on human communities entering the transition between hunter-gathering and primitive pastoralism provide evidence suggesting that a more sedentary existence tended to lead to increased child mortality and high levels of disease. A significant characteristic of this malnutrition was a reduced diversity in diet as humans switched from a mixture of meat, grains, and fruit to a diet dominated by grains (Cohen and Armelagos 1984). This change in diet was not entirely detrimental, because stor-

ing wheat, barley, and millet in earthenware pots may have created ideal conditions for *Streptomycetes* to develop. The tetracycline-producing *Streptomycetes* are effective against bacteria, rickettsiae, spirochetes, and some viruses (Bassett et al. 1980). Thus some ancient populations would have inadvertently dosed themselves with the first antibiotics during roughly the same historical period that they began to acquire their first significant bacterial and protozoan pathogens.

The first cities

The first significant aggregations of humans into sedentary agricultural settlements occurred around 5500 B.C. at Khuzistan in the valleys of the Tigris and Euphrates in what is now Iran (Chandler and Fox 1974). Irrigation allowed these first cities to grow; nevertheless, it took 2000 years for these early villages to develop into the city of Uruk (which had around 50,000 inhabitants at the peak of its influence) and the surrounding towns of Ur, Kish, Lagash, and Umma (each of which had between 10,000 and 20,000 inhabitants). The early settlements that appeared in the Indus River valley in India at this time and in China, Egypt, and Mesoamerica consisted more of towns and villages than cities. None of these conurbations were large enough to continuously sustain any of today's common childhood diseases.

Only when communication between small groups of neighboring towns began to be established is it likely that human populations became large enough to sustain direct life cycle bacterial and viral infections. It is in these first cities that the now common diseases of humans started to appear. Many of the first pathogens to infect humans evolved from diseases of domestic animals. Measles, for example, is closely related to two other morbilliviruses—canine distemper and rinderpest (a disease of cattle)—whereas smallpox probably evolved from cowpox. India is probably the original home of smallpox; traditions concerning this disease have long existed amongst the Brahmans. Rubella, typhoid, and dysentery also

Figure 4. Frequency distribution of family sizes in Great Britain 1870–1925 (after Wrigley 1969). Decreases in family size may have lead to changes in the average age at infection and the case mortality rate.

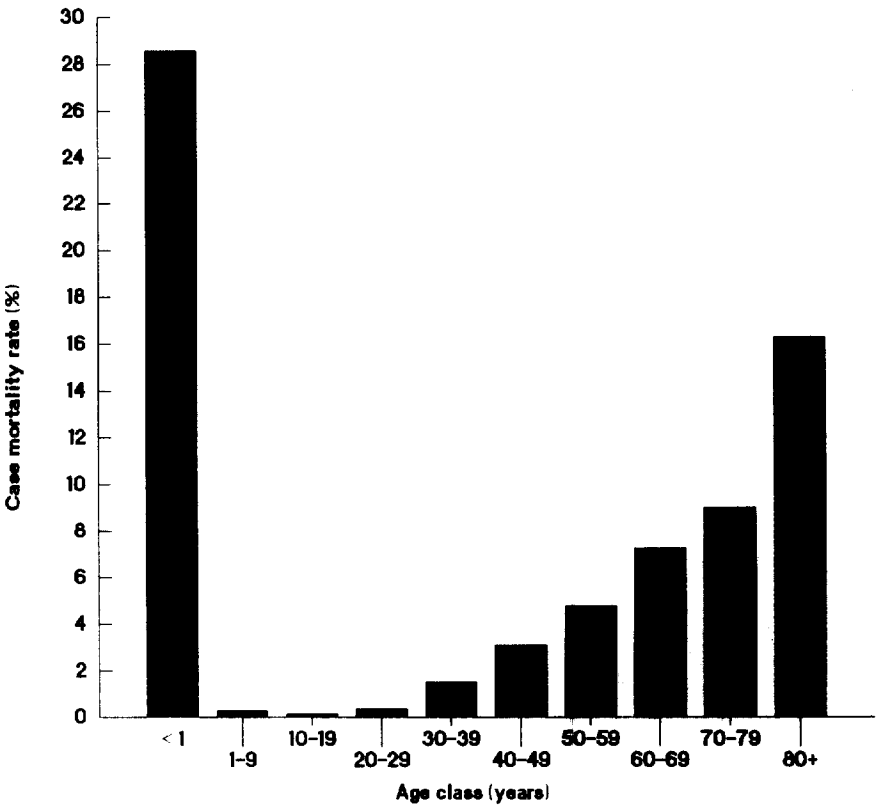

Figure 5. The age-dependent mortality rate due to an outbreak of measles in the population of the Faeroe Islands in 1846 (after Panum 1940).

arose during this period. Although written descriptions of many early disease outbreaks are almost impossible to decipher, rabies is known from writings in Babylonian tombs from approximately the twenty-third century B.C., while smallpox and tuberculosis were clearly described in Chinese writings from approximately A.D. 1000.

Hippocrates (460–377 B.C.) was probably the first person to record diseases with enough precision for them to be identified today as malaria, mumps, diphtheria, tuberculosis, and perhaps influenza. The Greeks were definitely aware of pulmonary tuberculosis—an epidemic of tuberculosis seems to have raged during Hippocrates' lifetime. Hippocrates also described many other recognizable infections, including malaria, so it is apparent that these diseases were present in Hellenic

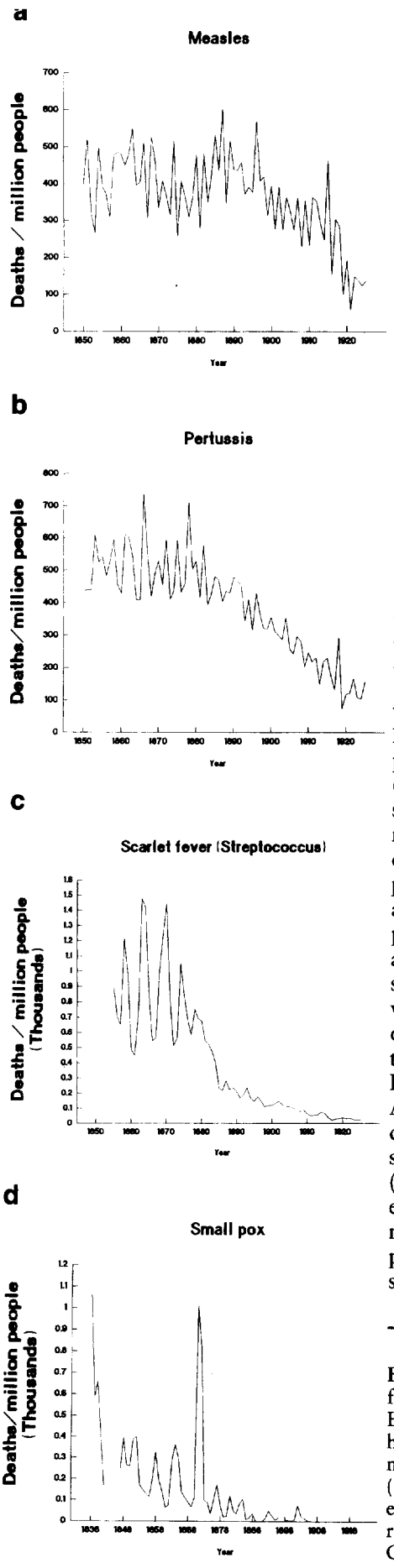

Figure 6. Decline in mortality rates from four common microparasitic diseases in England and Wales during the second half of the nineteenth century: (a) measles, (b) pertussis, (c) scarlet fever (Streptococcus), and (d) smallpox. In each case the figures given are death rate per million people infected (after Cumpston 1927).

times. Interestingly, none of Hippocrates's records indicate the presence of smallpox, measles, or bubonic plague in ancient Greece, which suggests that the size of ancient Greek and Egyptian cities may not have been large enough to continuously sustain measles and smallpox infections, and these pathogens may have died out.

It is likely that many of the plagues mentioned in the Old Testament arose when human populations became sufficiently aggregated to sustain epidemics of new pathogens. Because human populations had not previously been exposed to pathogens such as smallpox, rubella, and measles, levels of immunological resistance were low and mortality rates were high. After several visitations by a specific pathogen, levels of genetic resistance and herd immunity start to increase, leading to reduced pathogenicity and a more regular pattern of prevalence.

There is circumstantial evidence that the plague of Athens (430–429 B.C.), which killed 25% of the city's population, may have been measles (McNeill 1976); contemporary descriptions of the pathology closely resemble those from the first epidemic of measles in Hawaii. The population of Athens was only around 155,000 at the time of the plague; therefore, if the plague was an early outbreak of measles, the small size of the human population would have caused the disease to die out once the initial pool of susceptible individuals had been exhausted. In a smallpox epidemic in Rome in A.D. 165–180, 25% of the populace died, while 5000 a day died in a second epidemic in A.D. 251–266 (McNeill 1976). These and other epidemics may have played a significant role in strengthening the position of the growing Christian sect, whose members believed in assisting the sick, whom the rest of society sought to avoid. These caregivers, through repeated low-level exposures to the virus, may have built up their immunity to the pathogen, thus appearing protected by virtue of their religious convictions. Perhaps this so-called protection gave the Christians an advantage during the big Roman epidemics and convinced many nonbelievers to join their ranks.

If one examines the distribution of city sizes over the last 3500 years, it becomes clear that cities capable of maintaining an endemic measles infection were continuously present only from around the late seventeenth century (Figure 2). Cities capable of continuously supporting diphtheria (circa 50,000) have been present only since the early Christian era, while cities capable of supporting pertussis and streptococcus (scarlet fever) have been present only since the late Middle Ages (twelfth and thirteenth centuries). Although some large cities were present in late Roman times, these cities declined in the Dark Ages. Most of the cities in which today's diseases evolved were the size of large towns or the suburban districts of today's megacities. The data in Figure 2 suggest that sequential outbreaks of measles in different towns may have been an important historical mechanism for ensuring the persistence of pathogens in human populations before towns and cities were large enough to continuously support infections.

In the case of measles, persistence would have had to occur on time scales of the order of 10 to 18 centuries before cities were large enough to continuously sustain infections. Consequently, a curious distortion in the present understanding of the population biology of measles has been created. Most studies have focused on the dynamics of measles in large cities in the United States and Europe in the mid-twentieth century (Anderson et al. 1984, Bolker and Grenfell 1993, Grenfell 1992, Olsen and Schaffer 1990); yet for 90% of its history, measles has persisted by moving between ephemeral populations of susceptibles in largely rural populations (Cliff and Haggett 1980, 1984, 1988).

May and Anderson (1990) pro-

pose a mathematical model to explain this phenomenon as an explanation for the early period of the HIV/AIDS epidemic. The model suggests that it may have been possible for HIV to persist for periods of up to 100 years as a sequence of outbreaks that flared and died out in small villages. Providing a new outbreak was initiated in another village before the pathogen had completely died out in villages where outbreaks were occurring, then the disease could have persisted for a long time in populations of villages that, individually, would be too small to sustain a continuous outbreak. It would be intriguing to know if similar spatial heterogeneities allowed measles to persist for much of human history.

Diseases in the New World

The population of the New World probably numbered around 100 million before Columbus's arrival in the fourteenth century (McNeill 1976). In terms of disease status it resembled the Old World around the time of the birth of Christianity. Intestinal worms and protozoan infections prevailed, and traces of these parasites are often found in archaeological studies of bodies from pre-Colombian burial sites (McNeill 1976). The apparent lack of other pathogens in the New World may be due to differences in animal husbandry practices (Clutton-Brock 1987). In much of South America the traditional herd animals were llamas and alpacas. Because these animals live in small groups high in the mountains, there was little contact with humans and smaller populations of hosts within which diseases could become established. Important exceptions to this case were the protozoan infections that give rise to trypanosomiasis and its close relative leishmaniasis. These pathogens are endemic in many species of small mammals, and *Trypanosoma cruzi* is particularly prevalent in the populations of guinea pigs traditionally kept in large numbers by New World farmers as a source of meat. Triatomid bugs transmitted the pathogen between guinea pigs, humans, and reservoir hosts such as possums.

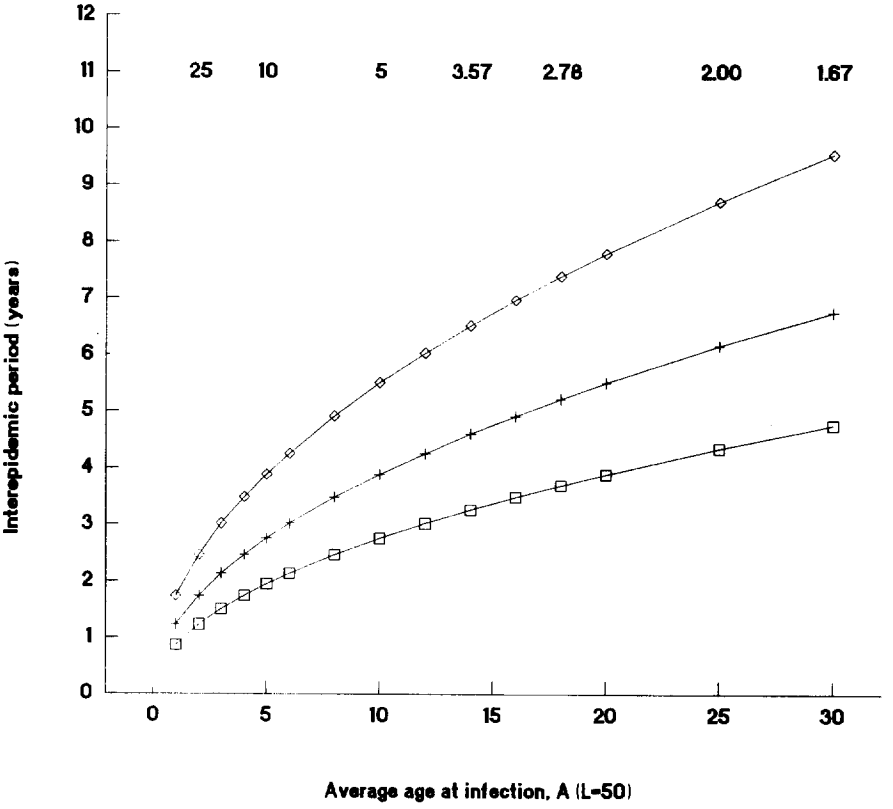

Figure 7. The relationship between average age of infection, A, and interepidemic period for three different periods for which the host is infected: one week (bottom line, □), two weeks (middle line, +) and four weeks (top line, ◊). The period of infection combines the incubation (or latent) and infectious periods of the pathogen. The values of R_0 corresponding to this particular age of infection are given above the main part of the graph; this calculation assumes life expectancy (L) = 50 years. Typical periods of infection for some common childhood diseases are: measles (12–16 days), mumps (16–24 days), pertussis (28–33 days), rubella (18–26 days), smallpox (10–15 days), and scarlet fever (15–23 days). These periods may be longer in malnourished or immunocompromised hosts.

Disease again worked in the favor of the Christians during the European invasion of the New World during the fifteenth and sixteenth centuries—an event that had immense epidemiological repercussions for the indigenous people of the Americas (McNeill 1976). The coming of the Europeans provides perhaps the best illustration of the effects of population movement on pathogen transmission. The introduction of smallpox, measles, and typhus to South American and Central American human populations, who had no natural resistance or immunity to these diseases, led to apalling numbers of deaths. Largely as a result of successive epidemics of these diseases, the population of Mexico fell from 20 million to approximately 3 million in the 50 years from 1518 until 1568 and then to 1.6 million in the next 50 years. The successful colonization of these continents probably owes considerably more to the pathogens that Europeans brought with them than to any of their more traditional weapons.

While malaria was prevalent in Europe at the time of the colonization of the New World, historical data suggest that the disease was absent in the Americas before the arrival of the Europeans. Moreover, the importation of slaves from highly endemic areas of Africa fueled epidemics that spread in waves across the continents, ravaging both the indigenous and colonist populations. In North America, the spread and

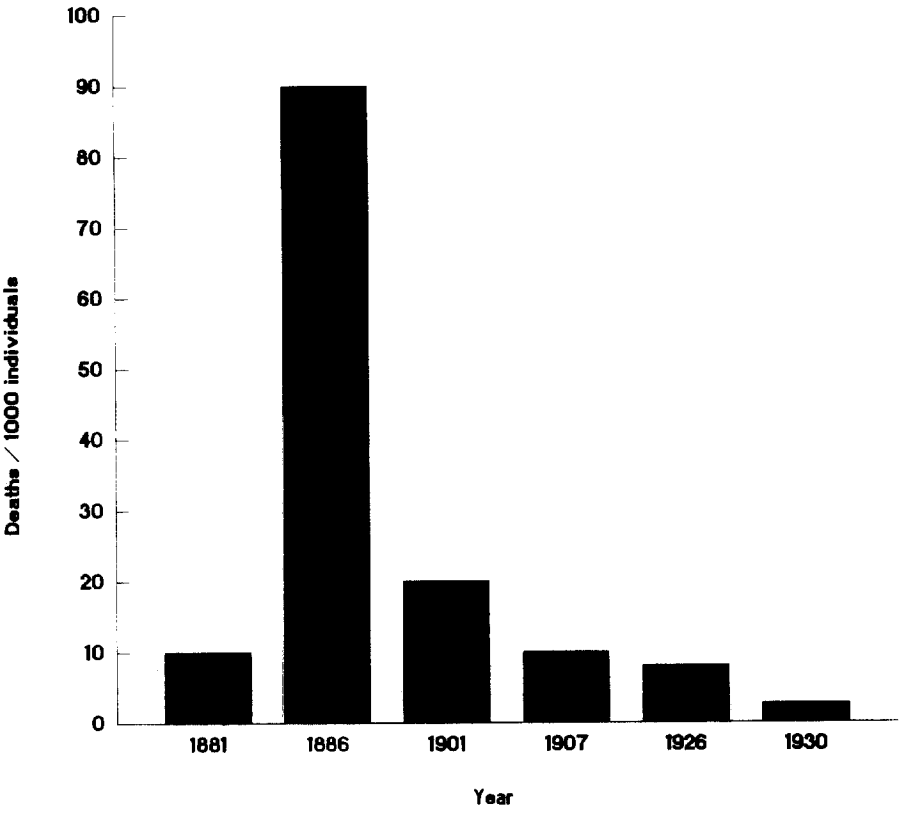

Figure 8. The mortality rate due to tuberculosis in three generations of Canadian plains American Indians after their forced adoption of a sedentary existence (after Ferguson 1933).

persistence of malaria were closely associated with pioneer life. Although epidemic levels of malaria had subsided in the northeast colonies by the mid-1700s, later settlers in the upper Mississippi Valley suffered high levels of mortality and morbidity. So bad was the situation that one writer declared that "it is unlikely that any amount of toil... will make this land habitable." The prevalence of malaria finally began to decrease in the latter half of the nineteenth century, probably due to improved housing, nutrition, and agricultural practices as the frontier became settled. In the southern states, however, the disease remained deeply entrenched as late as the 1930s and 1940s, when 4 million cases per year were being reported. This high endemicity seeded a new round of malaria epidemics in the northeast when Northern soldiers returned from the South at the end of the Civil War—epidemics that did not subside until the early 1900s.

Meanwhile, back in Europe

The first influenza epidemic in Europe occurred from 1556 to 1560 (McNeill 1976), at approximately the same time as a similar epidemic in Japan (1556). This directly transmitted microparasite has a short latent period and is highly contagious, requiring large populations in order to establish itself. Its first recorded appearance in the human population seems to have caused around 20% mortality (McNeill 1976). The great influenza pandemic that followed the World War I in 1918–1919 led to the deaths of 20 million people—more than had died in the war itself (McNeill 1976). This high mortality may have been associated with the low levels of nutrition in the human population following the wartime rationing and the high incidence of secondary infections that followed the war in Europe.

It is only from around 1650 that population statistics are reliable and that real estimates can be made of the birth and survival rates of human populations. These and similar data sets suggest that disease epidemics were a common consequence of increasing urbanization. The data for the city of London for the time period 1625–1750 illustrate two major outbreaks of plague (Figure 3; Appleby 1975). Although this disease died out after thatched roofs were replaced with slate roofs in the aftermath of the Great Fire of London, other pathogens became established in this period. So-called fever deaths may represent mortality due to influenza, while deaths due to consumption reflect the increasing incidence of tuberculosis. Smallpox remained endemic in the population at this time and caused regular epidemic outbreaks.

Spatial aggregation and average age at infection

Although diseases of childhood were well established in towns in the seventeenth and eighteenth centuries, they were still transmitted only intermittently in the smaller populations of rural areas. Before 1800, less than 2% of the European population lived in cities of 100,000 or more (McNeill 1976). Many contagious diseases were spread when boys and men joined the army or when children went to school. Even by the time of Napoleon (1812), it was noted that skinny and ill-fed urban recruits survived much longer than did large, muscular, and well-fed recruits from the countryside, who were repeatedly ill (McNeill 1976). This disparity in health was probably due to a reduced exposure to pathogens and hence a lower level of immunity in the smaller populations of isolated towns and villages. The earlier figure for the size of the world's cities illustrates the aggregation of Europe's growing human population into an increasing number of larger towns and cities (Figure 2). This large-scale change in spatial distribution was matched by a reduction in spatial aggregation at a lower scale.

Many demographers argue that the so-called demographic transition to smaller family sizes (Figure 4) is a direct consequence of increased health and hence increased

survival. Yet it should be noted as well that reductions in family size also lead to direct improvements in the health and welfare of women and perhaps to significant reductions in the rates of disease mortality within families. Studies of mortality from measles in families of different sizes indicate that mortality rates among children are higher in larger families, even when the effects of increased malnutrition in many larger families are controlled (Aaby et al. 1988, Garenne and Aaby 1990). In general, as more children in a family succumb to an infection, its pathology becomes worse. This outcome may be because susceptible children in larger families receive higher disease inocula from their infectious siblings, or it may reflect rapid selection that allows the virus to exploit the common genetic background of children in the same family. Probably, both of these effects operate. If they do, then the net impact of pathogens on the host population may have declined as human societies have shifted toward smaller families.

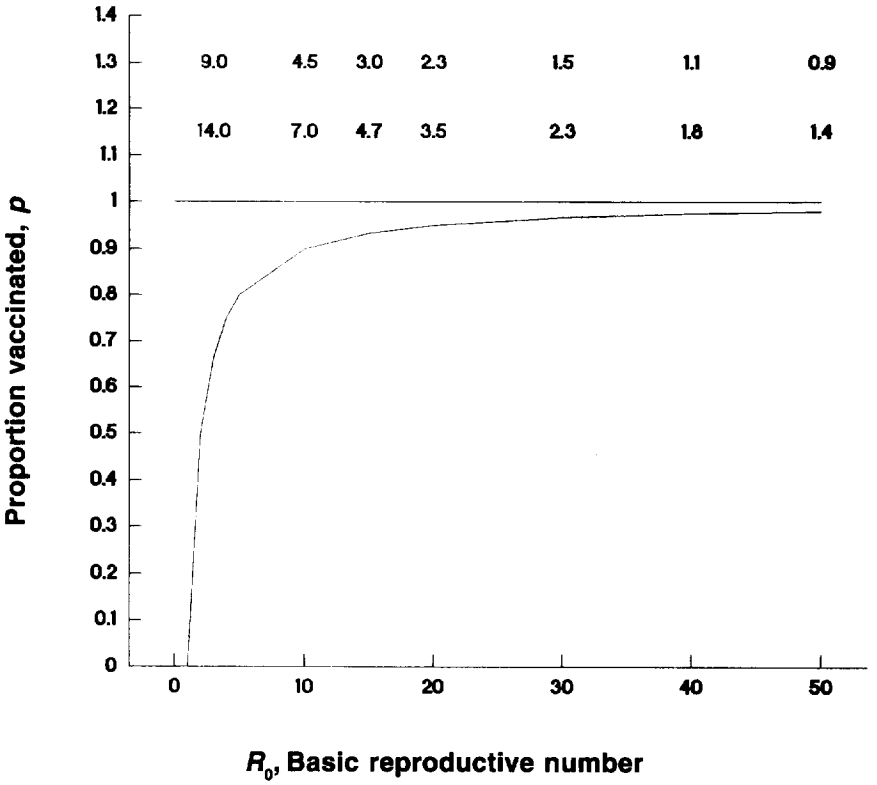

Figure 9. The relationship between the proportion of a community that must be immunized to eradicate an infection, p, and the basic reproductive rate, R_0. The figure also illustrates the average of infection, A, before vaccination for different values of R_0; this relationship is shown for L = 70 (lower figures) and L = 45 (upper figures). The infection is eradicated for values of p above the line.

Age-dependent mortality

The net impact of a pathogen on a population is influenced by the interaction between the average age of infection and the age-dependent mortality rate of the disease. If pathogen-induced mortality increases significantly with age, then the later ages of infection that characterize rural communities may well cause higher mortality in lower density rural populations. In contrast, mortality may be lower in urban populations where people are likely to be exposed to infection at earlier ages. Anecdotal evidence of this effect is provided by information on recruits to Napoleon's army.

A remarkable study of an outbreak of measles in an isolated island off the northern coast of Scotland (Panum 1940) offers quantitative evidence in support of age-dependent mortality. In 1846, when the human population of the Faeroe Islands numbered 7782 people, measles appeared and infected nearly everyone. A previous epidemic in 1781 had also infected almost the whole population, but the disease had disappeared completely, presumably due to a lack of susceptible persons. Measles was reintroduced to the population in 1846 by an infected carpenter who had been exposed to the disease just before leaving Copenhagen, Denmark. He became sick in April, soon after his arrival on the islands, and by October the disease had spread throughout the community. The 92 people who survived the 1781 epidemic escaped infection in 1846. Others escaped by rigid quarantine and isolation, but 95% of the rest of the population, approximately 6000 people out of a possible 6682, eventually contracted measles. The infection then disappeared from the islands, again presumably due to a lack of susceptible persons. The mortality rate of people of different ages, carefully recorded by Panum (1940), reveals an interesting age-dependent case fatality rate (Figure 5). Children younger than one year old tended to die from measles. The mortality rate then declined to approximately 0.3% in age classes up to age 30, after which it increased first to 1.5%, and then to 8%–10% in people older than 60. Thus the average age of infection is likely to be important in determining the impact a disease such as measles has on a population. If this relationship between mortality and age occurs for other infectious diseases, then infrequent outbreaks of disease in isolated and semi-isolated rural and island populations would lead to higher levels of mortality than would occur in cities, where the majority of individuals are exposed to common, so-called childhood diseases when still young.

Vaccination and the control of human disease

The migration of people into the industrial cities in Europe throughout the nineteenth century probably increased the rates of transmission of directly transmitted diseases. Be-

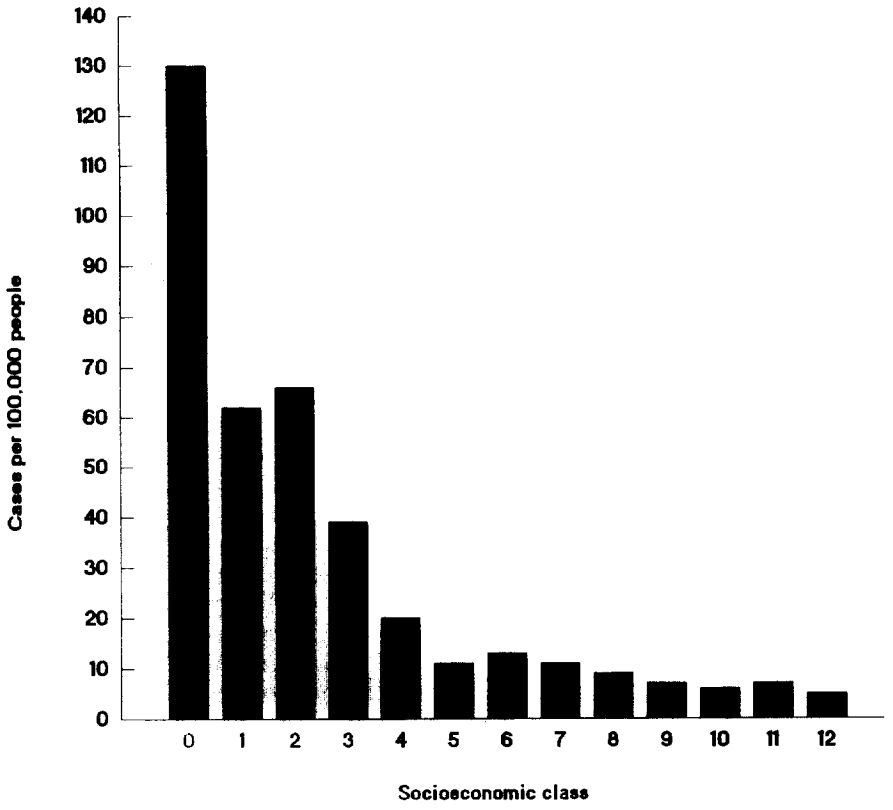

Figure 10. The incidence of tuberculosis in different socioeconomic sectors of the population in New York State in 1976 (Hinman et al. 1976).

fore 1900 it is unlikely that populations in cities were self-sustaining, because before this date urban populations were maintained only by immigration from the surrounding countryside, where 80%–90% of the population still lived. The mortality rates caused by common childhood pathogens began to decline during this period, particularly when vaccines (for smallpox) or treatments (for scarlet fever) became available (Figure 6). Notice in Figure 6 that several of the diseases exhibit a pronounced tendency to produce regular epidemic cycles. The period of these cycles can be roughly estimated using a simple formula (period = $2\pi(AD)^{1/2}$, where A is average age of infection and D is the duration of time for which an individual is infected; Anderson and May 1982). This relationship suggests that the interepidemic period is likely to decline as the average age at infection declines (Figure 7). Several of the data sets (measles and smallpox) illustrated in Figure 6 show an apparent shortening of interepidemic period as we move from the mid-nineteenth to the early twentieth century. This shorter period may reflect the younger age at first infection (and hence increasing R_0) as the human population became larger and more aggregated into urban areas. It may also reflect the effects of better nutrition in reducing the time that people were infectious.

Selection for increased disease resistance

There are almost no data available that can be used to examine selection for resistance in humans to the common childhood diseases such as measles, mumps, and rubella. Similarly, there are no studies that allow scientists to accurately examine reductions in the virulence of these pathogens (Ewald 1983). There is one historical study of tuberculosis (Ferguson 1933, 1934) that allows some quantification of the selection pressure that this pathogen may have placed on previously naive human populations. In a study of tuberculosis in American Indians of the Canadian plains, Ferguson (1933, 1934) records huge increases in mortality due to tuberculosis once the American Indians were removed from their normal nomadic existence and settled onto reservations (Figure 8). The mortality rate from tuberculosis, which had not previously been recorded for these tribes, rapidly increased to 90–140 cases per 1000 people within a few years of the reservations' founding. From 1882 through the 1930s, the mortality rate decreased from 90 cases per 1000 people to 20 cases, and then to 8 cases, per 1000 people (Ferguson 1933, 1934). In some years the mortality rate was 20 times the rate for the surrounding white population. A number of different factors may have caused this high mortality rate—inadequate diet after the disappearance of buffalo, poor housing, overcrowding, and general spiritual demoralization. Yet many European immigrants exposed to conditions as bad as those encountered by the American Indians did not have mortality figures of this magnitude. Careful analysis of the family trees of the Indians by Ferguson revealed that during the peak of the epidemic in the 1880s, although some families suffered high mortality, others suffered much lower mortality. Many of the surviving American Indian families were descendants of the families with lower mortality. This finding suggests that some selection for reduced susceptibility to tuberculosis operated.

Unfortunately, it may never be possible to determine the relative importance for declining disease mortality rates of changing age at first infection, smaller family size (and hence smaller inocula), better nutrition, and selection for resistance. Obviously improvements in nutrition and personal hygiene have been important, but changes in the spatial distribution of humans and selection for reduced virulence in the pathogens may also have played important roles in changing average age of infection and case mortality rate. Furthermore, most of the reduction in pathogen-induced mortality rates occurred before the full development of preventive medicine, suggesting that much of the reduction of the impact of disease may

have been due to the development of agricultural practices that made food cheaper and more plentiful and the development of more efficient means of sewage and waste disposal.

Vaccination

The development of vaccines for many directly transmitted microparasitic diseases of humans contributed to the final decline in juvenile mortality rates in developed countries during the present century. Nevertheless, these diseases are still major causes of death in the developing world. Mathematical population models can again help to explain why vaccination has been successful against some pathogens and why it may be less successful against others. Essentially, a successful vaccination scheme should reduce the size of the pool of susceptibles to below the threshold needed for the pathogen to sustain itself. More specifically, the proportion of the host population that needs to be vaccinated, p, is given by $p=1-1/R_0$ (Figure 9). It is thus possible to eradicate a disease by vaccinating only a proportion of a population rather than the whole population. However, this possibility also implies that although vaccination may be useful as a control against diseases such as rubella and measles, it is unlikely to be effective against diseases such as malaria, where a large proportion of the infant population is infected by the age of six months.

One of the major additional effects of vaccination is the reduction of infected individuals in the population due to the reduced number of contacts that infected individuals have with susceptible individuals. This effect is called herd immunity (Fine 1993, Fox 1983, Fox et al. 1971). As the percentage of the population that is immunized increases, there are linear decreases in the total incidence of the disease in the population and increasingly rapid decreases amongst the proportion of individuals not vaccinated.

Arita et al. (1986) assembled data on population densities for smallpox vaccination coverage in African and Asian countries during the late 1960s and early 1970s. These data indicate that smallpox disappeared early from countries in which the density of susceptible (unvaccinated) individuals fell below ten persons per square kilometer. In populations with a population density of less than 50 persons per square kilometer, this density roughly corresponds to 80% coverage. Infections of smallpox persisted in more densely populated regions, in particular Nigeria, with 58 persons infected per square kilometer; Pakistan, with 83 persons per square kilometer; India, with 175 persons per square kilometer; and Bangladesh, with 502 persons per square kilometer (Arita et al. 1986). To eradicate smallpox in these areas would have required coverage of the order of 98% to reduce the number of susceptible individuals to less than ten persons per square kilometer; such coverage was impractical. By 1970 vaccination policy changed to active case detection, contact tracing, and the breaking of individual chains of transmission. It was this policy change that eventually led to the worldwide eradication of smallpox (Arita et al. 1986, Fine 1993).

Infectious diseases today

Despite widespread advances in medical science, infectious diseases continue to have devastating consequences for human populations in many parts of the world. Although the incidence of measles in the United States has fallen by approximately 99% since the introduction of vaccination in 1963, annual resurgences continue even though approximately 98% of children in the United States have been vaccinated by the age of school entry. Analysis of surveillance data suggests that transmission has been continuous in several large urban populations, in particular those with large, poor, inner-city populations—such as New York, Los Angeles, and Chicago. Although transmission is only sporadic throughout the rest of the country, inner-city areas present an extremely difficult problem for public health providers because the social conditions for high vaccine coverage are least conducive in the areas where the highest uptake is required. As urbanization trends continue globally, disease is likely to continue to be maintained in cities, presenting a constant threat to the decreasing proportion of the human population living in rural areas.

In 1993 alone, tuberculosis killed 2.7 million people and infected another 8.1 million worldwide (Snider 1994). An estimated one-third of the world's population, or 1.7 billion people, is infected but has not yet developed the disease. Furthermore, the mortality rates of infected individuals are sharply related to socioeconomic status (Figure 10).

The present tuberculosis epidemic is expected to grow worse, especially in the developing world. This epidemic is partly driven by the increasing urban population, by the evolution of multidrug-resistant strains, and by the emergence of HIV/AIDS, which compromises the immune system of human hosts and makes them more susceptible to infectious diseases such as tuberculosis. In the United States and Europe, prisons and homeless populations now act as reservoirs for the resurgence of tuberculosis. These underserved subpopulations are now above the threshold sizes for maintaining diseases in areas where they had previously been significantly reduced.

While medical and sanitation technology have certainly helped minimize the impact of infectious diseases in the last two centuries, humans can only really claim victory over one pathogen—the smallpox virus. It seems inevitable that other diseases (e.g., measles and typhus) that may now be simmering quietly in scattered populations will reemerge as the susceptible subpopulations of major cities increase to threshold sizes.

The epidemiological and nutritional situation of many developing countries is similar in some ways to that of Europe in the seventeenth and eighteenth centuries (Livi-Bacci 1991). Although smallpox has been successfully eradicated, malaria, measles, rubella, and other childhood diseases are still major causes of death. Malaria, in particular, is showing a major resurgence. Following the widespread use of drugs and insecticides against the pathogen and its mosquito vector in the

mid-twentieth century, genetic resistance has evolved to these compounds, and malaria is returning rapidly to areas that had been declared free of the infection. The macroparasitic pathogens such as hookworm, ascaris, and schistosomiasis continue to debilitate many millions of people. The effects of all diseases in the developing nations are made worse by poor nutrition. The United Nations Children's Fund estimates that 50%–75% of all recorded deaths of infants and young children can be attributed to a combination of malnutrition and infection (UNICEF 1995).

As the only developed nation other than South Africa without a national health plan, the United States may be the first developed country to have to contend with reemerging diseases. For example, although malaria had been successfully controlled in most of the country by the mid-nineteenth century, large cities such as New York still supported endemic transmission of *Plasmodium falciparum* as late as the early 1940s. In an interesting parallel to current HIV epidemiology, falciparum malaria was maintained in the mid-twentieth century entirely among intravenous drug users (Blower et al. 1991, Most 1940). Today, with access to quality health care dependent on socioeconomic status, the rapid growth of susceptible subpopulations (e.g., homeless and prison populations) in large cities is not surprising. These subpopulations are susceptible not because they have no history of exposure to a particular pathogen (the majority of people in developed countries have not been exposed to many infectious pathogens) but because they are underserved by or inaccessible to the medical community (Berkelman et al. 1994).

No challenge to medical science more sharply illustrates the importance of developing an ecological understanding of infectious disease than the current pandemic of HIV/AIDS. Much of the progress in understanding HIV/AIDS comes from studies at the population level (Anderson and May 1991) and from extensions of these mathematical techniques to develop an understanding of the population dynamics of interactions between components of the immune system and the HIV virus in individual infected hosts (Nowak et al. 1991). These quantitative scientific insights into the biology and epidemiology of HIV have been obtained using extensions of models originally developed to study the ecology of infectious diseases in human and other animal populations (Anderson and May 1991). It is essential that policy for disease-control strategies in the twenty-first century build upon this understanding of diseases at the population level as well as at the molecular and cellular level.

Acknowledgments

We would like to thank Leslie Real for comments on an earlier draft and for inviting us to contribute to this special issue. We would also like to thank Bryan Grenfell, Bob May, Annarie Lyles, Ruth Berkelmen, and the students of Freshman Seminar 102 "AIDS, Anthrax and Worms: Disease in Human History" for many insightful and stimulating discussions. We also thank David Goodman and Louise Schaeffer of the Biology Library at Princeton University for never failing to find the obscure books we request. Andrew P. Dobson's work is partly sponsored by the Geraldine R. Dodge Foundation.

References cited

Aaby P, Bukh J, Dasilva MC, Lisse IM. 1988. Decline in measles mortality—nutrition, age at infection, or exposure. British Medical Journal 296: 1225–1228.

Anderson RM, May RM. 1979. Population biology of infectious diseases: Part I. Nature 280: 361–367.

_____. 1982. Directly transmitted infectious diseases: control by vaccination. Science 215: 1053–1060.

_____. 1991. Infectious diseases of humans: dynamics and control. Oxford (UK): Oxford University Press.

Anderson RM, Grenfell BT, May RM. 1984. Oscillatory fluctuations in the incidence of infectious disease and the impact of vaccination: time series analysis. Journal of Hygiene 93: 587–608.

Appleby AB. 1975. Nutrition and disease: the case of London, 1550–1750. Journal of Interdisciplinary History 6: 1–22.

Arita I, Wickett J, Fenner F. 1986. Impact of population density on immunization programmes. Journal of Hygiene 96: 459–466.

Bassett EJ, Keith MS, Armelagos GJ, Martin DL, Villanueva AR. 1980. Tetracycline-labeled human bone from ancient Sudanese Nubia (A.D. 350). Science 209: 1532–1534.

Berkelman RL, Bryan RT, Osterholm MT, LeDuc JW, Hughes JM. 1994. Infectious disease surveillance: a crumbling foundation. Science 264: 368–370.

Black FL. 1966. Measles endemicity in insular populations: critical community size and its evolutionary implication. Journal of Theoretical Biology 11: 207–211.

_____. 1975. Infectious diseases in primitive societies. Science 187: 515–518.

Blower SM, Hartel D, Dowlatabadi H, Anderson RM, May RM. 1991. Drugs, sex and HIV: a mathematical model for New York City. Philosophical Transations of the Royal Society of London B Biological Sciences 321: 171–187.

Bolker B, Grenfell BT. 1993. Chaos and biological complexity in measles dynamics. Proceedings of the Royal Society of London B Biological Sciences 251: 75–81.

Chandler T, Fox G. 1974. 3000 years of urban growth. New York: Academic Press.

Cliff AD, Haggett P. 1980. Changes in the seasonal incidence of measles in Iceland, 1896–1974. Journal of Hygience 85: 451–457.

_____. 1984. Island epidemics. Scientific American 250: 38.

_____. 1988. Atlas of disease distributions: analytic approaches to epidemiologic data. Oxford (UK): Basil Blackwell.

Clutton-Brock J. 1987. A natural history of domesticated mammals. Austin (TX): University of Texas Press.

Cohen MN, Armelagos GJ. 1984. Paleopathology at the origins of agriculture. New York: Academic Press.

Cumpston JHL. 1927. The history of diptheria, scarlet fever, measles, and whooping cough in Australia, 1788–1925. Canberra (AUT): Commonwealth of Australia, Department of Health.

Diekmann O, Metz JAJ, Heesterbeck H. 1990. On the definition and the computation of the basic reproductive ratio R_0 in models for infectious diseases in heterogeneous populations. Journal of Mathematical Biology 28: 365–382.

Dietz K. 1993. The estimation of the basic reproductive number for infectious diseases. Statistical Methods in Medical Research 2: 23–41.

Dobson AP. 1988. The population biology of parasite-induced changes in host behavior. Quarterly Review of Biology 63: 139–165.

Ewald PW. 1983. Host-parasite relations, vectors, and the evolution of disease severity. Annual Review of Ecology and Systematics 14: 465–485.

Ferguson RG. 1933. The Indian tuberculosis problem and some preventative measures. National Tuberculosis Association Transactions 29: 93–106.

_____. 1934. Some light thrown on infection, resistance and segregation by a study of tuberculosis among Indians. Transactions of the American Clinical and Climatological Association 20: 18–26.

Fine PEM. 1993. Herd immunity: history, theory, paractice. Epidemiologic Reviews 15: 265–302.

Fox JP. 1983. Herd immunity and measles. Reviews of Infectious Diseases 5: 463–466.

Fox JP, Elveback L, Scott W. 1971. Herd immunity: basic concept and relevance to public health immunization practices. American Journal of Epidemiology 94: 179–189.

Garenne M, Aaby P. 1990. Pattern of exposure and measles mortality in Senegal. Journal of Infectious Diseases 161: 1088–1094.

Garrett L. 1994. The coming plague: newly emerging diseases in a world out of balance. New York: Farrar, Straus & Giroux.

Grenfell BT. 1992. Chance and chaos in measles dynamics. Journal of the Royal Stastical Society B. 54: 383–398.

Hinman AR, Judd JM, Kolnik JP, Daitch PB. 1976. Changing risks of tuberculosis. American Journal of Epidemiology 103: 486–497.

Kermack WO, McKendrick AG. 1927. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London A 115: 700–721.

_____. 1932. A contribution to the mathematical theory of epidemics. Part II. The problem of endemicity. Proceedings of the Royal Society of London A 138: 55–83.

Levins R, Awerbuch T, Brinkman U, Eckardt I, Epstein P, Nakhoul N, Albuquerque de Possas C, Puccia C, Spielman A, Wilson MA. 1994. The emergence of new diseases. American Scientist 82: 52–60.

Livi-Bacci M. 1991. Population and nutrition: an essay on European demographic history. Cambridge (UK): Cambridge University Press.

MacDonald G. 1956. Theory of the eradication of malaria. Bulletin of the World Health Organization 15: 369–387.

May RM, Anderson RM. 1979. Population biology of infectious diseases: Part II. Nature 280: 455–461.

_____. 1990. Parasite-host coevolution. Parasitology 100: S89–S101.

McNeill WH. 1976. Plagues and people. New York: Anchor Press/Doubleday.

Most H. 1940. Falciparum malaria in drug addicts. American Journal of Tropical Medicine and Hygiene 20: 551–569.

Nowak MA, Anderson RM, McLean AR, Wolfs TFW, Goudsmit J, May RM. 1991. Antigenic diversity thresholds and the development of AIDS. Science 254: 963–969.

Olsen LF, Schaffer WM. 1990. Chaos versus noisy periodicity: alternative hypotheses for childhood epidemics. Science 249: 499–504.

Panum PL. 1940. Observations made during the epidemic of measles on the Faeroe Islands in the year 1846. Delta Omega Society.

Snider DE. 1944. Tuberculosis: the world situation. History of the disease and efforts to combat it. Pages 13–33 in Porter JDH, McAdam KPWJ, eds. Tuberculosis: back to the future. New York: John Wiley & Sons.

Wrigley EA. 1969. Population and history. New York: McGraw Hill.

[UNICEF] United Nations Children's Fund. 1995. State of the world's children 1995. Oxford (UK): Oxford University Press.