Books on cognate subjects

The mathematical theory of infectious diseases and its applications	N. T. J. BAILEY
Sampling methods for censuses and surveys	F. YATES
Experimental design: selected papers	F. YATES
Estimation of animal abundance and related parameters	G. A. F. SEBER
Statistical epidemiology in veterinary science	F. B. LEECH & K. C. SELLERS
Experiments: design and analysis	J. A. JOHN & M. H. QUENOUILLE
The advanced theory of statistics	SIR MAURICE KENDALL, A. STUART & J. K. ORD
Multivariate analysis	SIR MAURICE KENDALL
Stochastic point processes and their applications	S. K. SRINIVASAN
Analysis of categorical data	R. L. PLACKETT
Statistical method in biological assay	D. J. FINNEY

Complete catalogue available from Charles Griffin & Company Ltd

The Biomathematics of Malaria

Norman T. J. Bailey, M.A., D.Sc.

Unit of Health Statistical Methodology, World Health Organization, Geneva. Professor in the Division of Medical Informatics, University of Geneva. Formerly Professor Biomathematics, Cornell University Graduate School of Medical Sciences, and Member of the Sloan-Kettering Institute for Cancer Research

CHARLES GRIFFIN & COMPANY LTD
London and High Wycombe

Contents

Preface

The wide prevalence of parasitic diseases in many countries of the developing world is generally recognized as a major obstacle preventing hundreds, if not thousands, of millions of people from extricating themselves from a morass of misery and suffering. Interacting factors such as undernutrition, economic underdevelopment, environmental stress and pollution, as well as social and political conflicts, make the issues even more complex and their resolution more difficult. It follows that an understanding of the population dynamics of parasitic diseases is not likely to be sufficient in itself to lead to their eradication. And, unfortunately, such knowledge as we have is only at a preliminary stage.

There is now an extensive literature on the mathematical theory of infectious disease, with some applications to parasitic diseases in particular. This corpus of investigation provides a useful background to further development. It has certainly led to a clarification of concepts and thinking, and some mathematical models have been used to facilitate the choice of intervention strategies, though mainly in the area of bacterial disease control. In a few cases it has been possible to fit models to data in a statistically satisfactory way. The chief work has been with measles, but successful applications have also been made to infectious hepatitis and the common cold. And in the area of parasitic disease, the first successful statistical fitting of a model to field data was achieved for malaria by Dietz, Molineaux and Thomas in 1974. While all this is promising, and certainly intellectually exciting, it is not enough. The total research effort available is dissipated over a wide area, and most of it has been concerned more with mathematical niceties than with practical understanding and control.

As a consequence, any general work dealing with the mathematical theory of infectious disease will inevitably be selective in its subject-matter and unlikely to be closely related to real-life control situations. In the latter, about 95 percent of the effort tends to be concerned with planning, administration, financing, logistics, staff problems, haphazard data collection, computer print-outs, qualitative evaluation, etc., and only about 5 percent with applied mathematical thinking and the quantitative assessment of results.

The time has come, therefore, to supplement general theory by concentrating more effort on specific diseases. Those involving the greatest menace should receive most attention, and parasitic diseases in particular have a high-priority claim. Amongst these, pride of place should be accorded to malaria, whose victims number some one hundred and fifty million annually. Although in the 1950's it was confidently believed that the goal of total eradication could be achieved – a policy which subsequently turned out to be highly successful with smallpox – it now appears that malaria is much more resilient to external interference than had been realized, and at the present time there is a resurgence of the disease in spite of the continual increase of knowledge.

Before malaria is satisfactorily under control – eradicated in some areas, reduced to an acceptably low level of endemicity in others – there will have to be a greater degree of co-ordination between a wide range of disciplines, including clinical science, diagnosis and treatment, immunology, parasitology, entomology, epidemiology, mosquito control, preventive medicine, public health management, etc. There will also have to be more effective co-operation between ministries of health, research institutions and the pharmaceutical industry. There are clearly both national and international aspects, the latter being a special concern of the World Health Organization.

We have here a complex system consisting of a number of separately identifiable, though mutually interacting, component activities. All reputable workers in the field of malariology have first-hand knowledge of several of these activities, though few see the whole picture in all its quantitative detail, and perhaps none will fully understand the behaviour of the whole system as it operates within the economic, social and political constraints of a given society. This may explain some of the difficulties of predicting with sufficient accuracy the likely consequences of any chosen strategy of intervention. And if one cannot predict the consequences of a given strategy, one had no rational basis for choosing between the available alternatives.

One discipline which has a considerable contribution to make to most if not all of the substantive areas mentioned above is that which deals with quantitative aspects and the methods of drawing scientifically valid conclusions from the assumptions: it may be termed "biomathematics", i.e. mathematics applied in a specifically biological or medical context. The domain referred to includes statistics, modelling, operational research, cost-benefit analysis, systems analysis, system dynamics, medical informatics, etc. I believe that a concerted effort to develop this whole

area of quantitative support in the context of specific diseases would yield results of general value that are unobtainable through more diffused activities, as well as producing a significant contribution to the control of the disease in question.

So far as malaria is concerned, a great deal is already known about the biology and epidemiology. A sound basis is therefore available on which to build tentative mathematical descriptions. Indeed, work on the latter has been going on since Sir Ronald Ross's Threshold Theorem appeared in 1909, but only recently has it been possible to develop models that stood some chance of being acceptable to malariologists as well as offering the prospect of good predictive behaviour after further research and development. This work will be described in the book, starting with the urgency of the malaria situation today, and going on to the epidemiology of malaria, the historical background, the role of quantitative methods in general, the principles of modelling, the mathematical theory of host-vector diseases, applied malaria modelling to date, and statistical estimation problems.

The ensuing discussion is somewhat more speculative. First there is an account of the principles and potentialities of control theory. Then we examine the importance of sensitivity theory in assessing the way in which uncertainties in assumptions, especially those in any kind of model, whether explicit or not, may affect the conclusions. This opens up an area of great practical importance to overall policy and the choice of control strategies. Finally we consider how operational research and systems analysis can be used to provide a more precise understanding of the behaviour of the complex system of factors involved in the spread and maintenance of malarial infection, with a view to facilitating public health control. The utility of these latter approaches has yet to be demonstrated, but they have had great success in other fields, being essentially composed of a blend of common sense and the scientific method applied to problems of management, administration and decision-making in the control of processes in the real world.

A special reason for advocating the adoption of a systems approach is that, particularly in Africa, malaria may be only one of several parasitic infections simultaneously present in the community – a situation referred to as "polyparasitism". As many as seven such infections have been observed in a single village, including for example three types of filarial infections and two types of schistosomiasis, together with malaria and hookworm – less than 20 percent of those examined being free of all these infections, and some individuals having as many as five. The quantitative

epidemiological modelling of such a situation presents enormous difficulties since the various infections interact with one another. However, system-dynamics modelling is beginning to have some success in other complex areas such as aquatic ecosystems, and it may be possible to make similar advances with polyparasitism.

To sum up, I believe that malaria control programmes would have a much better chance of success if the decision-making and strategy selection were more securely based on scientifically validated epidemiological models, and if the planning and broad implementation were more in accord with the principles of operational research and systems analysis. The present book is designed to present existing methodology and future potentialities to biomathematicians who may have opportunities for serious work in this field, as well as to those who may be involved in collaborating with biomathematicians and in using the practical techniques developed.

I should like to take this opportunity to express my indebtedness to several friends and colleagues with whom I have had illuminating discussions over the years, especially Prof. Niels Becker, Prof. Klaus Dietz, Dr Paul Fine, Dr Louis Molineaux and Dr Ingemar Nåsell, some of whom saw and criticized early drafts of certain chapters; any remaining errors are of course my own responsibility. I should also like to thank all those authors who have continued to send me, over a long period of time, pre-publication drafts of their papers, as well as subsequent reprints. Thanks are also due to W. H. Freeman & Company for permission to print a revised version of the malaria cycle diagram that originally appeared in the article "Malaria" by Carlos A. Alvarado and L. J. Bruce-Chwatt in the May 1962 issue of *Scientific American*. Finally, I should like to thank Mrs Mary Martin for preparing an extremely clear and legible typescript.

NORMAN T. J. BAILEY

Geneva
March 1982

1 The world threat of malaria

We all hope that the great pestilential scourges of history are, and will remain, things of the past. The frightful mortalities incurred and widespread resultant desolation can still chill the spirit of anyone who reads the contemporary accounts. As is now widely known, there were in Europe in the fourteenth century approximately 25 million deaths from the Black Death alone out of a total population of some 100 million. In many regions whole villages and towns simply ceased to exist. Similar horrors were being enacted in China at about the same time. In 1331 an epidemic, also probably plague, is reported as having killed nine-tenths of the population of Hopei Province. Twenty years later, up to two-thirds of local populations were said to have died in epidemics in eight widely scattered areas of China.

Such examples are not isolated occurrences, but have been repeated in various forms over the centuries. Even nearer to our own time in the twentieth century the global mortality in the 1919 influenza pandemic was reckoned to be around 20 million in twelve months. Although most developed countries are now happily free from epidemic disasters on a grand scale, anxiety is not far below the surface. A few cases of Lassa fever or Legionnaire's disease are enough to cause disquiet. But what will happen if a new and virulent mutation of influenza suddenly appears? Or if rabies crosses the Channel into the United Kingdom, hitherto uninfected? Occasionally the virtually extinct smallpox virus escapes from a laboratory: what risks are run by a highly susceptible local population unprotected by vaccination? Again, what might be the consequences of the accidental release of supervirulent organisms specially manufactured in biological warfare laboratories?

Such matters are taken quite seriously in developed countries, and risks can only be minimized by eternal vigilance. Surveillance must be efficient and continuous, and public health intervention must be appropriate and rapidly implemented when needed.

But all of this pales almost into insignificance when compared with the load of misery and suffering perpetually borne by the populations of many Third World countries as a result of endemic infectious diseases, largely

parasitic in nature. The figures are staggering, though perhaps constant repetition is preventing some of us from fully grasping the implications: although over the centuries malaria has been eradicated from many parts of the world simply by draining swamps and marshes, and more deliberate malaria eradication programmes have further enabled many hundreds of millions to free themselves from the disease, there are still approximately 350 million people in the world living in areas where the infection is endemic and where no organized antimalarial measures have yet been undertaken. A very high proportion of these people are likely to be affected. Again, schistosomiasis is believed to affect around 200 million people, with possibly three times this figure being exposed to the infection. For hookworm disease the figure is approximately 450 million. Another rampant infection is trachoma, which is due to a bacterial rather than a parasitic organism. Methods of treatment have been improving, but the total number implicated may be of the order of 400–500 million.

Such figures are of course mere statistics. The human suffering caused by the diseases alone is catastrophic enough, but when it is remembered that the effects are often compounded by chronic undernutrition, economic underdevelopment, growing environmental stress and pollution, together with social and political conflicts, the result virtually amounts to an almost unrelieved state of disaster.

An added complication in many areas is the existence of what is referred to as "polyparasitism", namely the simultaneous occurrence of several forms of parasitic disease. This is particularly common in Africa. Up to seven infections have been observed in the population of a single village, including malaria, two kinds of schistosomiasis, three types of filariasis and hookworm, some individuals having as many as five of these infections and only 20 percent being free. Of course, this enormously complicates any attempts to understand the overall population dynamics since the diseases tend to interact with one another. However, the top priority is still clearly malaria, and for the purposes of the present study we shall concentrate attention on it.

A compelling review of the urgency of the world malaria position appeared in *World Health* (1978, Aug.-Sept., p. 38), and summarizes the situation as reported to the World Health Assembly earlier that year. See also the recent global assessment by WHO shown in the map in Fig. 1. Malaria in fact menaces more people in the world than any other single disease. The situation is actually deteriorating, with reported cases having doubled over a five-year period, but many cases are not reported at all. It is estimated that the global total of malaria victims now numbers around

one hundred and fifty million annually, while in Africa alone *a million children die every year from malaria* before they reach the age of five. This latter fact should by itself be sufficient to motivate continuous effort dedicated to effective action.

A critical situation also exists in South-east Asia where, after partial eradication, malaria has now re-established itself at its previous endemic levels in large areas of Bangladesh, India and Sri Lanka. In India itself, for example, mosquito control programmes had reduced the annual figure of 75 million cases in 1935 to about 60 000 in 1962. By 1974 the level had risen again to 3 million, and in 1976 stood somewhere in the region of 6 million.

It must be recognized that global malaria eradication programmes have partially succeeded, in that 500 million people have been freed from the threat of endemic malaria, thus allowing new socio-economic developments in many areas of the world. Nevertheless, the continual deterioration in many countries, resulting in some instances in a 30- to 40-fold increase in cases reported over a ten year period, is extremely disquieting.

There are many complex factors associated with this resurgence of malaria. For example, in 107 countries where malaria exists 62 have reported the occurrence of mosquito resistance to the usual insecticides. In addition, resistance of some parasite strains to the 4-aminoquinolines, the most commonly used antimalarial drugs, has been observed in 20 countries. There is thus an urgent need for the development and application of new insecticides and alternative forms of medication. It appears that the number of new insecticides offered by industry to the World Health Organization has dropped to zero, presumably because other products destined for other markets are more profitable. There are biological alternatives, such as genetic control or the use of larvicidal fish, but these developments are either still in their early stages or are not yet fully effective.

These destructive influences are further aggravated by economic difficulties, particularly the rapidly increasing costs of petrol and insecticides, which makes the financing of even limited malaria-control operations progressively uncertain.

This, briefly, is the world situation with regard to malaria. Considerable success in control and eradication has been achieved in the past, and maintained over a wide area of the globe. Optimism for the future has, however, been severely shaken in recent years with the progressive resurgence of the disease which still threatens hundreds of millions of people. It is no longer certain that existing methods of control or eradica-

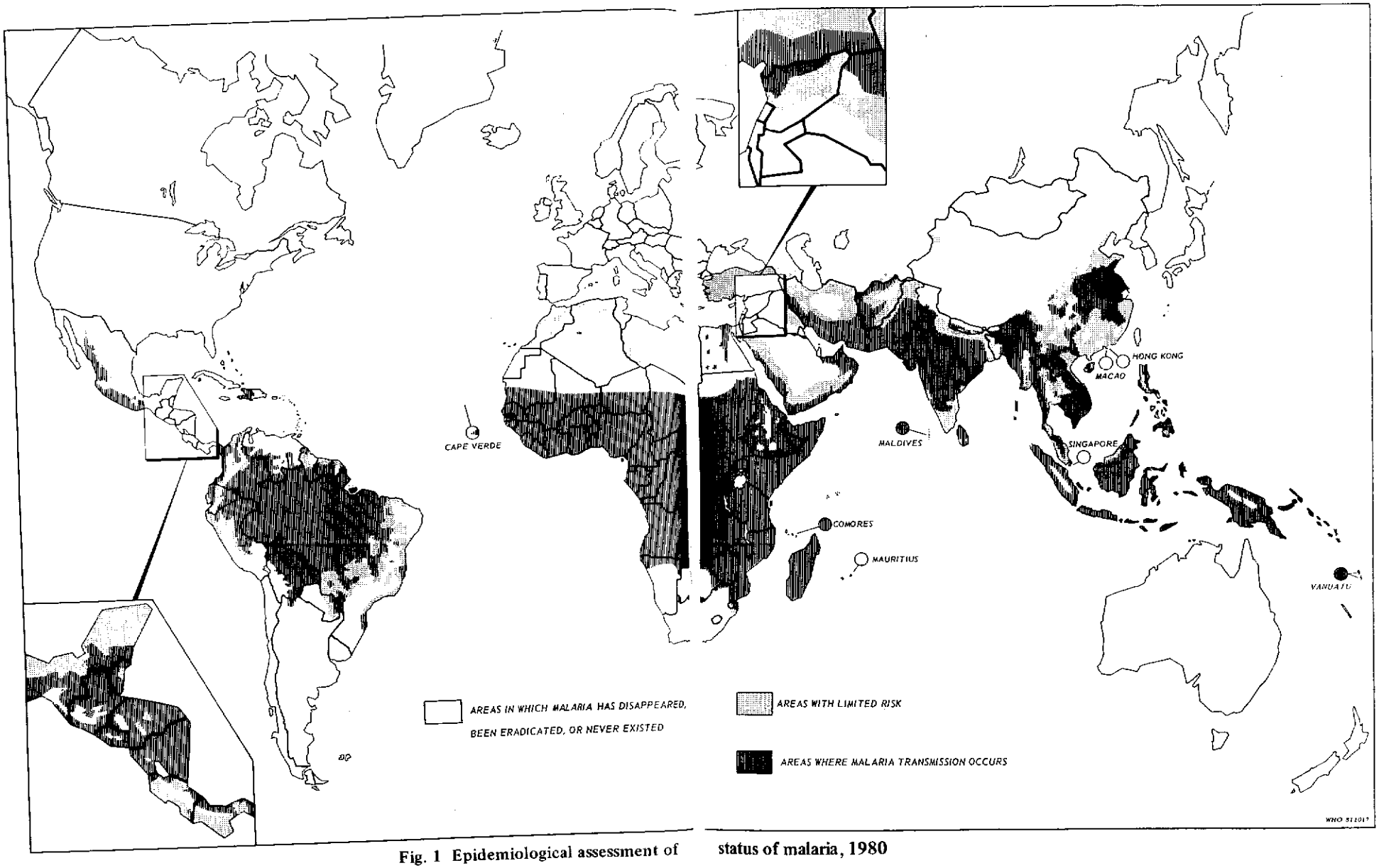

Fig. 1 Epidemiological assessment of status of malaria, 1980

Map published in *Vaccination Certificate Requirements for International Travel and Health Advice to Travellers*, World Health Organization, Geneva, 1982

tion will be able to stem the tide. The possible consequences for social, economic and political stability are incalculable.

It would, of course, be naïve to suppose that expanded biomathematical and operational research efforts would automatically contribute anything substantial to an alleviation of these gloomy prospects. At the same time they will be essential to underpinning new efforts to understand and control the highly complex network of interacting factors that are involved in the current recrudescence of malaria. Improved epidemiological knowledge is required, including a better understanding of the population dynamics of malarial spread. The selection of suitable control strategies must be made considerably more efficient. This means finding new ways of predicting with sufficient accuracy the likely consequences of any given strategy *in advance of actual implementation*. For if the consequences of alternative available strategies cannot be predicted to the point where we can effectively discriminate between them, there is no rational basis for any kind of choice.

It may therefore be expected that biomathematics and operational research will be able to make effective contributions to the renewed fight against malaria – provided that they avoid purely academic abstractions and concentrate on supporting in an integrated manner very practical undertakings specifically designed to choose optimal strategies for malaria control and to alleviate widespread human suffering.

2 The epidemiology of malaria

2.1 Introduction

The purpose of this chapter is to give a short account of the biology and epidemiology of malaria in such detail as is required for embarking on serious biomathematical work. Some of the most important results obtained in the past by renowned malariologists like Ronald Ross and George Macdonald have depended on only a relatively small number of key concepts being significantly described and related by certain fundamental mathematical equations. The general principles of such an approach, and in particular the role of so-called mathematical modelling, will be discussed in Chapters 4 and 5. For the moment we merely note that most epidemiologists are obliged to rely on some quantitative techniques in order to pursue their calling, and malariologists are no exception. Moreover, any work aimed at elucidating the population dynamics of an infectious disease must employ mathematical approaches – the central problem is how to devise methods that are simple enough to be mathematically manageable, as well as comprehensible to scientists who are not professional mathematicians, while at the same time giving sufficient emphasis to concepts which must be incorporated in our thinking if an acceptable degree of realism is to be achieved.

For biological, clinical and epidemiological detail the reader should consult appropriate standard textbooks on the topics in question. So far as broad ideas in infectious disease theory are concerned, there are a number of comparatively non-technical books worth reading such as Greenwood (1935), Zinsser (1935), Pickles (1939), Geddes Smith (1943), Winslow (1943), Hare (1954), Burnet & White (1972), and McNeill (1977). There is also a useful book by Lapage (1963) which is specifically devoted to parasitic diseases as a whole. Malaria itself has been covered in greater detail by Macdonald (1957, 1973) and Pampana (1969), while an excellent authoritative and up-to-date account is given in the recent book by Bruce-Chwatt (1980).

We shall therefore start below by reviewing briefly the basic concepts in the spread of infectious diseases where the transmission is from person to person, and then go on to consider the complications introduced by the

existence of an intermediate host or vector. The special aspects of malaria itself will then be examined, together with a range of influential factors deriving from various socio-economic and environmental activities.

It must, however, be emphasized that the discussion given in the present chapter is for general guidance only. It indicates the range and variety of the aspects that may have to be taken into account in trying to develop realistic biomathematical applications. For definitive statements, acknowledged authorities must be consulted, although it must be admitted that there are sometimes appreciable differences of opinion on important issues.

2.2 Basic concepts in the spread of infection

First of all let us be clear about the implications of the word "infection". We are dealing in this book primarily with malaria, and are thus concerned with one specific example of a disease that is "infectious" in the sense of being transmitted at some stage in the life-cycle of the relevant organism from an infected host to an uninfected susceptible. This broad definition leaves open the possibility of an intermediate animal or insect vector being involved.

Considerable variations in terminology will be found in the literature. "Infectious" in the above sense is common in English usage, but is often replaced by "contagious" in American usage. Again, "infective" is sometimes employed synonymously with "infectious", but is also used to indicate diseases that are caused by some identifiable invasive organism, as in bacterial endocarditis, where there is normally no case-to-case transmission. Similarly, "communicable", as opposed to "non-communicable", is another common synonym for "infectious" or "contagious". No single definition is perhaps entirely satisfactory, but in any case the biological and epidemiological context will normally make quite clear what is intended. And clarity is important since different underlying epidemiological mechanisms will have to be portrayed by different biomathematical representations.

There are many relatively simple infectious diseases, such as measles, influenza or diphtheria, where the infecting organism is a virus or bacterium and is readily transmitted from one individual to another by sufficiently close *physical contact*. The latter often occurs through the transmission of infected droplets, breathed out by one person and breathed in by another. But with some diseases the path of infection may

be through polluted water supplies, eating contaminated food, contact with infected clothing, etc.

An uninfected person, still in a susceptible state, is referred to as a *susceptible*. Such a person may have some natural resistance, or *immunity*, but if he does acquire an infection, he will not normally become immediately infectious to others. It usually takes time for the invading organisms to gain a foothold, and reach a point where they are established in sufficient numbers for a substantial emission of infectious material to occur. The time-interval between the receipt of the infection and the beginning of infectiousness is called the *latent period*. In some cases it is convenient to regard this as having zero duration, or alternatively as being of fixed length. More realistically, we might assign a statistical distribution to the length of the latent period, though this can lead to considerable mathematical complications, unless something simple like a negative-exponential curve is used.

When the latent period ends, the infected person becomes actively infectious, and is often called an *infective*, while the period of time during which infectious organisms are emitted is the *infectious period*. In simple quantitative studies we suppose the infectious period to be of fixed duration, and the intrinsic *infectiousness* of the infective during this period also to be constant. Indeed, in some models it is convenient to consider infectiousness as concentrated in a single point of time. On occasion, however, a more searching analysis may be required. The infectious period must then be considered as having variable length and tentatively assigned some statistical distribution, e.g. normal (gaussian). At the same time it may also be debated whether the infectiousness must not also be regarded as a quantity that varies over the infectious period, as well as depending to some extent on age, sex, etc.

While the infectious period starts when the infective begins actively to discharge infectious material in one way or another, it ends either in fact when this process stops biologically or virtually when the infective develops recognizable symptoms, is withdrawn from free circulation in the community, and is *isolated* more or less adequately from those who are susceptible. Thus, what should be taken as the effective infectious period depends partly on biological realities and partly on social behaviour. A highly infectious person, isolated in hospital and subjected to barrier nursing, is no longer capable of transmitting infection. But the same person lying in bed in an overcrowded dwelling in a densely populated community may be a considerable source of danger. A realistic quantitative model must therefore reflect, at least approximately, the circum-

stances actually obtaining in any community under study. It should be noted however that if the infectious period ends biologically before symptoms occur, it may be very difficult to collect any direct evidence as to the precise length and end-point of the infectious period.

We can now see quite easily how a complete cycle of events occurs which tends to lead to the maintenance of an infectious disease by its continual transmission from infectives to susceptibles. Whether only a minor outbreak takes place or a substantial *epidemic* builds up depends on the relative values of certain key parameters (we shall not pursue the exact quantitative relationships here). And in some circumstances, when there is a sufficient supply of new susceptibles, a balance may result with a steady *endemic* level of infection or disease.

It should be noted that in the above process of infectious spread, what is actually observed in general is not the occurrence of a new infection in a hitherto uninfected susceptible, but the recognition of a case of disease due to the appearance of symptoms in an infected person who has been circulating clandestinely in the community. That is, we only observe *removals*. Since the event of infection is usually unobservable, there are liable to be considerable mathematical complications in trying to estimate unknown, but epidemiologically significant, parameters from rather restricted data.

In special cases, however, the actual time at which infection occurred may be known with some accuracy, as when the only possible contact of a given susceptible with an infective can be traced to a visit to a particular household or social gathering on a clearly identifiable date. The interval between the receipt of infection and the onset of symptoms is usually referred to as the *incubation period*. Opportunities for direct observation of incubation periods are not common, but sometimes it is possible to make estimates by indirect means by fitting an appropriate model to observed dates of removal.

Another concept often used is the *serial interval*, which is the period of time between the observation of symptoms in one case to the observation of symptoms in a second case directly infected from the first. Although this can be the basic observable epidemiological unit in certain detailed community studies, it is a more complex concept than at first appears. It reflects to some extent the life-cycle of the infectious organism involved, but cannot be easily related directly to the mechanism of transfer.

It may therefore be more fruitful in a detailed analysis to use some breakdown of the incubation period into components, as implied in the

foregoing discussion. For example, when symptoms leading to removal appear either at the end of the infectious period or during it, the incubation period is precisely equal to the sum of the latent period and the part of the infectious period when the infective is still a source of danger to other susceptibles.

As we have seen, in the case-to-case transmission of infection there has to be some kind of physical contact between a susceptible and an infective. This is often defined more exactly in terms of the notion of *adequate contact*, which means a certain degree of proximity between two individuals such that, if one is susceptible and the other is infectious, a transmission of infection to the susceptible will actually take place. The infective will of course continue to be infectious until isolation or the end of his infectious period.

Adequate contact, as defined above, is thus a relationship that may occur by chance between any two individuals in a community. If this chance were so large as to amount to certainty then the disease would spread to all susceptibles more or less immediately after the first infection occurred. In fact there is a process of infection taking place over a period of time. The character of this process depends on the average chance of adequate contact, in general a probability less than one. This probability is really a composite, average index, depending on the resistance of the susceptibles to infection, the infectiousness of the infectives (involving the numbers and virulence of organisms transferred), the actual degree of proximity achieved both quantitatively and qualitatively, etc. In a group containing several infectives a particular susceptible will remain free of disease only in so far as he escapes adequate contact with any of them.

The notion of the chance of adequate contact is the point at which probability enters into the mechanism of infectious disease spread. In small groups we should expect statistical fluctuations to be appreciable, and this is in fact immediately evident to direct intuitive observations. Mathematically, we have to resort to some form of stochastic modelling. In large groups, at least as a first approximation, it may be possible to replace the chance of adequate contact between two individuals by an average rate of transmission between infectives and susceptibles.

In simple quantitative descriptions we may find that the unanalysed concept of adequate contact is perfectly satisfactory. But in more searching investigations it may be necessary to disaggregate some of the components and examine them separately. Thus we may want to distinguish such contributory factors as the infectiousness of the infective, the quantity of organisms transmitted, the virulence of the invading organisms,

the relative immunity of the susceptible, and the dependence of all these items on age, sex, social class, etc.

Immunity is in fact an item of special importance. The resistance of individual susceptibles to infection may vary considerably in practice, depending on the extent to which their internal biological and biochemical mechanisms can deal with invading organisms. Some immunity occurs naturally; in other cases it is built up gradually in response to previous exposures to small doses of infection; yet again it may be stimulated artificially and rapidly, as with certain vaccinations and immunizations. Moreover, a number of diseases are themselves a direct source of immunity. Measles, for example, usually gives rise to a permanent immunity in any individual who has been attacked. When measles cases have recovered, they continue to circulate in the community but contribute virtually nothing to the further spread of the disease, beyond diluting the population numbers to some extent. Influenza, on the other hand, confers immunity, but this is liable to be strain-specific, and protection is therefore limited. Again, in many specific epidemiological investigations some individuals may have forgotten earlier, possibly immunizing, attacks of a disease, while others may have lost an immunity resulting from a previous attack. All this makes immunity a somewhat elusive factor.

In looking closely later on at the detailed processes of the spread of malaria we shall find several instances of the need to use certain indexes both as they stand and in relation to their component parts. The problems of identifying varying levels of relative immunity in both susceptible and infected individuals will also have to be faced.

With some diseases, like poliomyelitis, tuberculosis or typhoid fever, there are certain individuals who, while apparently healthy themselves, harbour the infection and may transmit it to others. These are the so-called *carriers*. Because they are not readily observable, but may exist in large numbers, it is always important, in establishing the basic epidemiology of any infectious disease, to determine whether carrier states in one form or another are a significant aspect of the whole picture.

2.3 Reservoirs, vectors and intermediate hosts

The discussion of the previous section introduced some of the main concepts in the spread of infectious disease, and to simplify the presentation we concentrated on the more straightforward person-to-person types of transmission such as occur with many types of virus and bacterial diseases. As already hinted, complications can easily arise when there are

alternative, indirect paths of transmission involving polluted water supplies, contaminated food, contact with infected material, etc. In specific applications it may or may not be necessary to incorporate such features directly in the construction of an epidemiological model. If, for example, the main source of spread is through person-to-person contacts then it may be possible to ignore other sources of infection if these are sufficiently small. But in many types of food poisoning, case-to-case transmission may be virtually non-existent, while the local pattern of food preparation, distribution and disposal may be the really important factor.

Before examining our specific topic of malaria, it is worth looking at the general role played by animals and insects in the mediation of disease transmission. Take, for example, plague which has a very long history. In the so-called bubonic form of the disease, the plague bacillus *Pasteurella pestis* is normally acquired from the bites of infected fleas carried by certain infected rodents, especially the "black rat". The rats obtain the infection both from other infected rats, as well as from wild rodents whose burrows harbour the plague bacillus. Such animal populations, together with their fleas, provide natural *reservoirs* for maintaining the disease more or less indefinitely. Human infection and disease is thus an almost accidental by-product of a naturally occurring process.

There is also a pneumonic form of plague in which the infectious organisms are transmitted directly from person to person by droplet infection. Without specialized treatment pneumonic plague is apt to be fatal, and the disease is unlikely to maintain itself for long in the absence of a reservoir of infected rats.

Rabies is another disease which is kept going in animal populations, especially wild foxes, and which is transmitted on occasion to man, either directly from infected wildlife or through the intermediate agency of the domesticated dog.

The occurrence of natural disease pools is widespread: there are over 100 diseases known to be shared by man and wild animals, birds or fish, and also over 100 shared with domesticated animals (Hubbert *et al.*, 1975, ch. XCII). For this reason the total eradication of many diseases is infeasible, though by careful public-health management and proper attention to personal hygiene it is possible to achieve a considerable measure of control, resulting in either an acceptably low endemic level or only occasional localized outbreaks.

In addition to the broad concept of reservoirs or pools of infection, there is also the more specialized notion of a *vector*, which in some way or other actually carries infection from one host to another. Thus, typhus is

transmitted by fleas and lice, yellow fever and malaria by mosquitoes, sleeping sickness by tsetse flies, schistosomiasis by snails, etc. Since such vectors are very active in permitting the infecting organism to migrate to new hosts, there is usually some ecological adaptation of the organism's life-cycle to the physiology and behaviour of the vector so as to permit transference to occur more expeditiously. It may also happen, however, that the vector fulfils a purely mechanical role in moving the infective organism at some stage of its life-cycle to a new host without any actual development taking place. Again, a number of parasitic diseases have very complex life histories in which the infective organism is parasitic during each life-cycle on two or more different species. The fish tapeworm, for example, has its final adult form in man. Man is thus the *final host*, defined as being the site where fertilized eggs are produced. But there are also two *intermediate hosts*, which harbour larval stages, the first in certain crustacea and the second in freshwater fish. There is in addition a free-swimming non-parasitic stage of the cycle between man and the first intermediate host.

It is not intended to develop here any detailed account of such complications, but merely to introduce some of the chief concepts which may be helpful in setting the scene for a more specific discussion of the epidemiology of malaria.

2.4 The life-cycle of malarial parasites

Compared with several other parasitic diseases, the main outline of the transmission of malaria is relatively simple. Human malaria is directly caused by four species of protozoa in the genus *Plasmodium*, and they are transmitted to man only by female mosquitoes belonging to certain species of the genus *Anopheles*. Conversely, female mosquitoes can pick up the infection when they bite infected humans. Only female mosquitoes are involved since only they suck blood: males take their liquid meals from fruit juices and elsewhere. In the case of the malaria parasite it is, on the accepted definitions, the mosquito which is the final host, i.e. where the fertilized eggs are produced. Man is then the intermediate host harbouring the asexual phase of development. It is ironic that it is only man who suffers from malarial disease: the mosquitoes appear to be unaffected by their own role in the process.

Apart from the *Plasmodium* species causing malaria in man there are other species infecting many kinds of birds and animals. Generally speaking, however, these are not infective to man, or vice versa. The only

exception is the case of certain monkeys which can very occasionally act as an intermediate host for the human parasites instead of man himself.

Now the general course of the *Plasmodium* life-cycle in the species that infect man is shown diagrammatically in Fig. 2, and may be described as follows. Suppose that a human subject is bitten by a female anopheline mosquito. If the latter's salivary glands contain *Plasmodium* cells in the *sporozoite* stage these may be injected into the human blood-stream as the mosquito sucks the blood. Within about 30 minutes the sporozoites make their way in the blood to the liver, where they complete their growth and multiply asexually to produce *merozoites*. Some 40 000 of the latter can be produced in six days by a single sporozoite of the potent *Plasmodium falciparum* species. In the case of the latter parasite the merozoites then pass out of the liver into the blood, where they enter red blood-cells. Further multiple asexual reproduction then takes place. Daughter merozoites burst and destroy their host red cells before being liberated into the blood to repeat the process in other healthy red blood-cells. Repetition of this subcycle of events can produce millions of parasites. Symptoms of malaria may appear in the host when the total parasite load in the blood-stream is of the order of 1 000 million. As defined in Section 2.2, the *incubation period* is measured as the time elapsing between receiving the infection from, in this case, an infected female mosquito, and the appearance of clinical symptoms.

After a certain time the asexual multiplication of some merozoites comes to an end, and distinct male and female *gametocytes* are formed in red blood-cells. No new development of the parasite can take place in human blood. But if a female anopheline mosquito sucks up blood containing gametocytes these are retained and develop further, all other phases of the parasite that may be present in the blood simply being digested. Once in the mosquito's stomach the gametocytes give rise to male and female *gametes*, i.e. sex-cells equivalent to mammalian spermatozoa and ova. Fertilization of female gametes then produces *zygotes*, which go through a process of development that gives rise to a large number of new sporozoites. These sporozoites migrate to the salivary glands of the mosquito, where as many as 100 000 may be present, and the process begins all over again.

It can thus be seen that the life-cycle of the malaria-producing *Plasmodium* takes place entirely within the two hosts, man and mosquito, with no external free-living stages.

There are, however, some variations of detail between the different *Plasmodium* species. One particularly important difference is that, whereas

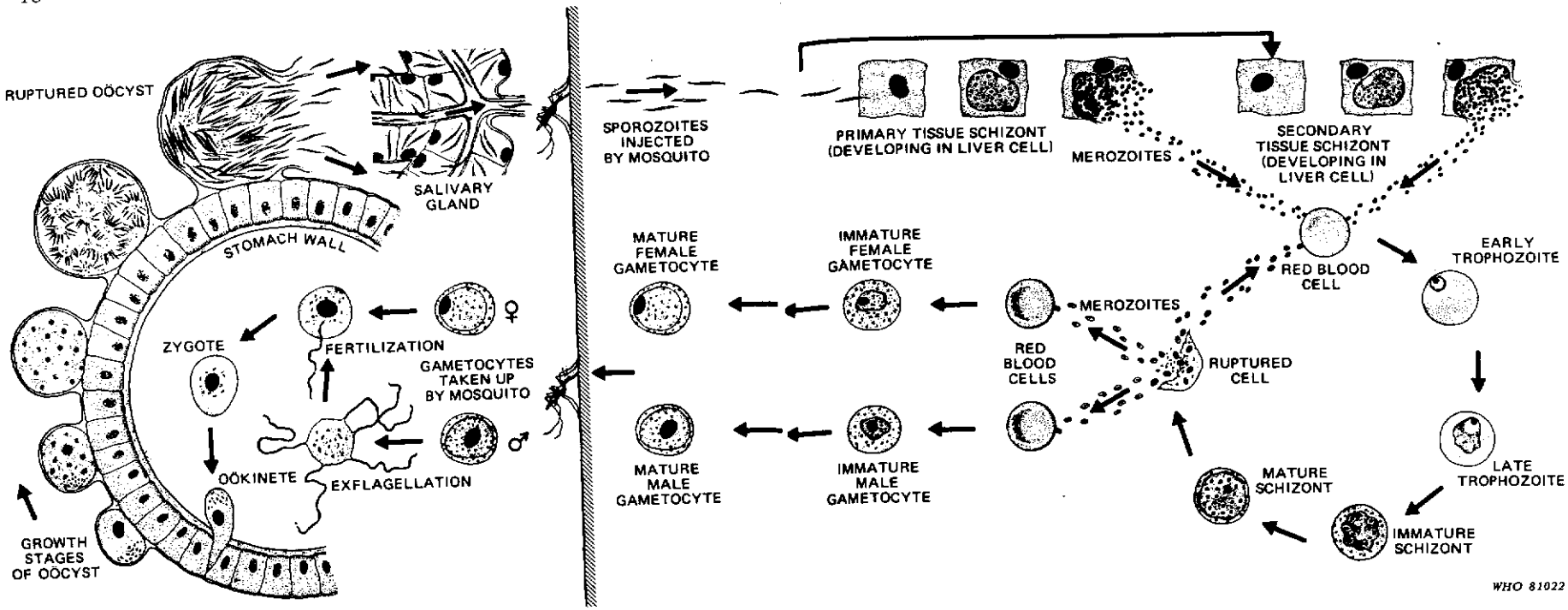

Fig. 2 The life-cycle of the malaria parasite

in the mosquito and in the human host

Reproduced with modifications from Alvarado and Bruce-Chwatt, *Scientific American*, May 1962, by permission of W. H. Freeman & Company

in *P. falciparum*, and perhaps in *P. malariae*, the merozoite infection of the liver clears completely and continues development in the red blood-cells, in the other two species, namely *P. vivax* and *P. ovale*, the merozoite infection of the liver may persist and continue asexual reproduction as before, possibly leading to the occurrence of relapses after periods of quiescence. This phenomenon is also of particular importance in the chemotherapy of malaria (see Section 2.7 below).

For authoritative and more detailed discussions of the biology of the malaria parasite, the reader may consult Lapage (1963, ch. 8), Pampana (1969, ch. 1), Bruce-Chwatt (1980, ch. 2) etc., or for a more medically oriented presentation, Nnochiri (1975, Sections 2.1, 3.1 and 4.1).

2.5 Types of malarial disease in man

While the life-cycles of all four *Plasmodium* species, namely *P. falciparum*, *P. vivax*, *P. malariae* and *P. ovale*, follow the general qualitative pattern indicated in the previous section, there are quantitative differences between them, and recognizably distinct effects occur in terms of human disease.

In well-defined, uncomplicated infections by a single species of parasite in people who have had no previous malarial infections there are certain cyclic patterns of symptom occurrence which characterize different types of malaria. Generally speaking, the typical initial symptoms of an attack

of malaria are chill, shivering, various aches and pains, etc., followed by fever, and then profuse sweating and a fall in body temperature. These attacks start at the end of the incubation period and correspond to a certain concentration of parasites in the blood, leading to the destruction of red cells and the liberation into the blood-stream of merozoites and the products of the red-cell destruction. These events trigger off reactions in the infected person, and the repetition of the sub-cycle of merozoite reproduction as described in the preceding section leads to a corresponding cycle of symptomatic attacks.

The different kinds of malarial infection are characterized by different lengths in the cycles of attack, the usual classification being based on the old Roman system of counting inclusively. This means that a cycle which is, mathematically and scientifically, strictly two days will be counted as three, i.e. with the first and last days both included, and thus described as *tertian*.

In *P. vivax* the fever is liable to be irregular for two or three days to begin with, after which it settles down to a two-day cycle and is therefore called *tertian malaria*, or *benign tertian malaria* in contrast with more malignant forms. *P. falciparum* is especially liable to cause severe attacks with a fatal outcome. Starting irregularly, the fever tends to recur at intervals between 36 and 48 hours, giving rise to the title of *subtertian malaria*, or sometimes *malignant tertian malaria*. With *P. malariae* there is a three-day cycle and the disease is called *quartan malaria*. Finally, there are the relatively mild infections of *P. ovale*. These are tertian in character and may be confused with *P. vivax*.

In practice, this picture may be more complicated than the classification suggests because a given individual may be infected more than once by different mosquitoes. There may then be distinct broods of parasites in the blood with possibly overlapping life-cycles, a phenomenon referred to as *superinfection*. Thus, individuals infected with *P. vivax* can have attacks every day. This is a *quotidian malaria*. If the cycles become synchronized after a time, the frequency again becomes tertian. Quotidian malaria is also common with *P. falciparum*, partly because it is liable to be irregular anyway, and partly because of possible superinfection. In addition, there may also be multiple infections with two or more different kinds of parasites, e.g. *P. falciparum* and *P. vivax*, which complicates the clinical pattern still further.

Along with variations between the different *Plasmodium* species in the intervals between successive attacks, there exist differences in the corresponding incubation periods. Estimates vary somewhat, but a recent

account (Bruce-Chwatt, 1980, ch. 2 and 3) gives the following figures for averages: *P. vivax* and *P. ovale*, 13–17 days; *P. falciparum*, 12 days; and *P. malariae*, 28–30 days. It is also noted that short incubation periods of the order of 5 days may occur with heavy inoculations of *P. falciparum* in non-immune subjects. In addition, very long incubation periods have been observed in some strains of *P. vivax*.

Generally speaking, there is liable to be a great deal of biological variation in any quantitative measurements of this sort. The incubation period tends to vary with both the size of the dose of infection received and the relative resistance of the individual to the parasites which subsequently develop. This resistance, or immunity, is a key concept in the epidemiology of malaria (as with many other diseases) and its magnitude depends on age, general health, nutritional status, genetic constitution, exposure of past generations to malaria, etc.

Some individuals may have an inborn or *innate immunity* which protects them from infection either wholly or partially. This could explain why a few people seem able to live disease-free in malarious areas, taking no precautions against infection and not contracting malaria. More important, however, is the widespread occurrence of *acquired immunity* which develops in response to infection, and tends to increase when attacks are repeated. This type of immunity is not only specific to each species of malarial parasite, but is also specific to individual strains. But, unless a sufficient number of booster inoculations are received, the immunity that has been acquired will gradually be lost. Finally, there is the possibility of some degree of *congenital immunity* operating in the newborn over the first few months of life due to the acquisition of maternal antibodies. This whole subject is a very complex one, and for an extensive discussion of details see Pampana (1969, ch. 3) and Bruce-Chwatt (1980, ch. 4).

The above account of malarial disease deals of course only with an outline of some of the main features. There are in addition a vast number of clinical aspects, including numerous complications, covering both diagnosis and treatment (see Nnochiri, 1975, Sections 2.1, 3.1 and 4.1, for recent summaries).

2.6 Environmental and geographical aspects of malaria

Environmental and geographical factors are of very considerable importance in the occurrence and distribution of malaria. Where there are no anopheline mosquitoes there can be no indigenous malaria (although

isolated cases may of course be imported). In general this tends to restrict malaria to regions lying roughly between latitudes 65° N and 40° S, and having altitudes less than about 3000 metres. There is a natural inhibition in mountain areas and in deserts where there is a lack of the water necessary to mosquito breeding. In particular, malaria is absent from Pacific islands east of longitude 170° E because there are no anopheline mosquitoes.

There are geographical differences in the distribution of both parasites and mosquitoes. Thus the malignant *P. falciparum* is the most frequent species throughout the tropics and sub-tropics, and may also occur in some temperate areas. *P. malariae* is distributed over the same range but is much less frequent. *P. vivax* on the other hand has a less restricted geographical distribution and is prevalent in many temperate climates as well. Finally, *P. ovale* is found mainly in tropical Africa but sometimes in the West Pacific.

Temperature is in fact a critical factor. Where the mean monthly temperature is less than about 16°C malaria is liable to be absent, even in the presence of anophelines, because it is too cold for the parasite to develop inside the mosquito, both mosquito and parasite behaviour being affected in different ways. More specifically, the sexual cycle of the parasite, often referred to as the *extrinsic cycle*, or *extrinsic incubation period*, and taking place in the mosquito from the acquisition of gametocytes to the production of sporozoites, varies with both parasite species and temperature. Authorities seem to differ amongst themselves, but the following figures (Bruce-Chwatt, 1980, ch. 2) are fairly typical. For *P. falciparum* the extrinsic cycle is about 10 to 11 days when the external temperature is 25°C, 15–17 days when it is 23°C and 22 days at 20°C. For any given temperature, the corresponding average cycle will be a few days shorter for *P. vivax*, e.g. 16 days at 20°C. Figures for *P. malariae* are 26–28 days round about 25°C and 30–35 days at 20°C, while *P. ovale* takes about 16 days at 25°C. Above 32°C the proportion of parasites surviving drops off sharply, while development is indefinitely retarded at a lower limit: 15°C for *P. vivax* and 19° for *P. falciparum.*

Such results can be regarded only as a general guide: they illustrate the point that colder weather slows things down and a level may be reached when the longevity of the mosquito is insufficient for the extrinsic cycle to run its course. Thus *P. vivax* has an advantage over *P. falciparum* at lower remperatures since its sexual cycle can be more easily completed in the time available.

It is clear that, with these considerable variations between different

parasite species and different temperatures, great care must be taken to make quantitative investigations very flexible in their applications, and to ensure that any specific epidemiological models take full account of local conditions.

As to the mosquito, there are some 2 000 species altogether of which about 400 are members of the genus *Anopheles.* Approximately 60 of these species can carry malaria under natural conditions, but a short list of those that are of chief importance is probably no more than 20 or 30. An additional complication is that some species originally regarded as a single entity, such as *Anopheles gambiae*, are now seen as a complex of several species. However, we do not need to dwell on this aspect here. The major vectors in Africa are the *A. gambiae* complex and *A. funestus*, while in India they are *A. maculatus* and *A. minimus*. In Central and South America *A. aquasalis*, *A. albimanus* and *A. darlingi* are of the greatest importance, and so on. Many other species, however, have an appreciable role to play. There are many differences between the species in their behaviour with regard to feeding habits and choice of breeding sites, and these characteristics may have to be reflected in some way in any quantitative analysis that may be undertaken. Thus while most mosquitoes prefer to feed at night, some feed indoors and some outdoors, certain species confine themselves to man while others feed on birds or animals as well, some like muddy water but others prefer clear water, etc.

On the whole the distribution and prevalence of malaria is very uneven. In many areas it is endemic. In others, previously clear of infection, there may be epidemic outbreaks. Again, the process of eradication may itself give rise to a very variable historical picture. In the "island countries" of the Carribean malaria eradication has been very successful, but "mainland countries" tend to be less fortunate. And in the island of Sri Lanka, where endemic malaria was brought under control after the 1947 campaign, adverse climatic changes after 1967 plus a slackening of vigilance led to serious subsequent outbreaks of epidemic malaria.

A certain standard classification of the level of endemicity should perhaps be mentioned at this point, because the terms employed frequently appear in the literature. This is based on the "spleen-rates" for children in the 2-9 year age-group. "Spleen-rate" is here defined as "the percentage of persons in whom enlargement of the spleen can be detected by palpation". When the rate in children is constantly over 75 percent, with a low rate in adults, the malaria is said to be *holoendemic*. On the other hand, if the child rate is constantly over 50 percent, with a high adult rate, the term *hyperendemic* is used. Two other definitions also used are *meso-*

endemic and *hypoendemic*, corresponding to child spleen-rates of 11-50 percent and 0-10 percent, respectively, without reference to adult values. This classification was developed in relation to the situation in Africa, and has given rise to some controversy. The first two terms are used fairly freely, but the second two have not received general acceptance.

2.7 Treatment and control

It is appropriate to conclude the outline of malaria epidemiology in this chapter with a brief summary of the main aspects of treatment and control. Naturally these are very specialized subjects requiring the full attention of experts. But in order to gauge the contributions, both actual and potential, of biomathematics to the whole area of malariology, it is necessary to provide at least an introductory outline. As in earlier sections, for full and authoritative statements the reader must consult standard textbooks.

First of all, it may be noted that malaria is an acute and chronic disease, and may be fatal if untreated or treated too late. On the other hand, it can usually be prevented or cured completely by the use of appropriate drugs. Inadequate treatment of certain forms of malaria may however result in surviving persons continuing to have the infection over a period of years. Such chronic cases are liable to suffer from enlarged spleens and debilitating anaemias. This leads to a vicious circle of disease causing social and economic stagnation, which in turn inhibits disease prevention and control.

Several drugs are available for treating individual cases or for mass drug administration. There are, however, a considerable number of complexities. One important distinction, already referred to in Section 2.4, is between *P. falciparum*, and perhaps *P. malariae*, where there is apparently no further development in the liver after the merozoites move out into the blood-stream, and the other two species, *P. vivax* and *P. ovale*, where such development may take place.

The relatively cheap drugs chloroquine and amodiaquine, belonging to the 4-aminoquinoline group of compounds, act directly on certain blood-cell stages of all four species of *Plasmodium*. In the malignant *P. falciparum* this will be highly effective, even though only asexual forms are destroyed, since the cycle of development is interrupted. But in the other three more benign species, while there may be an immediate clinical cure due to the destruction of both sexual and asexual stages in the blood, there is the possibility of relapse due to further development in the liver

which is untouched by the 4-aminoquinolines. For a complete radical cure in these three species additional treatment with other drugs may be required.

Chemoprophylaxis, for example, to protect travellers who are temporarily in malarious areas, is also an important health measure. The drugs mainly used are pyrimethamine and proguanil, in addition to the 4-aminoquinolines mentioned above.

Unfortunately, the whole picture of treatment and prophylaxis is becoming increasingly complicated by the emergence of drug resistance in the parasites, especially to the otherwise cheap and effective 4-aminoquinolines. This phenomenon of drug resistance is now widespread in South-east Asia and South America, and has been observed in at least 20 different countries. It is accepted that considerable care is needed in the treatment of malaria, which must be properly related to the major factors operating, e.g. the species of *Plasmodium* involved, the immune status of the individual, the possibility of clinical complications, the danger of serious side-effects from the drugs employed, the occurrence in many areas of drug resistance, etc.

Apart from attacking the disease in man, there is of course the alternative approach of taking measures directly against the mosquito. These primarily comprise the use of (a) residual insecticides like DDT or BHC, especially where the mosquito enters houses and feeds on man; and (b) antilarval measures, like draining swamps or spraying larvicides on pools of water which may be used as mosquito breeding sites. Unfortunately a serious insecticide resistance problem is also developing here. It has already been observed in 62 countries, out of 107 where malaria exists, that some mosquitoes can tolerate doses of insecticides that would normally be lethal. Since resistance is an inherited characteristic, the genes responsible are likely to spread by the injudicious application of insecticides. New compounds are desperately needed in many places, but not enough has been done to mobilize the available industrial potential for the required research and development.

Much can also be done to reduce contact between man and mosquito through proper siting of houses, use of mosquito-proof screens and netting, the wearing of protective clothing, and the application of insect repellants.

All of these aspects of cure or prevention have to be considered in attempts to bring malaria under control. Obviously complete *eradication* is preferable, but not always feasible. When the transmission of the parasite can be interrupted and the pool of infection eliminated, the

disease will disappear, even if the vector mosquitoes remain. There will, however, still be a danger of the parasites being reintroduced. In favourable circumstances it may be possible to eliminate the vector and prevent its reappearance. Failing these ideal achievements it may be more realistic to settle for the *control* of malaria, in the sense of operating preventive and curative measures in such a way that an acceptably low endemic level of disease is reached. This may entail the maintenance of the control measures in perpetuity. But if serious public health problems can be avoided, and socio-economic development be allowed to progress, then the benefits may easily far outweigh the costs.

It may be useful in this connexion to note that a tripartite classification of the form of malaria prevalence is frequently used. First there is the occurrence of *epidemic* malaria, particularly arising where malaria is imported into a previously non-malarious area, or one from which malaria has been eradicated. Many people, irrespective of age and sex, are likely to be affected. This constitutes an emergency and is normally met, where possible, by the mass administration of drugs to the whole affected population.

Secondly, there is what is called *stable* malaria, corresponding roughly to the holoendemic and hyperendemic conditions described at the end of Section 2.6. Young children are liable to be seriously affected, and the malariological factors involved are likely to be of such a magnitude that full-scale control cannot be achieved. It may, however, be possible to protect the most vulnerable age-groups.

Thirdly, there is the possibility of *unstable* malaria, corresponding approximately to the previous definitions of mesoendemic and hypoendemic areas. The fact of instability may mean that control, properly undertaken, can be achieved relatively easily.

3 Historical perspective

3.1 Short history of malaria

The history of malaria goes back a long way, and indeed there is evidence of even prehistoric man being affected. The disease was first recognizably described by Hippocrates in the fifth century BC. He not only identified clinical details, but was aware of the association between the disease and various environmental and seasonal factors. In particular, he noted a connexion with stagnant water, which we now recognize as implicating the mosquito. Later, an accurate description of the disease was given by Celsus round about AD 30. Recurrent epidemics continued right on through the Middle Ages, especially in Europe. These were widely referred to in the seventeenth and eighteenth centuries as the "ague". The Italian term *mal' aria*, "bad air", appeared in the eighteenth century, reflecting the common belief that the foul air arising from marshes was somehow involved in causing the disease. The French word *paludisme*, also indicating the association between the disease and swamps, was only introduced much later in the 1880's.

It is not clear to what extent a general awareness of the connexion between certain fevers and marshy localities directly encouraged the draining of swamps, as already practised by the Greeks and Romans as early as the sixth century BC: perhaps natural trends in economic and social development automatically led to a certain degree of disease limitation and control, for example, through the elimination of marshes and swamps in an attempt to improve agricultural land. Ross* (1909), however, certainly believed that "The ancients knew that *drainage reduces malaria*", and he also considered that a wide variety of preventive measures had been followed since the earliest times, but *unconsciously*.

The first big success in systematically combating malaria came with the discovery, at the beginning of the seventeenth century, of the use by Peruvian Indians of the bark of a certain tree. This was later named

* It should be noted that this pioneer publication of Ross's seems to have been consistently referred to in the literature by the date of the Preface (1908) and the name of the printer (Waterlow), instead of the more usual date of publication (1909) and name of publisher (Churchill).

6 General theory of host-vector diseases

6.1 The role of general theory

For those who have been reading through this book page by page, the strictures of Section 5.4 on the validation of models at the end of the preceding chapter will immediately throw into sharp relief the whole question of the value of general mathematical theory. I have already briefly discussed this contentious subject elsewhere in relation to infectious disease modelling (see Bailey, 1975a, ch. 12). This was particularly relevant in that book because the whole of Part 2 was devoted to general theory, though Part 3 dealt with specific applications. It was suggested that "a mathematician *qua* mathematician is entitled to devote whatever time he has available to any problems he finds personally interesting. But we should be deluding ourselves if we imagined we were working on the theory of actual diseases in the real world when we were merely absorbed in the abstract intricacies of problems derived from, but not intimately related to, practical needs".

These remarks apply with even more force to any study like the present one, that is explicitly devoted to a specific disease. In fact there are strong arguments for having a well-established theoretical background, but the ultimate justification of the latter depends on the extent to which it finally facilitates understanding and control of the real world.

Individual models that are intended to be finally incorporated into the thought and action of decision-makers must, I believe, pass through the five stages of validation indicated in Section 5.4, or at least something rather similar. However, it is of enormous value for biomathematicians engaged in the first two deductive phases of model construction and validation to be able to draw on a well-founded body of theoretical knowledge. Thus at the very first stage of trying to establish a model in broad qualitative terms, using a flow-chart representation for example, it is extremely helpful to know that certain proposed designs are or are not likely to work, that in some cases a direction proposed on epidemiological grounds will lead to intractable mathematics or excessively pro-

longed and expensive computing, or that some theoretically attractive idea will turn out to be insufficiently realistic.

Such considerations are relevant to an even greater degree at the second, more intensive, phase of theoretical analysis. Fully to work out the logical consequences of the initial assumptions underlying a model may easily involve a great deal of mathematical analysis. The gradual accumulation of knowledge into a body of general theory helps to avoid much duplication of research through existence theorems, threshold theorems, standard solutions to typical equations, knowledge of the number and type of steady states, occurrence or non-occurrence of oscillatory behaviour, etc.

As we shall see later, quantitative modelling can be highly relevant to epidemiological and public health practice, though it must be admitted that malaria models are only just beginning to emerge into the third phase of submission to statistical tests of fit to empirical data. A good deal of value, however, can be gleaned from the study of models that, though hypothetical, are nevertheless oriented towards at least a qualitative understanding of practical issues. The reason for this is that large populations are almost always complex and heterogeneous, and it is frequently impossible to expect any modelling to be achieved that is sufficiently realistic in detail to pass formal goodness-of-fit tests. At the same time, theoretical considerations may lead to practical insights into the mechanisms underlying the observed phenomena. And when this happens there is likely to be some contribution to an improvement in public health control.

In order to undertake the whole range of modelling expeditiously, we thus need a body of knowledge to guide us. We must know what models are already available, what their properties are, how mathematically tractable they are, whether convenient approximations exist, what computational methods are required, what expectations there are of exact mathematical results, whether simulation studies are likely to be useful, how to choose between deterministic and stochastic formulations, whether to include spatial factors, whether the models are reasonably robust to possible oversimplifications in the assumptions, whether commonly held epidemiological ideas are consistent with and follow from the commonsense assumptions usually made, whether modifications proposed in models for greater realism would in fact lead to appreciably different results, etc.

We also expect the general theory to indicate where the main gaps in our knowledge lie, what would be the most fruitful areas for new practice-oriented research, what kind of analytical and computational techniques need to be developed in order to save time and effort, what might be the

consequences for theoretical work of new developments in computer technology, etc. Conversely, we expect some indications of what is *not* worth doing, what would be mathematical refinement purely for its own sake, what kind of theorems would be valid under only very severe restrictions and would have little practical consequence, what sort of statistical data would be inherently uninformative and therefore useless as a basis for further analysis, etc.

No doubt the reader will add further items to the above list according to his inclinations; and may have his own views as to which individual models are realistic and useful, and which are hopelessly oversimplified; or which theorems provide genuine insight, and which are of only academic interest. Nevertheless the broad purposes of the general theory remain clear, and should be borne in mind when we come to the more practically oriented discussions of later chapters.

Next, it should be pointed out that the general theory of host–vector diseases should, by definition, be applicable to any disease involving the host–vector feature, such as plague, typhus, malaria, schistosomiasis, etc. However, this broad applicability is achieved only at a cost of ignoring many specific aspects of individual diseases: for example, the existence of animal reservoirs in plague, or the free-swimming larval stages of schistosomiasis. On the other hand, in some diseases, such as malaria, it is obligatory for the parasite to spend a part of its life-cycle inhabiting the vector, with the infection switching directly back and forth between host and vector. The general theory of the present chapter thus has specific relevance to malaria, though it may achieve no more than a broad qualitative interpretation. In fact many of these models were originally investigated with malaria in mind. When we want to study any given disease in greater depth, more realistic models are needed that incorporate a finer degree of detail. Subsequent chapters will accordingly develop the modelling required for practical epidemiological and public health use.

Finally, it may be observed that the treatment of two populations interacting in the host–vector manner defined above, i.e. with cross-infection between the groups but none within groups, can be regarded as a special case of an even more general theory involving several interacting groups, in which there is both an intra-group as well as an inter-group spread of infection (see Bailey, 1975a, Sections 5.10 and 6.9). However, most of these discussions are either too formal or too algorithmic to provide much insight into the structure of the phenomena concerned. We shall therefore adopt an intermediate position of confining attention in the present chapter to two populations, which are most conveniently thought of as representing the human hosts and the mosquito vector.

6.2 Deterministic epidemics

First, we shall consider the formulation of a basic deterministic model that suffices to describe in broad qualitative terms the occurrence of epidemic outbreaks in typical host–vector situations. Comparatively little work has been done on such models, although Kermack & McKendrick (1927) devoted some discussion to the subject over 50 years ago, and in fact obtained a generalization of their well-known single-population threshold theorem (see Bailey, 1975a, Section 6.2). Kermack & McKendrick also allowed the possibility of a further generalization covering variable infection- and recovery-rates, but this extension seems to involve complications that are unnecessary for our purposes.

We shall suppose that there are two interacting populations. The first representing the human hosts, consists at time t of x susceptibles, y circulating infectious individuals or infectives, and z individuals who have in some sense been removed from the infection process and are isolated, dead or recovered and immune, written as (x, y, z) for short. We shall assume the total population size to be $x + y + z = n$.

The corresponding quantities in the second population, representing the intermediate host or vector, are x', y' and z', with $x' + y' + z' = n'$, or (x', y', z') in compact notation.

Now in order to achieve an approximate resemblance to a simple host–vector mode of transmission, such as we find with malaria, we want the human susceptibles to acquire the infection only from infectious vectors, and conversely. As already mentioned in Section 4.4 in connexion with the spread of transmission within a single population, it is common to assume that the rate at which the available susceptibles produce new infectives is jointly proportional to an infection-rate (or contact-rate) and the numbers of relevant susceptibles and infectives (as in the Law of Mass Action). The assumption is generally regarded as reasonable to epidemiological common sense, and leads to models that have the right kind of behaviour (see Bailey, 1975a). Of course, in its simplest form as stated above, it implies the homogeneous mixing of individuals within the population. (When heterogeneity of mixing is suspected, various modifications may be required.) We therefore suppose that the number of new infections occurring in the human population in time-interval Δt to be $\beta x y' \Delta t$, where β is the infection-rate.

Since we want the converse arrangement to hold for vectors, namely that susceptible vectors are infected by human infectives, we take the number of new infections occurring in susceptible vectors in time-interval Δt to be $\beta' x' y \Delta t$.

Let us, in addition, assume overall removal-rates for the two populations to be γ and γ', respectively, so that the numbers of removals occurring in time Δt are $\gamma y \Delta t$ and $\gamma' y' \Delta t$ for humans and vectors, respectively. Again, there are implications underlying the adoption of such constant removal-rates, but they are probably reasonably satisfactory as indications of average figures in relatively large communities.

Given the above transition-rates for the occurrence of new infectives and removals in the two populations, we can immediately write down the usual differential equations, or equations of motion, for the dynamic process involved as

$$\left.\begin{aligned} \frac{dx}{dt} &= -\beta x y', & \frac{dx'}{dt} &= -\beta' x' y, \\ \frac{dy}{dt} &= \beta x y' - \gamma y, & \frac{dy'}{dt} &= \beta' x' y - \gamma' y', \\ \frac{dz}{dt} &= \gamma y, & \frac{dz'}{dt} &= \gamma' y', \end{aligned}\right\} \tag{6.2.1}$$

since $\Delta x = -\beta x y' \Delta t$, etc., the initial states at $t = 0$ being $(x_0, y_0, 0)$ and $(x'_0, y'_0, 0)$.

One implication about the possibility of a proper epidemic outbreak can be drawn from equations (6.2.1) almost immediately. It is intuitively obvious that the rates of change in the numbers of infectives will have to be initially positive in both populations. Hence we must have

$$\beta x_0 y'_0 > \gamma y_0, \qquad \beta' x'_0 y_0 > \gamma' y'_0.$$

If we can assume that both y_0 and y'_0 are small, we have $x_0 \doteqdot n$ and $x'_0 \doteqdot n'$. Thus

$$\beta n y'_0 > \gamma y_0, \qquad \beta' n' y_0 > \gamma' y'_0,$$

from which it follows, by eliminating y_0, that

$$\beta n y'_0 > \frac{\gamma\gamma' y'_0}{\beta' n'}.$$

Therefore we must have the condition

$$nn' > \frac{\gamma\gamma'}{\beta\beta'} \tag{6.2.2}$$

for an epidemic build-up to occur.

This constitutes a threshold requirement in the initial density of susceptibles for an epidemic outbreak. We have assumed of course that y_0 and y'_0 are both positive. If one is zero, corresponding to the introduction of a small number of infected humans only or infected mosquitoes only, a slightly modified argument is needed. The zero quantity will clearly increase initially (though the non-zero one will decrease), and a sufficiently short time after the start of the process both y and y' will be positive. We can therefore consider the process as starting from this point, and the previous argument applies. A more rigorous discussion goes as follows.

Dividing the first equation on the left of (6.2.1) by the third on the right, yields, after integration,

$$-\log(x/x_0) = \beta z'/\gamma'. \tag{6.2.3}$$

Similarly,

$$-\log(x'/x'_0) = \beta' z/\gamma. \tag{6.2.4}$$

Let us now define the *intensity* of any epidemic that occurs as the proportion of susceptibles that finally contracts the disease. Since all these susceptibles are eventually removed, this means that intensities of i and i' in the first and second populations, respectively, will be given by

$$i = z_\infty/n, \qquad i' = z'_\infty/n'. \tag{6.2.5}$$

It follows that the specifications of the two populations at $t = \infty$ are $(n - ni, 0, ni)$ and $(n' - n'i', 0, n'i')$, provided that we can take y_0 and y'_0 as negligibly small. Substituting this into (6.2.3) and (6.2.4) then gives

$$\left.\begin{aligned} \beta n' i'/\gamma' &= -\log(1 - i), \\ \beta' n i/\gamma &= -\log(1 - i'). \end{aligned}\right\} \tag{6.2.6}$$

If we now expand the logarithmic expressions in (6.2.6), retaining the first two terms only, and multiplying the corresponding sides of the two equations so obtained, we find

$$\frac{\beta\beta' nn' ii'}{\gamma\gamma'} \doteqdot (i + \tfrac{1}{2}i^2)(i' + \tfrac{1}{2}i'^2).$$

After cancellation of the factor ii', rearrangement gives

$$\frac{nn'}{\rho\rho'} - 1 = \tfrac{1}{2}(i + i'), \tag{6.2.7}$$

where we have ignored the second-order term in ii', and have written ρ and ρ' for the two *relative removal-rates* given by

$$\rho = \gamma/\beta, \qquad \rho' = \gamma'/\beta'. \tag{6.2.8}$$

It is evident from (6.2.7) that, since i and i' must both be positive or zero, we must have $nn' > \rho\rho'$ for a true epidemic to occur, confirming the conjecture in (6.2.2). There are, therefore, no separate thresholds for man and vector, but there is a joint threshold for the product, $\rho\rho'$, of the two relative removal-rates.

Let us write π for the product of the number of susceptibles in the two populations, so that $\pi = xx'$. We then have $\pi_0 \doteqdot nn'$, and therefore

$$\begin{aligned} \pi_\infty &= x_\infty x'_\infty \\ &\doteqdot \pi_0(1 - i - i'), \end{aligned} \tag{6.2.9}$$

neglecting the product ii'. Substituting for $i + i'$ from (6.2.7) and writing

$$\pi_0 = \rho\rho' + \epsilon \tag{6.2.10}$$

then gives

$$\pi_0 - \pi_\infty \doteqdot 2\epsilon. \tag{6.2.11}$$

This is a modified form of Kermack & McKendrick's single-population threshold theorem which they obtained for the case of two suitably interacting populations. It shows that if there is an epidemic outbreak the product of the initial numbers of susceptibles is eventually reduced to a value as far below the threshold as it originally was above it, at least to a first approximation.

The exact values of the individual intensities, i and i', are of course given by the solutions of (6.2.6). Approximate values for small epidemics are easily found by solving the equations obtained by retaining only the first two terms of the logarithmic expansions. We find

$$i \doteqdot \frac{2\epsilon}{n(n' + \gamma'/\beta)}, \qquad i' \doteqdot \frac{2\epsilon}{n'(n + \gamma/\beta')}, \tag{6.2.12}$$

where we have made use of the fact that $\rho\rho' \doteqdot nn' - \epsilon$.

Of course, the above treatment is somewhat heuristic in nature, and in any case deals only with first approximations. It would, therefore, be interesting to investigate the situation more exactly along the lines followed by D. G. Kendall in analysing the single population situation (see Bailey, 1975a, Section 6.2).

For the time being, we first note the general feature of the existence of a threshold, which determines whether or not a true epidemic outbreak occurs. This is clearly related to the need for public health authorities to be able to manipulate human and vector densities, and the relevant contact and removal parameters, so as to maintain the populations in sub-threshold conditions.

However, a word of warning is necessary about the interpretation of thresholds, especially in regard to the implications for possible interventions. If the parameters β, β', γ and γ', in the above model are independent of n and n', then we can consider in a straightforward way the effect of changing some of them by public health action so as to ensure that $nn' < \rho\rho'$, or better still, $nn' \ll \rho\rho'$. Thus reducing the size of the vector population n' is always a step in the direction of preventing disease.

But suppose that some of the parameters depend on n or n'. In malaria, for example, according to commonly accepted ideas we may suppose that the mosquito vector exhibits a certain man-biting rate b'. Then in unit time x' susceptible mosquitoes bite $b'x'$ people, of whom $(b'x')(y/n)$ are affected by malaria. Let f be the proportion of the latter who are actually infectious. Then the rate at which newly infected mosquitoes occur is $b'fx'y/n$. In other words, $\beta' = b'f/n$. Similarly, in unit time, y' infected mosquitoes, of whom a proportion f' are infectious, bite $b'y'$ people, of whom $(b'y')(x/n)$ are susceptibles. The rate at which newly infected people occur is thus $b'f'xy'/n$, i.e. $\beta = b'f'/n$.

This is of course a departure from homogeneous mixing as previously conceived. Nevertheless, equations (6.2.1) still hold with the above definitions of β and β'. But the condition $nn' > \rho\rho'$, necessary for a true epidemic to occur, becomes, on substituting for β and β', $nn' > \gamma\gamma' n^2/b'^2 ff'$ or $n'/n > \gamma\gamma'/b'^2 ff'$. Thus it is now the *ratio* of the vector population to the host population that is the critical quantity, and not the product. This threshold result, applicable to malaria, in fact goes back to the original work of Ross (1911), long before the studies of Kermack & McKendrick (1927). But for a more extensive discussion, see Chapter 7.

In order to make the specification of the model a little more realistic, so far as any application to malaria might be concerned, we should have to interpret all mosquito removals as deaths, since immunity to infection by the parasite does not appear to occur and the concept of isolation is not applicable in principle. Similarly, in the human population the mortality from malaria might be relatively low, though morbidity could be high, and effective isolation in tropical contexts might also be low: removals would

thus consist largely of those who had recovered and were immune. However, immunity may in practice be lost by some individuals, though we could assume the rate to be negligibly small during the course of a particular epidemic outbreak. All this shows how careful one has to be about accepting the validity of rather general model formulations. However, with such provisos in mind, one can tentatively pursue the implications of this model.

Suppose we assume, therefore, that the specification of the human population is the same as before, leading to the equations in the left-hand column of (6.2.1). Over the period of any epidemic that might occur it is assumed that the demographic features of birth and death are negligible, and that recovered infectives acquire immunity, at least temporarily.

In the mosquito population, on the other hand, with a relatively short life-cycle we must include the demographic features. Removals involve death only, with no isolation or recovery. Since the female mosquito death only, with no isolation or recovery, we assume that the death-rate γ' appears unaffected by her parasite load, we assume that the death-rate γ' operates equally on susceptibles and infectives. The class of removals z' can thus be ignored, and we can put $x' + y' = n'$, where n' is constant over the time-interval considered. In order to maintain a constant mosquito population we must also have a birth-rate γ' to balance the death-rate. It would be natural to suppose that all newborn mosquitoes were initially susceptibles.

The resultant dynamic equations are then easily seen to be

$$\left.\begin{aligned} \frac{dx}{dt} &= -\beta x y', & \frac{dx'}{dt} &= -\beta' x' y + \gamma' y', \\ \frac{dy}{dt} &= \beta x y' - \gamma y, & \frac{dy'}{dt} &= \beta' x' y - \gamma' y'. \\ \frac{dz}{dt} &= \gamma y, \end{aligned}\right\} \qquad (6.2.13)$$

All these expressions follow immediately from the rates of infection, removal, birth and death, as specified above. It should, however, be remarked that the form of the right-hand side of the first equation in the second column arises as follows. Starting with the exact specification, for the rates of infection, death and birth, in that order, and using $n' = x' + y'$, we have $dx'/dt = -\beta' x' y - \gamma' x' + \gamma' n' = -\beta' x' y + \gamma' y$. We now have two linearly independent equations in the first column, but only one in the second column.

Elementary intuitive consideration of the equations for dy/dt and dy'/dt, on the grounds that a build-up of infectives is initially required in both populations for a proper epidemic outbreak, gives the same threshold result as before, namely that we must have $nn' > \rho\rho$. This follows immediately from the fact that the right-hand sides of the two equations in the second line of (6.2.13) are, algebraically, precisely the same as the corresponding equations in (6.2.1). The resultant inequalities therefore have the same form as before.

A more detailed analysis of these equations might be worth pursuing, to obtain for example an approximate expression for the total size of any epidemic that occurs.

A far more complex model (although the authors refer to it as a "simple model") than any of the above has been developed by Olaofe & Olaofe (1975). This distinguishes five basic groups for human hosts and five analogous groups for mosquito vectors, namely: susceptibles; infected individuals in a latent condition; active infectives; infectives who have recovered and are immune; and those infectives who have died. Apart from the obvious transitions it is also assumed that some individuals will recover directly from the latent stage to the immune state.

The states of the system are specified in terms of 10 variables which refer to the numbers in each group, using either a cumulative function or a density function, as appropriate. The times spent by individuals in the latent and infectious groups are also accounted for explicitly, and upper limits are set to the times that may be spent in these groups. Transitions from one group to another are basically defined in terms of probabilities, and the Reed-Frost formula appears at one point in the expression for the density of newly latent persons. However, in general, the probabilities of the transitions are interpreted in terms of transition-rates, so that an essentially deterministic model is specified.

These assumptions, sketched here only in broad outline, lead to a somewhat elaborate formulation of the dynamics of the system in terms of a set of integro-differential equations. Although rather intractable mathematically, numerical solutions can be found with the aid of a computer. The authors give some results for arbitrarily chosen parametric values and population sizes, and it is gratifying to see the appearance of typical bell-shaped epidemic curves. Further developments, designed to include birth- and death-rates, superinfection, geographical spread, etc., could be readily undertaken. The authors also believe that it would be easy to predict the peak of an epidemic as soon as the latter process was known to have begun. This claim requires some practical substantiation, especially as one would expect any such predictions to be very sensitive to un-

certainty about the values of unknown parameters, quite apart from the large biological variation usually found in real situations.

6.3 Stochastic epidemics

Having looked at the broad structure of the kinds of deterministic models we are led to in a general approach to host–vector diseases, we can next examine the consequences of introducing probability elements. The arguments used at the beginning of the previous section can in fact be generalized rather easily to provide a stochastic formulation, in which the continuous rates of infection and removal are replaced by algebraically similar probabilities of discontinuous transitions taking place in short intervals of time.

We can, accordingly, introduce the random variables $X_1(t)$ and $Y_1(t)$ to represent the numbers of susceptibles and infectives, respectively, for the first population, with corresponding variables $X_2(t)$ and $Y_2(t)$ for the second population. The probability of a new infection occurring in the first population in time Δt is $\beta_1 X_1 Y_2 \Delta t$. When this transition occurs X_1 decreases by one unit and Y_1 increases by one unit. Again, the probability of a removal in Δt is $\gamma_1 Y_1 \Delta t$, with Y_1 decreasing by one unit and X_1 remaining unchanged. Similar considerations apply to the second population, *mutatis mutandis*. We also define the relative removal-rates

$$\rho_1 = \gamma_1/\beta_1, \qquad \rho_2 = \gamma_2/\beta_2, \tag{6.3.1}$$

corresponding to (6.2.8).

Let $p(r_1, s_1; r_2, s_2; t)$ be the probability of there being r_1 susceptibles and s_1 infectives in the first population, and r_2 susceptibles and s_2 infectives in the second population, at time t. We define the probability-generating function as

$$P(x_1, y_1; x_2, y_2; t) = \sum_{r_1, s_1} \sum_{r_2, s_2} p(r_1, s_1; r_2, s_2; t)\, x_1^{r_1} y_1^{s_1} x_2^{r_2} y_2^{s_2}. \tag{6.3.2}$$

Using standard methods (e.g. Bailey, 1964, Section 7.4), we can write down immediately the partial differential equation satisfied by the probability-generating function, namely

$$\begin{aligned} \frac{\partial P}{\partial t} &= \beta_1 y_2 (y_1 - x_1) \frac{\partial^2 P}{\partial x_1 \partial y_2} + \gamma_1 (1 - y_1) \frac{\partial P}{\partial y_1} \\ &\quad + \beta_2 y_1 (y_2 - x_2) \frac{\partial^2 P}{\partial x_2 \partial y_1} + \gamma_2 (1 - y_2) \frac{\partial P}{\partial y_2}, \end{aligned} \tag{6.3.3}$$

with initial condition

$$P(x_1, y_1; x_2, y_2; 0) = x_1^{n_1} y_1^{a_1} x_2^{n_2} y_2^{a_2}, \tag{6.3.4}$$

assuming that the epidemic starts with n_1 susceptibles and a_1 infectives in the first population, and n_2 susceptibles and a_2 infectives in the second population.

So far, no explicit analysis is available of this stochastic formulation, which no doubt presents considerable difficulties. A threshold theorem can, however, be obtained, corresponding to the single population case treated in Section 6.5 of Bailey (1975a). Unfortunately, we cannot as yet achieve the degree of sophistication present in Whittle's single-population theorem, but must confine ourselves to an approximate discussion originally due to Bartlett (1955, p. 129) and extended by him to the present host-and-vector model (Bartlett, 1964, 1966).

According to this approach, we assume that the initial populations of susceptibles, n_1 and n_2, are sufficiently large for the populations of infectives to be approximately subject to a bivariate birth-and-death process. The probability of a new infection occurring in the first population in Δt is now $\beta_1 n_1 Y_2 \Delta t$, and the probability of a removal is $\gamma_1 Y_1 \Delta t$. The corresponding quantities for the second population are $\beta_2 n_2 Y_1 \Delta t$ and $\gamma_2 Y_2 \Delta t$. Even this simplified linear process is not readily soluble, but the extinction phenomena can be treated by viewing the model as a continuous-time branching process. Standard theorems can then be applied (e.g. Sevastyanov, 1951; Mode, 1971).

From the branching process point of view, the "offspring" arising from a single infective individual of the first population in Δt may be classified as follows. Let the numbers of infectives newly formed in the first and second populations be S_1 and S_2, respectively. If the individual is removed, with probability $\gamma_1 \Delta t$, we have $S_1 = 0 = S_2$. If an infection of a susceptible from the second population occurs, with probability $\beta_2 n_2 \Delta t$, we have two individual infectives, one from each population, so that $S_1 = 1 = S_2$. The balance of probability, $1 - (\beta_2 n_2 + \gamma_1)\Delta t$, is allotted to the case when no transition occurs, with the original infective simply remaining, and then $S_1 = 1, S_2 = 0$. We then form the expectation

$$\underset{\Delta t}{E}\, y_1^{S_1} y_2^{S_2} = f_1(y_1, y_2)$$

to give the joint probability-generating function for the numbers of "offspring" in the two populations arising in Δt from one infective of the first population. A similar argument applies to "offspring" of infectives in

the second population. It is easily seen that the functions so defined are explicitly given by

$$\left.\begin{aligned} f_1(y_1, y_2) &= y_1 + \{\gamma_1 - (\beta_2 n_2 + \gamma_1) y_1 + \beta_2 n_2 y_1 y_2\}\,\Delta t, \\ f_2(y_1, y_2) &= y_2 + \{\gamma_2 - (\beta_1 n_1 + \gamma_2) y_2 + \beta_1 n_1 y_1 y_2\}\,\Delta t. \end{aligned}\right\} \tag{6.3.5}$$

We now define the extinction probability $p_1(t)$ to be the probability that the epidemic process is already over at time t, i.e. $s_1(t) = 0 = s_2(t)$, given that the process started at $t = 0$ with $a_1 = 1$ and $a_2 = 0$. Conversely, we define $p_2(t)$ to be the probability of extinction at time t, given the starting point at $a_1 = 0$ and $a_2 = 1$.

Actually, it is quite possible to write down the backward differential equations for p_1 and p_2, using elementary probability considerations. These simultaneous equations can be solved quite easily in the limiting case of $dp_1/dt = 0 = dp_2/dt$, giving two roots for each variable. Unfortunately, this method does not tell us which roots to choose. The branching process theory, on the other hand, specifies that the probabilities of ultimate extinction, $p_1(\infty)$ and $p_2(\infty)$, are given by the *smallest* non-negative solutions of the equations

$$g_1(y_1, y_2) = 0, \qquad g_2(y_1, y_2) = 0, \tag{6.3.6}$$

where

$$f_1 = y_1 + g_1 \Delta t, \qquad f_2 = y_2 + g_2 \Delta t. \tag{6.3.7}$$

Substituting from (6.3.7) into (6.3.5) shows that the equations (6.3.6) are

$$\left.\begin{aligned} \beta_2 n_2 y_1 y_2 - (\beta_2 n_2 + \gamma_1) y_1 + \gamma_1 &= 0, \\ \beta_1 n_1 y_1 y_2 - (\beta_1 n_1 + \gamma_2) y_2 + \gamma_2 &= 0. \end{aligned}\right\} \tag{6.3.8}$$

Explicit solutions are readily found to be

$$\left.\begin{aligned} p_1(\infty) = 1 \quad \text{or} \quad & \frac{\gamma_1(\beta_1 n_1 + \gamma_2)}{\beta_1 n_1 (\beta_2 n_2 + \gamma_1)}; \\ p_2(\infty) = 1 \quad \text{or} \quad & \frac{\gamma_2(\beta_2 n_2 + \gamma_1)}{\beta_2 n_2 (\beta_1 n_1 + \gamma_2)}. \end{aligned}\right\} \tag{6.3.9}$$

It follows from (6.3.9) that for $p_1(\infty) < 1$ we must have $n_1 n_2 > \rho_1 \rho_2$, and the same condition for $p_2(\infty) < 1$. Otherwise we have $p_1(\infty) = 1 = p_2(\infty)$. Thus the chance of ultimate extinction of the process stemming from any given initial infective is unity if $n_1 n_2 \leqslant \rho_1 \rho_2$. And so extinction is certain for any process starting with a_1 infectives in the first population and a_2 infectives in the second population if $n_1 n_2 \leqslant \rho_1 \rho_2$. The threshold value, involving the joint product $\rho_1 \rho_2$ of the relative removal-rates, is thus identical with the deterministic result obtained in the previous section.

When $n_1 n_2 > \rho_1 \rho_2$, on the other hand, the chance of the process not becoming extinct, starting from the initial values (a_1, a_2), is

$$\begin{aligned} & 1 - \left\{\frac{\gamma_1(\beta_1 n_1 + \gamma_2)}{\beta_1 n_1 (\beta_2 n_2 + \gamma_1)}\right\}^{a_1} \left\{\frac{\gamma_2(\beta_2 n_2 + \gamma_1)}{\beta_2 n_2 (\beta_1 n_1 + \gamma_2)}\right\}^{a_2} \\ & \quad = 1 - \left(\frac{\rho_1}{n_1}\right)^{a_1} \left(\frac{\rho_2}{n_2}\right)^{a_2} \left(\frac{\beta_1 n_1 + \gamma_2}{\beta_2 n_2 + \gamma_1}\right)^{a_1 - a_2}. \end{aligned} \tag{6.3.10}$$

This expression is therefore the approximate chance of a major epidemic building up. Note that (6.3.10) cannot be written solely as a function of ρ_1 and ρ_2, or of $\rho_1 \rho_2$; though if we define "cross" relative removal-rates

$$R_1 = \frac{\gamma_1}{\beta_2}, \qquad R_2 = \frac{\gamma_2}{\beta_1}, \tag{6.3.11}$$

then (6.3.10) can be put in the form

$$1 - \left(\frac{R_1}{n_1}\right)^{a_1} \left(\frac{R_2}{n_2}\right)^{a_2} \left(\frac{n_1 + R_2}{n_2 + R_1}\right)^{a_1 - a_2} \tag{6.3.12}$$

which is in terms of R_1 and R_2 only, but not of $R_1 R_2$.

The foregoing discussion probably represents the simplest and most straightforward approach to the stochastic modelling of host-vector diseases. However, it may be of importance for practical applications to look more closely at the departures from homogeneous mixing mentioned in Section 6.2 in connexion with the more detailed modelling of malaria.

In the deterministic case, the basic equations are unchanged in essence, though some redefinition of parameters may be necessary. If total population sizes remain constant, the same kind of adjustments should be valid for the stochastic versions as well. But if the totals vary, some rethinking of the mixing and contact mechanisms may be necessary.

6.4 Small epidemics in large populations

The bivariate birth-and-death approximation used in the previous section to obtain a simple stochastic threshold theorem may also be applied to

study processes which entail only small epidemics in large populations, or perhaps the early stages of major epidemics in large populations. We might, for example, want to examine the consequences of introducing a parasitic disease involving a vector into a region or population previously unaffected.

As already mentioned, this means that we can regard the first population of infectives as subject to a birth-and-death process with probabilities of a birth or death in Δt being $\beta_1 n_1 Y_2 \Delta t$ and $\gamma_1 Y_1 \Delta t$, respectively, the corresponding values for the second population being $\beta_2 n_2 Y_1 \Delta t$ and $\gamma_2 Y_2 \Delta t$. In this case (Griffiths, 1972) the differential equation (6.3.3) reduces to

$$\frac{\partial P}{\partial t} = \{\beta_2 n_2 y_1 (y_2 - 1) + \gamma_1 (1 - y_1)\} \frac{\partial P}{\partial y_1} + \{\beta_1 n_1 y_2 (y_1 - 1) + \gamma_2 (1 - y_2)\} \frac{\partial P}{\partial y_2}, \tag{6.4.1}$$

with initial condition

$$P(y_1, y_2; 0) = y_1^{a_1} y_2^{a_2}. \tag{6.4.2}$$

Although (6.4.1) is not readily soluble, mean values are easily found. If we put $y_1 = e^{\theta_1}$ and $y_2 = e^{\theta_2}$ in (6.4.1), we obtain the following partial differential equation for the moment-generating function M:

$$\frac{\partial M}{\partial t} = \{\beta_2 n_2 (e^{\theta_2} - 1) + \gamma_1 (e^{-\theta_1} - 1)\} \frac{\partial M}{\partial \theta_1} + \{\beta_1 n_1 (e^{\theta_1} - 1) + \gamma_2 (e^{-\theta_2} - 1)\} \frac{\partial M}{\partial \theta_2}. \tag{6.4.3}$$

Alternatively, equation (6.4.3) may be written down directly by the usual standard procedure.

Now let $\mu_1(t)$ and $\mu_2(t)$ be the stochastic means of $Y_1(t)$ and $Y_2(t)$, respectively. Equating coefficients of θ_1 and θ_2 on both sides of (6.4.3) then gives

$$\left.\begin{aligned} \frac{d\mu_1}{dt} &= \beta_1 n_1 \mu_2 - \gamma_1 \mu_1, \\ \frac{d\mu_2}{dt} &= \beta_2 n_2 \mu_1 - \gamma_2 \mu_2. \end{aligned}\right\} \tag{6.4.4}$$

These equations are of course precisely the same as the corresponding deterministic equations for the approximating process being considered, as expected for a linear system.

Now let us assume that extinction is certain, i.e. that $n_1 n_2 \beta_1 \beta_2 \leqslant \gamma_1 \gamma_2$ (see Section 6.3). Eliminating μ_2 from (6.4.4) gives

$$\frac{d^2\mu_1}{dt^2} + (\gamma_1 + \gamma_2) \frac{d\mu_1}{dt} + (\gamma_1 \gamma_2 - n_1 n_2 \beta_1 \beta_2) \mu_1 = 0, \tag{6.4.5}$$

the quantity μ_2 obviously satisfying the same equation. The required solution of (6.4.5) can be written in the form

$$\mu_1(t) = A_1 e^{\lambda_1 t} + A_2 e^{\lambda_2 t}, \tag{6.4.6}$$

where λ_1 and λ_2 ($\lambda_1 \geqslant \lambda_2$, say) are the roots of

$$\lambda^2 + (\gamma_1 + \gamma_2)\lambda + (\gamma_1 \gamma_2 - n_1 n_2 \beta_1 \beta_2) = 0. \tag{6.4.7}$$

Now at $t = 0$ we have

$$\mu_1(0) = a_1, \qquad \mu_2(0) = a_2. \tag{6.4.8}$$

And substituting (6.4.8) in (6.4.4) gives

$$\frac{d\mu_1(0)}{dt} = \beta_1 n_1 a_2 - \gamma_1 a_1. \tag{6.4.9}$$

The initial conditions (6.4.8) and (6.4.9) enable us to calculate the constants A_1 and A_2 in (6.4.6). We easily find

$$A_1 = \frac{-(\lambda_2 + \gamma_1) a_1 + \beta_1 n_1 a_2}{\lambda_1 - \lambda_2}, \qquad A_2 = \frac{(\lambda_1 + \gamma_1) a_1 - \beta_1 n_1 a_2}{\lambda_1 - \lambda_2}. \tag{6.4.10}$$

The stochastic mean $\mu_1(t)$ is thus expressed by (6.4.6), where the constants A_1 and A_2 are defined in (6.4.10), and λ_1 and λ_2 are the roots of (6.4.7). The corresponding expression for $\mu_2(t)$ is obviously derived simply by interchanging the suffixes 1 and 2.

The average total epidemic sizes can also be obtained fairly easily. In stochastic terms we introduce random variables $Z_1(t)$ and $Z_2(t)$ to represent the total number of removals in each of the two populations at time t. It is then quite straightforward to write down an extension of (6.4.3) for the joint moment-generating function of Y_1, Y_2, Z_1 and Z_2. Equating the coefficients of the linear terms in the dummy variables then yields two equations in addition to (6.4.4). If we write $\nu_1(t)$ and $\nu_2(t)$

for the stochastic mean numbers of removals, we find

$$\frac{d\nu_1}{dt} = \gamma_1\mu_1, \qquad \frac{d\nu_2}{dt} = \gamma_2\mu_2. \tag{6.4.11}$$

Once again, these are identical with the deterministic values and could be written down immediately since the process is still linear.

We quickly find

$$\begin{aligned}\nu_1(\infty) &= \gamma_1 \int_0^\infty \mu_1(t)\,dt \\ &= -\frac{\gamma_1(A_1\lambda_2 + A_2\lambda_1)}{\lambda_1\lambda_2}, \quad \text{using (6.4.6)}, \\ &= \frac{\gamma_1\gamma_2 a_1 + n_1\beta_1\gamma_1 a_2}{\gamma_1\gamma_2 - n_1 n_2\beta_1\beta_2},\end{aligned} \tag{6.4.12}$$

using (6.4.10) and substituting the usual expressions for $\lambda_1 + \lambda_2$ and $\lambda_1\lambda_2$ derived from (6.4.7).

Let us define

$$\phi_1 = \frac{\gamma_1}{n_2\beta_2}, \qquad \phi_2 = \frac{\gamma_2}{n_1\beta_1}. \tag{6.4.13}$$

Substitution in (6.4.12) gives

$$\nu_1(\infty) = \frac{\phi_1\phi_2 a_1 + \phi_1 a_2}{\phi_1\phi_2 - 1}. \tag{6.4.14}$$

Similarly,

$$\nu_2(\infty) = \frac{\phi_2 a_1 + \phi_1\phi_2 a_2}{\phi_1\phi_2 - 1}. \tag{6.4.15}$$

Thus the mean epidemic sizes w_1 and w_2, not counting the initial numbers of infectives a_1 and a_2, are given by

$$\left.\begin{aligned} w_1 &= \nu_1(\infty) - a_1 = \frac{a_1 + \phi_1 a_2}{\phi_1\phi_2 - 1}, \\ w_2 &= \nu_2(\infty) - a_2 = \frac{\phi_2 a_1 + a_2}{\phi_1\phi_2 - 1}. \end{aligned}\right\} \tag{6.4.16}$$

These values correspond with those given by Griffiths (1972, equation (22)) for the special case $a_1 = 1, a_2 = 0$.

Actually, Griffiths gives a more extensive discussion of the whole joint probability-generating function of the cumulative numbers of removals to extinction, i.e. the final epidemic sizes, by using the following device. Although a continuous-time bivariate birth-and-death process is primarily under consideration, the final epidemic states can be regarded as resulting from an associated discrete-time branching process. The latter thus implies a sequence of discrete generations, although the members of each generation may be born or die at any "time" point (i.e. measured on a continuous scale).

When, alternatively, $n_1 n_2\beta_1\beta_2 > \gamma_1\gamma_2$, extinction is not certain. There is then no overall limiting distribution of epidemic sizes, though, as pointed out by Griffiths, we may be interested in those realizations that do become extinct. Applying the results of Section 5 of Waugh (1958), Griffiths shows that the instantaneous transition probabilities of the given process *conditional upon extinction* are obtained by replacing the rates $n_1\beta_1$, $n_2\beta_2$, γ_1 and γ_2, by $\gamma_1\omega$, γ_2/ω, $n_1\beta_1/\omega$ and $n_2\beta_2\omega$, respectively, where

$$\omega = \frac{n_1\beta_1 + \gamma_2}{n_2\beta_2 + \gamma_1}. \tag{6.4.17}$$

With these substitutions, therefore, we can apply the foregoing results to epidemic processes that do not become extinct, at least so far as the behaviour of only minor outbreaks is concerned.

Further discussion of some of these problems is given by Griffiths (1973) in his development of multivariate birth-and-death processes as large population approximations to epidemic processes, using especially branching process representations.

As in the two previous sections, we might find it necessary in certain more specific applications to redefine the parameters of the above model so as to allow for some kinds of nonhomogeneous mixing.

6.5 Spatial spread of an epidemic

So far we have been able to ignore the implications of the geographical distribution of populations by assuming homogeneous mixing, or some modification of it. This assumption is quite a reasonable approximation for many diseases occurring in small groups, e.g. the occurrence of measles in families or school classrooms. It may even be adequate for small villages.

But when one considers larger populations spread out over the countryside, it is clear that a newly introduced source of infection often tends to produce its initial effect through a gradual spread from a focus. Susceptibles who are situated a long way from this focus evidently have a smaller chance of catching the infection than those who are in closer contact with the disease.

Of course, the situation is complicated in practice by the existence of large towns, which though separated may have an appreciable volume of traffic between them. This suggests the possibility of taking large towns as the basic geographical units, having homogeneous mixing within towns, but involving contacts between towns related to the known migration patterns. Some theoretical work for single-population diseases has been done on the basis of postulating several interrelated groups (see Bailey, 1975a, Section 6.9). A good deal of effort of a similar kind has also been made in recent years in the USSR to study the spread of influenza, using transportation records to estimate the densities of inter-city traffic (see Bailey, 1975a, ch. 19). This has led to the development of methods for predicting the spread of influenza from one city to another.

If, however, we want an approach, perhaps more appropriate to malaria, that will deal with a population diffused in a large number of small villages, rather than a small number of large towns, then we must incorporate a two-dimensional distribution of susceptibles, infectives and removed individuals, for both hosts and vectors, directly into our model.

Bartlett (1956) originally did this for recurrent epidemics of single population diseases with person-to-person transmission, by considering first a deterministic model for two distinct groups and then immediately generalizing to a spatially continuous system. A particular application showed how a disease, freshly introduced into a susceptible area, would give rise to a wave of infection travelling out from the focus. Bartlett also indicated how to develop a stochastic version of this model using a probability-generating *functional* to handle the position-dependent Markov process, the mathematical complexities becoming somewhat formidable. For a summary of Bartlett's work see Chapter 9 of Bailey (1975a).

More recently, Radcliffe (1973) has developed a generalization of Bartlett's approach to cover the initial geographical spread of host-vector epidemics. He starts by proceeding directly to the stochastic formulation and develops the algebraically complex theory in some detail. Eventually, an examination of the mean functionals demonstrates, subject to certain simplifying assumptions, that if the process once started from a focus does

not die out we again obtain a spreading wave of infection. Radcliffe then briefly introduces the deterministic version of the model and shows that this has the same general behaviour as that predicted by the mean functionals in the full stochastic model.

For purposes of exposition here we shall simplify the treatment by dealing only with the deterministic model in its own right, and shall apply the method used by Bartlett (1956) and extended by Radcliffe (1973) to study the spread of disease from a single focus. For a further description of this method applied to the single population case, see Section 9.2 of Bailey (1975a).

Let us therefore return to the deterministic specification of Section 6.2, in particular equations (6.2.1). Since we are now going to deal with the limiting case of spatially continuous distributions of hosts and vectors, we must redefine x, x', y and y' to represent spatial densities. The spatial co-ordinates themselves will be represented by (ξ, η).

We shall use β and β' for the corresponding infection-rates, but shall assume that a susceptible host at any point is liable to be infected by infective vectors only in the relatively immediate neighbourhood, and conversely for susceptible vectors being infected by infective hosts. We assume therefore that the spatial influence of such infectives is both local and isotropic. This leads to the introduction of the standard operator,

$$\nabla^2 \equiv \frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \eta^2}, \tag{6.5.1}$$

in equations (6.5.2) below with multiplying factors α and α' for the host and vector populations, respectively. The removal-rates are similarly defined as γ and γ', but require no spatial feature.

It is also easy to introduce a certain degree of migration amongst the infectives of both host and vector populations, again supposing this to be purely local and isotropic, and requiring parameters ϕ and ϕ' for the two populations, respectively. Possible movements amongst susceptibles are ignored.

The appropriate analogue of the second line of equations (6.2.1), relating specifically to infective hosts and infective vectors, is then easily seen to be

$$\left.\begin{aligned} \frac{\partial y}{\partial t} &= \beta x(y' + \alpha' \nabla^2 y') - \gamma y + \phi \nabla^2 y, \\ \frac{\partial y'}{\partial t} &= \beta' x'(y + \alpha \nabla^2 y) - \gamma' y' + \phi' \nabla^2 y'. \end{aligned}\right\} \tag{6.5.2}$$

This is the obvious extension to the host and vector situation of the second line in equation (9.1) in Bailey (1975a) which relates only to a single population.

Since we propose to look only at the initial stages of the spread of infection from a given focus, we can suppose that the densities of the susceptibles remain, at least approximately, unchanged. If, therefore, we assume in addition that the populations of hosts and vectors are uniformly distributed we can write $x = n$, $x' = n'$. Substituting this restriction in (6.5.2) gives

$$\left.\begin{aligned} \frac{\partial y}{\partial t} &= \beta n(y' + \alpha'\nabla^2 y') - \gamma y + \phi\nabla^2 y, \\ \frac{\partial y'}{\partial t} &= \beta' n'(y + \alpha\nabla^2 y) - \gamma' y' + \phi'\nabla^2 y' \end{aligned}\right\} \tag{6.5.3}$$

These equations are linear in y and y' and hold some promise of solution. It turns out, however, that the migration terms give rise to difficulties. Suppose we then put $\phi = 0 = \phi'$, to give

$$\left.\begin{aligned} \frac{\partial y}{\partial t} &= \beta n(y' + \alpha'\nabla^2 y') - \gamma y, \\ \frac{\partial y'}{\partial t} &= \beta' n'(y + \alpha\nabla^2 y) - \gamma' y' \end{aligned}\right\} \tag{6.5.4}$$

which retains the essential characteristics of infection and removal, including a local and isotropic spatially distributed influence for the interaction between any given susceptible host and nearby infective vectors, and conversely.

The simplest way to eliminate the otherwise awkward terms in $\nabla^2 y$ and $\nabla^2 y'$ is to use a bivariate Fourier transform with respect to the spatial co-ordinates, e.g.

$$g(p, q) = \int_{-\infty}^{\infty}\int f(\xi, \eta) \exp(ip\xi + iq\eta)\, d\xi\, d\eta, \tag{6.5.5}$$

with inverse

$$f(\xi, \eta) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int g(p, q) \exp(-ip\xi - iq\eta)\, dp\, dq. \tag{6.5.6}$$

Let M and M' be the Fourier transforms of y and y', respectively. Then multiplying both the first and second lines of (6.5.4) by $\exp(ip\xi + iq\eta)$, and integrating with respect to ξ and η, gives

$$\left.\begin{aligned} \frac{\partial M}{\partial t} &= \beta n\{1 - \alpha'(p^2 + q^2)\} M' - \gamma M, \\ \frac{\partial M'}{\partial t} &= \beta' n'\{1 - \alpha(p^2 + q^2)\} M - \gamma' M'. \end{aligned}\right\} \tag{6.5.7}$$

This is a straightforward pair of simultaneous linear differential equations with solutions of the form

$$\left.\begin{aligned} M &= a\, e^{\lambda_1 t} + b\, e^{\lambda_2 t}, \\ M' &= a'\, e^{\lambda_1 t} + b'\, e^{\lambda_2 t}, \end{aligned}\right\} \tag{6.5.8}$$

where a, a', b and b' are arbitrary constants depending on the initial conditions, and the eigenvalues λ_1 and λ_2 are given by the roots of the quadratic

$$\lambda^2 - (\gamma + \gamma')\lambda + \{(\gamma\gamma' - \beta\beta' nn') + \beta\beta' nn'(\alpha + \alpha')(p^2 + q^2)\} = 0, \tag{6.5.9}$$

retaining only linear terms in α and α' which we can suppose to be small. The two roots, to first order in α and α', are easily found to be

$$\left.\begin{aligned} \lambda_1 &= \tfrac{1}{2}(\omega - \gamma - \gamma') - \theta(p^2 + q^2), \\ \lambda_2 &= -\tfrac{1}{2}(\omega + \gamma + \gamma') + \theta(p^2 + q^2), \end{aligned}\right\} \tag{6.5.10}$$

where

$$\omega^2 = (\gamma - \gamma')^2 + 4\beta\beta' nn', \qquad \theta = \beta\beta' nn'(\alpha + \alpha')/\omega. \tag{6.5.11}$$

We are now in a position to operate on M and M' in (6.5.8), using the inverse Fourier transformation given by (6.5.6). Terms involving $e^{\lambda_2 t}$ lead to

$$(2\theta t)^{-1} \exp\left\{-\tfrac{1}{2}(\omega + \gamma + \gamma')\, t + \frac{\xi^2 + \eta^2}{4\theta t}\right\},$$

which tends to zero at $t \to \infty$ and thus represents only transient effects. On the other hand, terms in $e^{\lambda_1 t}$ transform back to

$$(2\theta t)^{-1} \exp\left\{\tfrac{1}{2}(\omega - \gamma - \gamma')\, t - \frac{\xi^2 + \eta^2}{4\theta t}\right\},$$

which is clearly not transient if $\omega > \gamma + \gamma'$, i.e. if

$$nn' > \frac{\gamma\gamma'}{\beta\beta'},$$

the condition already obtained in Section 6.2 for a major epidemic to build up.

Consider first the human hosts. The non-transient part of the solution is

$$y = \left(\frac{a}{2\theta t}\right) \exp\left\{\tfrac{1}{2}(\omega - \gamma - \gamma')\, t - \frac{\xi^2 + \eta^2}{4\theta t}\right\}. \tag{6.5.12}$$

If y_R is the total amount of infection *outside* a circle of radius R, we have

$$y_R = \int_{\xi^2 + \eta^2 > R^2} y\, d\xi\, d\eta$$

$$= (2\pi a) \exp\left\{\tfrac{1}{2}(\omega - \gamma - \gamma')\, t - \frac{R^2}{4\theta t}\right\}. \tag{6.5.13}$$

It follows from (6.5.13) that

$$R = \{2\theta(\omega - \gamma - \gamma')\}^{1/2}\, t \left\{1 - \frac{2\log(y_R/2\pi a)}{(\omega - \gamma - \gamma')\, t}\right\}^{1/2}. \tag{6.5.14}$$

Thus, as $t \to \infty$, we have $R \to \{2\theta(\omega - \gamma - \gamma')\}^{1/2} t$. For any given value of y_R, the circle of radius R grows at an approximately constant rate R/t when t is large. This rate is the *velocity of propagation*, and has the limiting value

$$\left[2\beta\beta' nn'(\alpha + \alpha') \left\{1 - \frac{\gamma + \gamma'}{\{(\gamma - \gamma')^2 + 4\beta\beta' nn'\}^{1/2}}\right\}\right]^{1/2}. \tag{6.5.15}$$

For the insect vectors the argument is essentially the same, except that the arbitrary constant a is replaced by a'. With this modification, (6.5.14) still holds and leads to the same limiting value of the velocity of propagation shown in (6.5.15).

We can expect that this kind of discussion provides some insight into the early stages of epidemic spread from a focus, although not too much emphasis should be laid on it, partly because of the initial simplifying assumptions in the model and partly because of the additional approximations made later in the cause of mathematical tractability.

The more complex stochastic treatment by Radcliffe (1973) is of course more realistic, and allows the calculation not only of mean densities but also of higher-order product densities. But here again considerable simplification has to be used in order to achieve any explicit results. Thus at present, the expressions for mean values of infective densities for both hosts and vectors depend, in deterministic and stochastic treatments alike, on assuming only local infection and no migration of susceptibles or infectives. In applications to malaria this may be too restrictive, at least so far as the mosquito hosts are concerned.

While the above discussion applies only to the opening stages of epidemic spread from a focus, we should like to know under what circumstances we may expect the disease to become generally rife over the whole plane, i.e. to result in a *pandemic*. In fact a pandemic threshold theorem was originally derived by D. G. Kendall (in the discussion on Bartlett, 1957) for the single population case, involving an infinite uniform two-dimensional distribution of individuals over the plane. Radcliffe (1976) has shown how Kendall's treatment may be generalized to the host and vector situation, though he rightly points out that the model is more appropriate to a viral disease with immunity in both hosts and vectors. We might, however, take the assumptions to be approximately valid, as already indicated in Section 6.2, by supposing immunity to be conferred on the human hosts at least temporarily during the spread of the epidemic, while interpreting all vector removals as deaths.

Suppose we therefore consider a generalization of (6.2.1) in which x, y, z, x', y' and z' are defined as proportions satisfying

$$x + y + z = 1, \qquad x' + y' + z' = 1, \tag{6.5.16}$$

with the actual densities of human and vector populations being σ and σ' individuals per unit area, respectively. We can thus take the numbers of individuals in an area dS surrounding a point P who are susceptible, infected or removed as $\sigma x\, dS$, $\sigma y\, dS$ and $\sigma z\, dS$, respectively, for the human population, and $\sigma' x'\, dS$, $\sigma' y'\, dS$ and $\sigma' z'\, dS$, respectively, for the vector population.

We also define the spatially weighted averages, $\tilde{y}$ and $\tilde{y}'$, defined by

$$\left.\begin{aligned} \tilde{y}(P, t) &= \iint \lambda(PQ)\, y(Q, t)\, dS \\ \text{and} \\ \tilde{y}'(P, t) &= \iint \lambda'(PQ)\, y'(Q, t)\, dS, \end{aligned}\right\} \tag{6.5.17}$$

where $\lambda(PQ)$ and $\lambda'(PQ)$ are suitably chosen non-negative weighting coefficients satisfying

$$\iint \lambda(PQ)\,dS = 1 = \iint \lambda'(PQ)\,dS. \tag{6.5.18}$$

Equations (6.2.1) are then replaced by

$$\left.\begin{aligned} \frac{\partial x}{\partial t} &= -\beta\sigma x\tilde{y}', & \frac{\partial x'}{\partial t} &= -\beta'\sigma' x'\tilde{y}, \\ \frac{\partial y}{\partial t} &= \beta\sigma x\tilde{y}' - \gamma y, & \frac{\partial y'}{\partial t} &= \beta'\sigma' x'\tilde{y} - \gamma' y', \\ \frac{\partial z}{\partial t} &= \gamma y, & \frac{\partial z'}{\partial t} &= \gamma' y'. \end{aligned}\right\} \tag{6.5.19}$$

In this specification we are using a simple extension of Kendall's original notation, though Radcliffe (1976) employs a somewhat more sophisticated version.

We next define spatially weighted averages, $\tilde{z}$ and $\tilde{z}'$, in the same way as (6.5.17). Multiplying the two equations in the third line of (6.5.19) by $\lambda(PQ)$ and $\lambda'(PQ)$ respectively, and integrating over the plane, gives

$$\frac{\partial \tilde{z}}{\partial t} = \gamma\tilde{y}, \qquad \frac{\partial \tilde{z}'}{\partial t} = \gamma'\tilde{y}'. \tag{6.5.20}$$

If we now divide the first equation on the left of (6.5.19) by the second equation on the right of (6.5.20), and conversely, we obtain

$$dx/d\tilde{z}' = -\beta\sigma x/\gamma', \qquad dx'/d\tilde{z} = -\beta'\sigma' x'/\gamma, \tag{6.5.21}$$

these equations being seen as applying to a *fixed point P*.

Let the initial conditions be

$$\left.\begin{aligned} &x(P,0) = 1-\epsilon(P),\ y(P,0) = \epsilon(P),\ z(P,0) = 0, \\ &x'(P,0) = 1-\epsilon'(P),\ y'(P,0) = \epsilon'(P),\ z'(P,0) = 0. \end{aligned}\right\} \tag{6.5.22}$$

Equations (6.5.21) can therefore be integrated to give

$$\left.\begin{aligned} x &= (1-\epsilon)\exp(-\beta\sigma\tilde{z}'/\gamma'), \\ x' &= (1-\epsilon')\exp(-\beta'\sigma'\tilde{z}/\gamma). \end{aligned}\right\} \tag{6.5.23}$$

We can now use (6.5.23) and (6.5.16) to eliminate x and x' from the two

equations in the third line of (6.5.19), giving

$$\left.\begin{aligned} \frac{\partial z}{\partial t} &= \gamma\{1 - z - (1-\epsilon)\exp(-\beta\sigma\tilde{z}'/\gamma')\}, \\ \frac{\partial z'}{\partial t} &= \gamma'\{1 - z' - (1-\epsilon')\exp(-\beta'\sigma'\tilde{z}/\gamma)\}. \end{aligned}\right\} \tag{6.5.24}$$

These two equations are sufficient, with the initial conditions for z and z' in (6.5.22), to determine z and z' for all P and t.

Let us now specialize the initial conditions in (6.5.22) to involve a focus of infection introduced by a small source of infected human hosts. We can suppose this initial infection to be uniformly spread over a circle, centred on the original O, and having radius a, so that

$$\left.\begin{aligned} \epsilon(P) \equiv \alpha > 0, \qquad & OP < a, \\ \equiv 0, \qquad & OP > a, \end{aligned}\right\} \tag{6.5.25}$$

while

$$\epsilon'(P) \equiv 0. \tag{6.5.26}$$

(We could, if we wished, make converse assumptions for an outbreak started by the introduction of infected vectors.) It is also convenient to impose some restrictions on the weighting factors λ and λ', e.g.

$$\left.\begin{aligned} \lambda(PQ) > 0, \qquad & 0 \leqslant PQ \leqslant b, \\ = 0, \qquad & PQ > b; \\ \lambda'(PQ) > 0, \qquad & 0 \leqslant PQ \leqslant b', \\ = 0, \qquad & PQ > b'. \end{aligned}\right\} \tag{6.5.27}$$

It can be seen from (6.5.19) that as $t \to \infty$, we must have x and x' decreasing, with z and z' increasing. Since these four quantities, being proportions, are bounded, they must have limits given by X, X', Z and Z', respectively. Further, it can be seen that the limits Y and Y' of y and y', respectively, must be zero. Thus $X = 1 - Z$ and $X' = 1 - Z'$. The ultimate behaviour of the epidemic must therefore be determined by $Z(P)$ and $Z'(P)$.

From (6.5.24) we immediately obtain the limiting values of z and z', as $t \to \infty$, given by

$$\left.\begin{aligned} Z &= \alpha + (1-\alpha)\{1 - \exp(-\beta\sigma\tilde{Z}'/\gamma')\}, \qquad && OP < a, \\ &= 1 - \exp(-\beta\sigma\tilde{Z}'/\gamma'), && OP > a, \end{aligned}\right\} \tag{6.5.28}$$

and

$$Z' = 1 - \exp(-\beta'\sigma'\tilde{Z}/\gamma), \tag{6.5.29}$$

using (6.5.25) and (6.5.26), and where $\tilde{Z}$ and $\tilde{Z}'$ are spatially weighted averages of Z and Z'. It is clear that $Z(P) > 0$ if $OP < a$, and we can show that the inequality also holds if $OP > a$. For if A is a point such that $Z(A) = 0$, it follows from the second equation of (6.5.28) that $\tilde{Z}'(A) = 0$. The condition on λ' in (6.5.27) then implies that $Z'(P) = 0$ almost everywhere inside a circle with centre A and radius b', and wherever $Z'(P) = 0$ equation (6.5.29) requires that $\tilde{Z}(P) = 0$. Using the condition on λ in (6.5.27) then defines a circle with centre A and a radius of at least b, within which $Z(P) = 0$ almost everywhere. Continuing this argument in the direction AO leads to a contradiction in a finite number of steps. Hence $Z(P) > 0$ *everywhere*. Repeated use of (6.5.28) easily shows that we also have $Z'(P) > 0$ *everywhere*. This yields the important result that the effect of the spreading epidemic will ultimately be felt over the whole plane.

If, again following Kendall, we conjecture that $Z(P)$ and $Z'(P)$ decrease steadily as $OP \to \infty$, at least for P sufficiently distant from the origin, it follows that $Z(P) \to \zeta \geqslant 0$ and $Z'(P) \to \zeta' \geqslant 0$, where

$$\left.\begin{aligned} \zeta &= 1 - \exp(-\beta\sigma\zeta'/\gamma'), \\ \zeta' &= 1 - \exp(-\beta'\sigma'\zeta/\gamma). \end{aligned}\right\} \tag{6.5.30}$$

These equations always have the solution (0, 0), but if at the origin in the (ζ, ζ')-plane the slope of the tangent to the second curve in (6.5.30) is greater than the slope of the tangent to the first curve, i.e. if $\beta'\sigma'/\gamma > \gamma'/\beta\sigma$ or $\sigma\sigma' > \gamma\gamma'/\beta\beta'$, there will be a unique positive solution (ζ, ζ'). In terms of the relative removal-rates defined in (6.2.8), the required condition is simply $\sigma\sigma' > \rho\rho'$. Thus we have a host-vector analogue of Kendall's pandemic threshold theorem, stating that

(i) there will be a pandemic if and only if the product of the two initial population densities $\sigma\sigma'$ *exceeds* the threshold given by the product of the two relative removal-rates $\rho\rho'$;

(ii) when there *is* a pandemic the severities, ζ and ζ', are given by the unique non-zero solutions of (6.5.30).

On the whole, any attempt to cope with the spatial spread of disease inevitably leads to considerable complexity. There is, however, a growing volume of literature incorporating the spatial aspect, though most of it deals with person-to-person transmission as opposed to vector-borne

disease. For a general review of this subject, see Bailey (1980a). It may be that some of the ideas developed in the context of simpler infectious diseases could be usefully extended to cover parasitic situations as well, and malaria in particular.

A few papers have appeared recently dealing with rather abstract, but generalized, approaches to vector-borne diseases with spatial spread, such as Cooke (1979) and Marcati & Pozio (1980). The latter authors consider, in particular, that their model could apply to the growth and spread of malaria. They show that, under certain specified conditions, the disease would die out if the recovery rate were less than equal to threshold value; otherwise the disease would spread until a homogeneous distribution of infected humans was achieved.

6.6 Endemic models

In the principal deterministic epidemic model discussed in Section 6.2 a major epidemic outbreak can only occur if the product of the initial numbers of susceptibles in the host and vector populations is above the threshold value defined by the product of the two relative-removal rates. Such an epidemic is characterized by an incidence-rate of new infections that first rises to a peak, and then falls away again asymptotically to zero, as the stock of susceptibles is rapidly depleted to a point at which the outbreak is virtually over. The properties of the analogous stochastic model are a little more complicated, but in essence involve the occurrence of a major outbreak only with a certain probability when the threshold is surpassed.

The partial exhaustion of the stock of susceptibles appears to be an essential factor in the occurrence of a true epidemic, i.e. an outbreak of limited extent. As soon as we allow the susceptible group to be replenished, either through the admission of new births, all considered to be initially susceptible, or by all or some removed individuals being transferred to the susceptible class, then we are likely to find at least one non-null stable equilibrium level of infectives to which the system will tend if an appropriate threshold condition is fulfilled. When such an equilibrium is realized in practice, at least approximately, we say that the disease in question is *endemic*, reserving the title of *epidemic* to outbreaks of limited duration.

The acquisition and loss of immunity to malaria are actually very complex phenomena (see end of Section 2.5), and we shall discuss these matters in more detail when we come to investigate models in Chapters 7

and 8 that attempt to deal more realistically with various aspects of the finer structure of the infection processes involved. In the meantime we merely note the slightly unsatisfactory situation in which contradictory assumptions are made in more general theory according to convenience. Thus in Sections 6.2 to 6.5, dealing with epidemic phenomena, the removal of human infectives is assumed to entail immunity, at least temporarily during the course of a given epidemic outbreak, in order to avoid replenishing the stock of susceptibles and permitting a stable level of disease. On the other hand, some of the more simplified treatments of endemic situations assume that there is no substantial development of immunity, and that removed infectives are in effect cured of the disease and are returned to the group of susceptibles. Such assumptions certainly make for mathematical tractability in analysing the equilibrium states and their associated stabilities.

In order to cover a range of alternative possibilities, let us extend the equations in (6.2.13) to include, for the human population, both the demographic aspects of birth and death, as well as the possibility of immunity developing in some infected individuals but not in others. We shall retain, as sufficiently realistic, the previous dynamics of the mosquito population, involving only a constant death-rate γ' for both susceptibles and infectives balanced by a similar birth-rate producing only susceptibles.

Let us assume therefore that all three human groups, of susceptibles, infectives and removals, are subject to the same death-rate λ, balanced by a birth-rate λ which produces only susceptibles. In addition, we can assume that infectives and removals are transferred back to the susceptible class at rates μ and ν, respectively. The relevant equations are easily seen to be

$$\left.\begin{aligned} &\frac{dx}{dt} = -\beta x y' + (\lambda+\mu)y + (\lambda+\nu)z, && \frac{dx'}{dt} = -\beta' x' y + \gamma' y', \\ &\frac{dy}{dt} = \beta x y' - (\gamma+\lambda+\mu)y, && \frac{dy'}{dt} = \beta' x' y - \gamma' y'. \\ &\frac{dz}{dt} = \gamma y - (\lambda+\nu)z, \end{aligned}\right\} \quad (6.6.1)$$

This system of equations accordingly allows a variety of assumptions about immunity to be included, with special simplified cases to be represented by putting some of the relevant parameters γ, μ and ν equal to zero. As before, there are two linearly independent equations in the first column, but only one in the second.

Equilibrium states of the system in (6.6.1) are given in the usual way by setting the differential coefficients equal to zero. Three independent equations are

$$\left.\begin{aligned} \beta x y' &= (\lambda+\mu)y + (\lambda+\nu)z, \\ \beta x y' &= (\gamma+\lambda+\mu)y, \\ \beta' x' y &= \gamma' y'. \end{aligned}\right\} \quad (6.6.2)$$

Substituting $z = n - x - y$ and $x' = n' - y'$ in (6.6.2) then gives the following three equations in x, y and y':

$$\left.\begin{aligned} \beta x y' &= (\lambda+\nu)(n-x) + (\mu-\nu)y, \\ \beta x y' &= (\gamma+\lambda+\mu)y, \\ \beta'(n'-y')y &= \gamma' y'. \end{aligned}\right\} \quad (6.6.3)$$

Although these equations have nonlinear components, explicit solutions can be found fairly easily. Thus eliminating $\beta x y'$ from the first two equations gives one linear equation in x and y. Again, rewriting the last equation as $\beta' n' y = y'(\beta' y + \gamma')$ and multiplying corresponding sides of this equation and the second line of (6.6.3), allows the product yy' to be cancelled to yield a second linear equation in x and y. The solutions of the pair of linear equations are easily found to be:

$$\left.\begin{aligned} x &= \frac{(\gamma+\lambda+\mu)\{n\beta'(\lambda+\nu) + \gamma'(\gamma+\lambda+\nu)\}}{\beta'\{n'\beta(\gamma+\lambda+\nu) + (\lambda+\nu)(\gamma+\lambda+\mu)\}}, \\ y &= \frac{(\lambda+\nu)\{nn'\beta\beta' - \gamma'(\gamma+\lambda+\mu)\}}{\beta'\{n'\beta(\gamma+\lambda+\nu) + (\lambda+\nu)(\gamma+\lambda+\mu)\}}, \\ y' &= \frac{(\lambda+\nu)\{nn'\beta\beta' - \gamma'(\gamma+\lambda+\mu)\}}{\beta\{n\beta'(\lambda+\nu) + \gamma'(\gamma+\lambda+\nu)\}}, \end{aligned}\right\} \quad (6.6.4)$$

to which we can add, for completeness, the null solution

$$x = y = y' = 0. \quad (6.6.5)$$

It is clear from (6.6.4) that there can be a non-null endemic level only if

$$nn' > \frac{\gamma'(\gamma+\lambda+\mu)}{\beta\beta'}. \quad (6.6.6)$$

This corresponds closely to the epidemic threshold condition in (6.2.2), except that the original removal-rate for the human hosts is now replaced by the sum of the death-rate and the removal-rates to the two groups con-

taining immune removals and non-immune susceptibles, i.e. by the "total" removal-rate for human infectives.

The threshold condition (6.6.6) can also be obtained somewhat heuristically in the same way as (6.2.2), by assuming y_0 and y_0' both small, with $x_0 \doteq n$ and $x_0' \doteq n'$, and requiring the initial rates of change in the numbers of infective hosts and vectors to be positive.

Now it is obvious that the null endemic level is stable since all infection has disappeared, and we have not, in the above model, envisaged any importation of infection from outside the community. But, for the solution given by (6.6.4), it is of some practical importance to know whether the equilibrium is stable or unstable. Epidemiological intuition suggests stability, since enduring endemic states are widespread in nature and the model is not totally remote from reality.

The investigation of stability problems can involve extensive mathematical work, and each individual model is liable to require special treatment. For a good general text on the relevant qualitative theory of the sets of ordinary differential equations that frequently arise from continuous-time models, see Brauer & Nohel (1969).

Fortunately a number of special problems in communicable disease modelling have been studied in some depth by Hethcote (1974, 1975, 1976), and one or two of these apply in particular to host-vector models. Assuming the absence of immunity, we have for the human hosts what Hethcote calls an SIS model, i.e. involving the transitions: susceptible → infective → susceptible, as opposed to SIR models which involve: susceptible → infective → recovery with immunity. For the mosquito vector, the pattern of transitions is simpler still, being represented by SI, since there is no recovery, with or without immunity. If, therefore, in the model leading to equations (6.6.1) we abolish the human removal group by putting $z = 0$ and $\gamma = 0 = \nu$, we are left with only two independent equations given by

$$\left.\begin{aligned} \frac{dy}{dt} &= \beta(n-y)y' - (\lambda+\mu)y, \\ \frac{dy'}{dt} &= \beta'(n'-y')y - \gamma'y'. \end{aligned}\right\} \tag{6.6.7}$$

These equations have steady-state solutions which can be found either by direct solution, or by putting $\gamma = 0 = \nu$ in (6.6.4). We find the non-null equilibrium point to be

$$y = \frac{nn'\beta\beta' - \gamma'(\lambda+\mu)}{\beta'(n'\beta+\lambda+\mu)}, \qquad y' = \frac{nn'\beta\beta' - \gamma'(\lambda+\mu)}{\beta(n\beta'+\gamma')}. \tag{6.6.8}$$

Hethcote (1974) investigated the properties of this system by standard phase-plane analysis. Careful examination of the (y, y')-plane portraits shows that if $nn'\beta\beta' \leqslant \gamma'(\lambda+\mu)$ the origin is asymptotically stable, and is in fact the only equilibrium point if $nn'\beta\beta < \gamma'(\lambda+\mu)$; while if $nn'\beta\beta > \gamma'(\lambda+\mu)$ the origin is unstable, but the equilibrium point given by (6.6.8) is asymptotically stable. It can also be shown that there are no periodic solutions contained entirely within the relevant region given by $0 \leqslant y \leqslant n$, $0 \leqslant y' \leqslant n'$.

Thus, suppose we define the removal-rates as

$$\rho = (\lambda+\mu)/\beta, \qquad \rho' = \gamma'/\beta'. \tag{6.6.9}$$

Then if the product of the two removal-rates is above the threshold nn', the infection eventually disappears; but when the product is below the threshold the disease remains endemic, with levels given by (6.6.8).

For the more general model of (6.6.1), we can define the removal-rates as

$$\rho = (\gamma+\lambda+\mu)/\beta, \qquad \rho' = \gamma'/\beta', \tag{6.6.10}$$

and may conjecture that the same kind of asymptotic stability will be obtained as in the simpler model when $\rho\rho' < nn'$. Some support for this view comes from considering an algebraically similar set of equations obtained by Bailey (1979, equations (34)) in describing the population dynamics of gonorrhoea when allowance is made for the inclusion of both symptomatic and asymptomatic infectives, the former being supposed to be under treatment and effectively out of circulation.

The relevance of the venereal disease modelling is that there is the same kind of "criss-cross" infection switching back and forth between two populations (at least for heterosexual contacts) as in the host-vector situation that we have been considering for malaria. However, in the latter case the mosquito dynamics have been deliberately kept simple in accordance with generally accepted ideas. It is interesting to note that the correspondence between venereal and parasitic diseases was first noted by Ross (1911, p. 685).

Thus, a similar conjecture for the gonorrhoea model was made by Bailey (1979) in regard to stability, and it was subsequently shown by Wichmann (1979) that the endemic state was asymptotically stable, at least for values of the parameters in the critical region near the threshold (linear approximations having been used near the equilibrium points). Further development of this work would obviously increase the stock of general theory having some practical relevance.

Although the model described by equations (6.6.7) was shown by Hethcote (1974) to have no strictly periodic solutions, i.e. closed paths in the phase-plane, as mentioned above, it may be asked whether this excludes all periodic components. Could there be, for example, a damped oscillatory path spiralling in to the equilibrium point? The last two paragraphs of the paper by Radcliffe (1974) appear to demonstrate such a possibility in relation to a deterministic model that is virtually identical with the one we are discussing, though derived from a large-population approximation to a stochastic formulation. However, as confirmed by private communication, this is due to a slip as we can easily show. (The period of oscillation given by Radcliffe is $2\pi/z$, where $z = (c - b^2)^{1/2}$, but using the values of c and b quoted we find $c - b^2 < 0$.)

Suppose in our model we write, in the neighbourhood of the equilibrium point (y_0, y_0') given by (6.6.8),

$$y = y_0(1 + u), \qquad y' = y_0'(1 + u'), \tag{6.6.11}$$

where u and u' are small. Substitution into (6.6.7) yields, to first order in u and u', two simultaneous first-order homogeneous linear differential equations in u and u' whose eigenvalues are real and negative: in the absence of conjugate complex values no periodic terms can arise.

A more complex model than any of the foregoing has been developed by Adler (1976). This distinguishes five groups of individuals for the human hosts, namely: susceptibles, infected but latent, infectious, recovered and immune, and dead; while for the mosquito vectors there are just three groups, namely: susceptible, infectious and dead. It is supposed that susceptible humans who become infected pass into the latent group and then into the infectious group. They are subsequently removed by death, recovery with immunity, or recovery without immunity which means re-entry to the susceptible group. Normal birth and death are ignored in the human population, but are included in the mosquito dynamics.

This is very much in the same style as the epidemic model of Olaofe & Olaofe (1975) described at the end of Section 6.2: indeed the latter authors acknowledge Adler as having initiated their problem. Although Adler describes his model as referring to vector-borne *epidemics*, it is clear from the formulation, particularly the partial replenishment of the stock of human susceptibles through a degree of recovery without immunity, that the model is more applicable to *endemic* situations. However, this matter is complicated by the slow steady loss of the human population through deaths from malaria.

The states of the system are specified in terms of eight variables which give the number of individuals in each group, using either a cumulative function or a density function, as appropriate. In addition to the time-variable for the process as a whole, humans are subclassified according to the time spent in latent or infectious groups, while mosquitoes are distinguished by their ages. Upper limits are set to the time humans may spend in the latent and infectious groups, and to the lifetime of mosquitoes. The permissible age-dependent transitions from one group to another are basically defined in terms of probabilities, which are then interpreted as transition-rates to provide a deterministic model.

The above assumptions, sketched here only in broad outline, lead to a somewhat elaborate formulation of the dynamics of the system in terms of a set of integro-differential equations. Relatively little explicit analytical development seems to be possible because of mathematical intractability, but a rigorous investigation of solution uniqueness and the admissible classes of weight functions, as well as the proof of a threshold theorem, have been provided by Sowunmi (1977; and an unpublished paper).

Fortunately, however, numerical solution of the equations is possible using a computer. Adler gives a number of results based on an arbitrary choice of parameters, but including the interesting adoption of a periodic saw-tooth function for the mosquito birth-rate, based on an annual cycle. Since one of the salient features of mosquito dynamics in the field is an often massive variation in mosquito density due to climatic changes affecting birth-rate, this is a most welcome inclusion in a general theoretical discussion. The human group of infectives shows, under a variety of different assumptions as to parametric values, strong, stable, though damped, oscillations related to the transition probabilities but hardly influenced at all (except in extreme cases) by large swings in the mosquito birth-rate. This result clearly requires further investigation.

A somewhat similar approach has also been developed by Elderkin *et al.* (1977), dealing primarily with the steady state of an age-dependent malaria model. Fairly intricate mathematical investigations were carried out dealing with questions of existence, uniqueness and properties of solutions of the basic equations. The authors also refer to additional theorems to be presented elsewhere.

More insight into the structure of models with fluctuating mosquito birth-rates might be obtained by incorporating this feature directly into one of the simpler models discussed earlier. Take, for example, the model in which human recovery involves no immunity, for which the relevant equations were given in (6.6.7). We can continue to assume a constant

total human population, but must allow for variations in the mosquito population. Equations (6.6.7) therefore still stand, but n' must now be a function of time $n'(t)$.

We could thus assume a constant death-rate γ', together with a periodic birth-rate $\gamma'(1 + k \sin \omega t)$. The net rate of mosquito increase is thus

$$\frac{dn'}{dt} = n'k\gamma' \sin \omega t,$$

from which we obtain

$$n' = C \exp \{ - (k\gamma'/\omega) \cos \omega t \}, \tag{6.6.12}$$

the constant C depending on initial conditions.

Substituting for n' from (6.6.12) into the second line of (6.6.7) then gives two equations in two unknowns. The functional form of n' looks rather unpromising, and it remains to be seen whether the qualitative behaviour of these equations can be deduced analytically. Alternatively, it might be reasonably simple to assume that, as a result of environmental pressures, the mosquito death-rate remained as a constant γ' while the total population varied cyclically like, for example, $n' = N'(1 + k \sin \omega t)$. Again, one could try using a saw-tooth function, varying linearly between upper and lower limits, for either the birth-rate (as in Adler, 1976) or the total population. For further ideas on the investigation of the effects of seasonal variations in contact-rates, at least in the context of diseases like measles with case-to-case disease transmission, see Dietz (1976).

There is also the question of investigating stochastic versions of the various endemic models. Little has been done in this direction so far, presumably because of the difficulty of achieving any real mathematical development after the initial stage of formulating the basic equation for the relevant probability- or moment-generating function. This initial stage is itself a matter of some complexity, as shown by considering a stochastic version of the simplest model, typified in deterministic form by equations (6.6.7).

Since both infectives and susceptibles can fluctuate to some extent independently in hosts and vectors, there are altogether four basic random variables to be taken into account. In the human population there are five types of transition, corresponding to infection of susceptibles, removal of infectives, birth of susceptibles, and deaths of both susceptibles and infectives. Even with the simplification assumed for mosquitoes, there are still four types of transition involved.

There is perhaps little point in simply writing down the resulting partial differential equation, although this can easily be done directly using a standard method, as with the stochastic epidemic model dealt with in Section 6.3. In fact, Radcliffe (1974) does this for a very similar model in which the transition probabilities of birth are constants not depending on the population sizes. His subsequent development is, however, restricted to deriving the deterministic approximation relevant to the large-population situation already discussed above.

Another possibility is to undertake the kind of studies made by Bartlett (1956) on stochastic models of recurrent epidemics for single populations, using a variety of approximation techniques (see also Bailey, 1975a, Sections 7.4 to 7.7). The significance of Bartlett's work is that stochastic models appeared to have important implications for the behaviour of recurrent epidemics even in quite large groups, e.g. towns with populations of the order of 250 000, for which deterministic models might otherwise have been thought adequate.

For the sake of completeness, perhaps some mention should also be made of the application of host-vector modelling to a different field, namely the transmission of ideas, by Goffman & Newill (1964). Their model includes the continuous introduction in both populations of new susceptibles and infectives, as well as the direct removal of susceptibles. Whether these notions have any special implications for epidemiology remains to be seen.

7 Elementary population dynamics of malaria

7.1 Introduction

Chapter 6 dealt in a broad manner with the quantitative theory of host-vector diseases in which there was a compulsory switching back and forth of the relevant parasite between host and vector. Much of this was applicable to malaria, but only in a general way regarding the qualitative nature of thresholds or the spatial spread from an initial focus. The time has now come to examine models which are much more disease specific, and which can be accepted as having a more realistic approach to malaria as such.

It is no accident that most of the work in the previous chapter has been developed principally by professional mathematicians of one kind or another. The centre of gravity of these activities is thus well inside the field of mathematics itself, and the discussion has a strong tendency to veer towards the abstruse. The present chapter, on the other hand, deals with a largely distinct line of evolution. Its general direction is principally due to the epidemiologists Ronald Ross (1909 and later) and George Macdonald (1950a and later), who also had appreciable competence as amateur mathematicians. Indeed, Ross's (1911) seminal work on the mathematics of the population dynamics of malaria considerably antedates the later, more theoretically oriented, stream of research. The approach discussed in detail below represents therefore a much closer integration of collaborative activities between (a) epidemiologists who have some mathematical skills and (b) biomathematicians who have chosen to work in a strongly medical and epidemiological environment. It is my personal belief that a continuation of such integrated studies holds the greatest promise for successful and useful work in the future, not forgetting however that the qualitative insights of the general theory can provide a valuable background against which many thorny technical problems can be reviewed.

7.2 Ross's Threshold Theorem for malaria

The first mathematical formulations of the population dynamics of malaria appeared in the now classic work of Sir Ronald Ross (1909 and

later). Ross had discovered, about 1898 in India, the crucial importance of the mosquito in the life-cycle of the parasite, and had received the Nobel Prize for Medicine in 1902. (For an extensive historical and methodological review of Ross's quantitative work, see Fine, 1975a.)

An approximate quantitative analysis of the principal epidemiological factors involved in the transmission and maintenance of malaria led Ross in his lengthy report on malaria control in Mauritius (Ross, 1909) to distinguish the following ingredients (here keeping for convenience to Ross's own notation):

- p: average population in the locality;
- m: average proportion of the population infected;
- i: proportion of infected individuals who are infectious;
- a: average number of mosquitoes per person in the locality per month;
- b: proportion of uninfected mosquitoes which feed on man;
- s: proportion of mosquitoes which survive through the extrinsic incubation period;
- f: proportion of infectious mosquitoes which feed on man;
- r: recovery-rate of infected individuals, per month.

It then followed that the number of new infections delivered per month would equal the product of all these quantities except the last, namely $pmiabsf$ or $pmiab^2s$ if f and b could be regarded as equal. The number of healthy persons so infected per month would then be $pmiab^2s(1-m)$. Similarly, it can be seen that the number of recoveries in infected persons per month would be rmp.

Thus if recoveries exactly balanced new cases, giving an endemic state, we would have

$$pmiab^2s(1-m) = rmp,$$

leading to

$$m = 1 - \frac{r}{b^2sai}. \tag{7.2.1}$$

Ross used the tentative estimates: $r = \frac{1}{5}$, $s = \frac{1}{3}$, $i = \frac{1}{4}$ and $b = \frac{1}{4}$ to yield the approximate rule of thumb

$$m = 1 - \frac{40}{a}. \tag{7.2.2}$$

This formula allowed one to see how the actual endemic malaria rate would be related to the number of different anopheline mosquitoes per

person per month. If a were less than about 40 the malaria rate would vanish altogether. Ross insisted, however, that the calculation was "useful, not so much for its numerical results, but because it gives precision to our ideas", and he was quite well aware of the large number of disregarded practical influences that would modify any detailed conclusions that might be drawn.

Such reasoning constituted in essence Ross's proof that a suitable reduction of the mosquito population, rather than complete eradication, would be sufficient to eliminate malaria. He succinctly stated this result (Ross, 1911; end of Section 28, p. 164) in the following terms:

> "We may therefore conclude,
> (1) That the amount of malaria in a locality *tends* towards a fixed limit determined by the number of malaria-bearing mosquitos and by other factors.
> (2) That if the number of malaria-bearing Anophelines is below a certain figure, that limit will be zero."

These conclusions constitute what we might justifiably now designate as Ross's Threshold Theorem for Malaria, especially in view of the variety of threshold results now recognized in infectious disease dynamics. This result has been considered by Moshkovskii (1969) as having even greater importance than Ross's fundamental work on the role of the mosquito in malaria transmission. While we might think that such a comparison was invalid, there is no doubt about the significance of both discoveries. Moreover, nearly forty years were to pass before any appreciable improvements were to be made in the quantitative formulation of threshold phenomena by the more sophisticated analysis of George Macdonald (1950a and later).

It should be noted that what Ross called "the mosquito theorem" really referred to his own clear demonstration in 1898 of the role of the Anopheline mosquito in the transmission of malaria. This can be seen from the evidence summarized in Section 7 of Ross (1909), where it is also stated that "the theorem has now been before men of science for nearly ten years ...". The actual mathematical work first appeared in Ross (1909) and was subsequently refined in Ross (1910, 1911).

7.3 Elementary deterministic models

Elementary mathematical representations of malarial population dynamics had already been given by Ross (1909, 1910) in terms of exponential growth and simple difference equations. But in Ross (1911) we find a far

more elaborate discussion embodied in the Addendum, classified as Section 66 and entitled "Theory of Happenings". This involves an extensive general mathematical epidemiological investigation, including difference equations, differential equations, approximations, numerical illustrations, appeals to data etc., with specific applications to malaria as well. In particular, subsection (14) gives the derivation of what we would now regard as the essential differential equations specifying the simplest deterministic models in continuous time.

Ross in fact set up two slightly different models, one involving two first-order differential equations, the other involving only one. The latter must, however, inevitably involve a greater degree of approximation. An extensive mathematical review of these equations and their properties has been given by Lotka (1923), but for present purposes we need consider only the essential points as set out below.

The derivation of Ross's principal differential equations for both human and mosquito populations was of course based directly on his own epidemiological experience. Various simplifying assumptions were made, such as the absence of immunity in recovered individuals. Although there is an obvious parallel with the general, but rather abstract, discussion of Section 6.2, it will be more convenient here to consider Ross's own ideas *ab initio* in explicit malariological terms.

We first define basic parameters, conforming to Ross's original concepts, but adopting a notation that is more consistent with our own previous discussions. Thus for the human population we have:

n: total population size;
y: total number of infected individuals;
f: proportion of infected who are infectious;
γ: recovery-rate;
μ: birth-rate;
ν: death-rate.

A similar set of definitions, using primes, i.e. n', y', f', γ', μ' and ν', refers to the mosquito population. We also need a time variable given by t.

If we assume the restricted form of homogeneous mixing already introduced in Section 6.2 for specific application to malaria, we need the additional parameter b' for the man-biting rate in mosquitoes. It then follows that in time Δt the y' mosquitoes will make $b'f'y'\,\Delta t$ infectious bites, of which a fraction $(n-y)/n$ are on susceptible humans. The number of new infections in humans in Δt is then $b'f'y'(n-y)\,\Delta t/n$. We also have in Δt the numbers of recoveries and deaths given by $\gamma y\,\Delta t$ and $\nu y\,\Delta t$, respec-

tively. If all new births are assumed to be susceptible, the birth-rate does not appear explicitly at this stage. It follows immediately that the basic differential equation describing the rate of change of the human infected population is

$$\frac{dy}{dt} = \frac{b'f'y'(n-y)}{n} - (\gamma + \nu)y. \tag{7.3.1}$$

A similar argument for the mosquito population shows that the number of new infections in mosquitoes in Δt is $b'fy(n'-y')\,\Delta t/n$. This yields the equation

$$\frac{dy'}{dt} = \frac{b'fy(n'-y')}{n} - (\gamma' + \nu')y'. \tag{7.3.2}$$

These two equations are not exactly symmetrical with regard to the presence and absence of primes since only b' exists, the transmission of disease from mosquito to man, and vice versa, being controlled in both cases by the mosquitoes' man-biting habit. Symmetry would occur only if man bit mosquitoes!

Ross himself further supposed that in man ν was very small in comparison with γ, while in the mosquito γ' could be ignored compared with ν'. If a constant mosquito population is also assumed, the birth- and death-rates must balance and ν' can be replaced by μ'. With these approximations in mind, we can rewrite (7.3.1) and (7.3.2) as

$$\left.\begin{aligned} \frac{dy}{dt} &= \frac{b'f'y'(n-y)}{n} - \gamma y, \\ \frac{dy'}{dt} &= \frac{b'fy(n'-y')}{n} - \mu'y', \end{aligned}\right\} \tag{7.3.3}$$

which are essentially the same as the forms given by Ross, using approximations included in his Addendum (Ross, 1911, Section 66). The equilibrium solutions of (7.3.3) are obtained by putting $dy/dt = 0 = dy'/dt$, giving

$$\left.\begin{aligned} b'f'y'(n-y) &= n\gamma y, \\ b'fy(n'-y') &= n\mu'y'. \end{aligned}\right\} \tag{7.3.4}$$

One obvious solution is the null result given by

$$y = 0 = y', \tag{7.3.5}$$

when all infection has disappeared from both human and mosquito popu-

lations. The non-null solution is easily obtained by dividing through the equations in (7.3.4) by yy', and solving the resultant simultaneous equations which are linear in y^{-1} and $(y')^{-1}$. We quickly find

$$y = \frac{n'b'^2ff' - n\gamma\mu'}{b'f(n\gamma + n'b'f')/n}, \qquad y' = \frac{n'b'^2ff' - n\gamma\mu'}{b'f'(\mu' + b'f)}, \tag{7.3.6}$$

as given by Ross with minor differences in notation.

While Ross dealt only with the equilibrium solutions of (7.3.4), an extensive investigation of these nonlinear equations was later made by Lotka (1923) who obtained approximate time-dependent solutions near the equilibrium points. If we define the malaria-rate in man as $m = y/n$ and the density of infected mosquitoes per head of human population as $u = y'/n$, together with the overall mosquito density per human as $a = n'/n$, then (7.3.6) can be replaced by the more convenient forms

$$m = \frac{ab'^2ff' - \gamma\mu'}{b'f(\gamma + ab'f')}, \qquad u = \frac{ab'^2ff' - \gamma\mu'}{b'f'(\mu' + b'f)} \tag{7.3.7}$$

the null result in (7.3.5) being replaced by

$$m = 0 = u. \tag{7.3.8}$$

A neat simplification of the formulae in (7.3.6) has been pointed out by Nåsell (1980a). This involves the use of two dimensionless "transmission factors", given by

$$T_1 = \frac{b'f'n'}{\gamma n}, \qquad T_2 = \frac{b'f}{\mu'}, \tag{7.3.9}$$

the substitution of which into (7.3.6) yields the symmetrical expressions

$$\left.\begin{aligned} P_1 \equiv \frac{y}{n} &= \frac{T_1T_2 - 1}{T_1T_2 + T_2}, \\ P_2 \equiv \frac{y'}{n'} &= \frac{T_1T_2 - 1}{T_1T_2 + T_1}. \end{aligned}\right\} \tag{7.3.10}$$

In terms of observed estimates $\hat{P}_1$ and $\hat{P}_2$ of P_1 and P_2, respectively, we could estimate T_1 and T_2 as

$$\hat{T}_1 = \frac{\hat{P}_1}{\hat{P}_2(1-\hat{P}_1)}, \qquad \hat{T}_2 = \frac{\hat{P}_2}{\hat{P}_1(1-\hat{P}_2)}. \tag{7.3.11}$$

A seven-parameter system is thus conveniently reduced to a two-parameter system, yielding potentialities for further simplified analysis (see also Section 8.4 on hybrid models). It may be noted that the use of transmission factors had already been introduced by Nåsell & Hirsch (1973) in their study of schistosomiasis.

For any given set of parametric values only one solution could be achieved in practice. Thus, if we define a quantity R as

$$R \equiv \frac{ab'^2ff'}{\gamma\mu'} = T_1T_2, \tag{7.3.12}$$

then it could be shown that when $R \leqslant 1$ the null point in (7.3.8) was stable. Thus a few malaria cases introduced into a malaria-free population would not provoke any epidemic outbreak or endemic persistence of disease: the system would soon revert to a malaria-free situation.

If, on the other hand, $R > 1$ then the introduction of a small amount of disease would cause the prevalence to rise to a stable endemic level whose precise value was determined by (7.3.7).

The quantity R is in fact what Macdonald (1952b, 1957), in a somewhat different specification of the fundamental malaria dynamics, designated the *basic reproduction-rate*. This is, in effect, the average number of secondary cases arising from a single primary case in a very large population of susceptibles. A direct heuristic derivation is as follows.

The human recovery-rate is γ, so the average time spent in the infected state is $1/\gamma$. During the latter time the average number of mosquito bites received from a susceptible mosquitoes each with a biting-rate of b' is ab'/γ. Of these bites only a proportion f are infectious for the mosquito. Thus there will be an average of $ab'f/\gamma$ mosquitoes infected by a given human primary case. Now each of these mosquitoes survives for a time $1/\nu'$, or $1/\mu'$ if the mosquito birth- and death-rates are assumed equal. The number of infectious bites inflicted, largely on human susceptibles, thus totals $b'f'/\mu'$. And so the total number of secondary cases must be

$$(ab'f/\gamma).(b'f'/\mu') = ab'^2ff'/\gamma\mu' = R.$$

Lotka's (1923) work thus put Ross's pioneering advances on a more rigorous mathematical basis. And from a modern point of view we could appeal to current methods of stability analysis, as already indicated in Section 6.6, to give results of wider validity. There is an obvious analogy between (7.3.3) and (6.6.7), so that the work of Hethcote (1974) implies that in the present case the null solution is asymptotically stable when $R \leqslant 1$, while if $R > 1$ the null position is unstable but the endemic level is

asymptotically stable. There are moreover no periodic solutions contained entirely within the relevant region.

In principle, however, the conclusions of Ross's Threshold Theorem for Malaria are vindicated. The control of malaria depends on reducing the basic reproduction-rate to below unity. This may be achieved by a variety of means, including antimalarial drugs, mosquito nets, insecticides and larvicides. And the point is underlined by all the different forms of mathematical analysis that reduction of the mosquito population alone below a certain critical level would be sufficient to eliminate malaria, i.e. without the need to attempt a complete eradication of the mosquito.

Finally, we should mention for the sake of completeness the alternative, single, first-order differential equation set up by Ross (1911, Section 66, p. 678) to describe the course of malaria in an infected community. As noted by Ross, such an equation could only be an approximation to a host–vector situation in which *two* basic (first-order) equations are really needed to describe the behaviour of the *two* species involved. Of course, if we eliminate y' from (7.3.3) we can obtain a single second-order equation in y and t. Similarly for y' and t. In general, such a second-order equation cannot be reduced to a single first-order equation.

If we go back to Section 7.2 and adopt the elementary approach used by Ross in arriving at his Threshold Theorem, it can be that the increment in the number of infected humans in some suitable unit of time can be written

$$\Delta(mp) = pmiab^2s(1-m) - rmp, \tag{7.3.13}$$

an expression of which Ross made considerable use.

If we refer (7.3.13) to a time-interval Δt, we can immediately derive the differential equation

$$\frac{dm}{dt} = b^2saim(1-m) - rm. \tag{7.3.14}$$

When $dm/dt = 0$, we obtain quite satisfactorily the previous equilibrium result in (7.2.1). But for really adequate time-dependent results, there are too many over-simplifications involved, including omission of continuous changes in the mosquito population.

Lotka (1923) has examined this situation at great length, giving an extensive comparative mathematical analysis of Ross's two models and pointing out the various explicit and implicit assumptions involved. The interested reader should consult Lotka's work for details. Here it is sufficient to note that the two-equation model of Ross discussed above is more

relevant to the main line of development. It is of course over-simplified in many respects, especially in regard to the phenomenon of superinfection which we shall discuss in the next subsection, but it does however portray some of the main features of malaria in a very clear manner, and has been used by Dietz (1975) as a first approximation in certain investigations into the use of control theory (see Section 10.3).

7.4 Infection and superinfection in humans

In Section 2.5 we referred briefly to the possibility of *superinfection*, i.e. the occurrence in a given individual of distinct broods of parasites in the blood corresponding to reinfections by different mosquitoes at different times. When this happens there is not just one class of infected humans: the latter should really be further classified at least according to the number of broods flourishing simultaneously side by side. There are obviously considerable opportunities for highly complex biological descriptions, leading to intractable mathematics. If and when superinfection is likely to occur it should be treated in an adequate and consistent manner. At the same time, undue attention to the fine structure of the process must be avoided if a useful type of population dynamic analysis is to be developed.

Strangely enough, the concept of superinfection, though simple in essence, has given rise to a good deal of confusion. There are of course different ideas on the best practical way of modelling the phenomenon. But there are also misunderstandings by some researchers of the explanations of other researchers with different ideas. And there have even been misunderstandings between statisticians and epidemiologists supposedly working with the same model! There are some historical facts in the literature, and much historical conjecture. Thus we shall now probably never know for sure what the protagonists Ross or Macdonald really thought about certain controversial issues. Those readers who would, however, like to study the subject in more detail should consult Paul Fine's (1975a, b) splendid historical review papers.

For our purposes it is important to be aware of the principal methods of describing superinfection, the development of the corresponding mathematical formulations, the variety of theoretical implications, and their success or failure in adequately accounting for actual data.

Although Ross himself did not deal explicitly with superinfection, it seems likely that he had something of this nature in mind. Thus in the "Theory of Happenings" (Ross, 1911, Section 66) there are several references to the theory of *repeated* happenings, including bitings and infections, in which a specific identification is made of the numbers or proportions of a population to whom an event has happened once, twice etc. (e.g. subsections (3), (4), (8), (15), (16)). In particular, use is made of the Poisson series to represent the appropriate population distribution.

In view of subsequent developments, it is useful to start at this pre-superinfection stage of modelling. Ross's essential contributions are enmeshed in a mass of partly theoretical and partly applied mathematical discussion. For, although specifically committed to the prevention of malaria (Ross, 1911), he not only embarked on the more general quantitative epidemiological analysis covered by the "Theory of Happenings" but later developed the "Study of *a priori* Pathometry" (Ross, 1915, 1916; Ross & Hudson, 1917a, b). The notion of a "happening" was intended to deal with any kind of event occurring in a population, such as birth, death, infection, recovery, etc. Hence the use of the initial letter "h" to represent the rate at which happenings occurred.

In the context of malaria, h can be used to represent the rate at which susceptible individuals contract the disease. Although h can for the moment be regarded as a suitably chosen fixed parameter, it must in any finer analysis be seen as being a time-dependent variable highly related to the epidemiological status of the co-existing mosquito population. Ross also defined a recovery-rate r for infected individuals. Ignoring the variety of generalizations Ross was working with, his basic treatment of the malaria situation regarding humans included only the transitions of infection and recovery just mentioned. Thus, suppose that a proportion x of the population is infected, and a proportion $1-x$ is susceptible. Then in time Δt, the proportion x is increased by an amount $h(1-x)\,\Delta t$ due to infection, and decreased by $rx\,\Delta t$ due to recovery. The relevant differential equation is accordingly

$$\frac{dx}{dt} = h(1-x) - rx = h - (r+h)\,x. \tag{7.4.1}$$

Assuming h and r to be constant, the solution of (7.4.1) is

$$x = \frac{h}{r+h} - \left(\frac{h}{r+h} - x_0\right) e^{-(r+h)\,t}, \tag{7.4.2}$$

where x_0 is the initial value of x at $t = 0$, thus yielding a typical "catalytic" curve of the type later discussed by Muench (1959).

If we now concentrate attention on a cohort of newborn susceptibles,

we can put $x_0 = 0$, to give

$$x(t) = \frac{h}{r+h}(1 - e^{-(r+h)t}). \tag{7.4.3}$$

This equation therefore defines the prevalence-rate of infection in terms of age, rising from $x = 0$ at $t = 0$ to the asymptotic value $x = h/(r+h)$ as $t \to \infty$. There was little serious criticism of this result until Macdonald (1950a,b) began to examine the applicability of the curve to various bodies of field data on the prevalence of malaria in successive age-groups of children living in malarious areas. It appeared that the curve could be made to fit the data, but the resultant estimates of the recovery-rate r were often greatly below anything that seemed credible on other grounds. For example, the calculated recovery-rates were usually below 0·001 per day and often below 0·0005. This would imply a duration of continuous parasitaemia from a single infection lasting for a period of the order of three to six years, a result generally regarded as quite inadmissible by malariologists, at least for *P. falciparum*.

Macdonald's (1950a) big step forward was to show that satisfactory results could be obtained by using a model incorporating some degree of superinfection. He first made the *assumptions*:

"(a) The amount of infective material to which the population is exposed remains unchanged.

(b) The existence of infection is no barrier to superinfection, so that two or more broods of organisms may flourish side by side, the duration of infection due to one being unaltered by others."

The first of these statements, implying a constant threat from infected mosquitoes, is of course a considerable over-simplification but perhaps quite reasonable as a first approximation. The second statement, explicitly introducing superinfection, seems to mean that separate broods should be considered as distinct entities, each with a certain average probability of being eliminated per unit of time.

Secondly, the following *definitions* were stated:

"(a) The inoculation rate is the proportion of the population receiving infective inocula in unit of time and will be referred to as h. Thus $h\,dt$ is the proportion inoculated in time dt.

(b) The recovery rate is the proportion of affected people (who have received one infective inoculum only) who revert to the unaffected group in unit of time, and will be called r."

A mortality rate m was also specified, but for present purposes we can assume that this is small enough to be ignored.

In his paper Macdonald (1950a) attributed the basic algebraic derivations to his distinguished colleague, Dr J. O. Irwin. Hence the convenient label "Macdonald-Irwin model" introduced by Nåsell (1980c) to refer to this particular formulation. Unfortunately, different interpretations of the superinfection concept are possible, and the mathematical form adopted does not precisely reflect the biological process as Macdonald appears to have perceived it. An unambiguous argument was not therefore forthcoming, and, indeed, no explicit justification was included at all in the "Mathematical Statement" in Appendix 1 of Macdonald's (1957) authoritative book published later – only sympathetic references to the previous work.

In general heuristic terms it was argued as follows. First, suppose $h < r$. Then in time Δt all members of the whole population, whether infected or not, exhibit new infections at rate $h\,\Delta t$. Thus the infected proportion of the population x increases by $h\,\Delta t$. In the same interval there are also $rx\,\Delta t$ recoveries, and x decreases by this amount. Conversely, when $h > r$ it may be supposed that once infected, an individual would never recover, since new infections would accrue faster than any brood could be eliminated. New infections would accordingly arise only from the susceptible fraction $1 - x$ and could amount to $h(1-x)\,\Delta t$. The resultant equations are therefore

$$\left.\begin{aligned} \frac{dx}{dt} &= h - rx, \qquad h \leqslant r, \\ &= h(1-x), \qquad h \geqslant r. \end{aligned}\right\} \tag{7.4.4}$$

If we take $x_0 = 0$ when $t = 0$, the relevant solutions of (7.4.4) are

$$\left.\begin{aligned} x(t) &= \frac{h}{r}(1 - e^{-rt}), \quad h \leqslant r \\ &= 1 - e^{-ht}, \qquad h \geqslant r, \end{aligned}\right\} \tag{7.4.5}$$

which may be compared with Ross's result in (7.4.3). When the formulae in (7.4.5) were applied by Macdonald (1950b) to data on *P. falciparum* excellent agreement was found. A realistic recovery-rate $r \doteqdot 0{\cdot}005$ was first deduced from data on untreated *P. falciparum* cases resulting from a single infection, allowing for an initial time-lag. Analyses were then made of various bodies of data on parasitaemia-rates in infants, yielding values of h

of about the same order as r, some less than r and some greater (e.g. $h = 0{\cdot}0025, 0{\cdot}0044, 0{\cdot}013$, etc.).

It is evident that, at least for $h \leqslant r$, the basic equations and their solutions differ between the Ross and the Macdonald-Irwin models only in respect of the relevant parametric values, e.g. comparing (7.4.1) and (7.4.3) with (7.4.4) and (7.4.5). If the recovery-rates in the Ross and Macdonald-Irwin formulations are designated by r_1 and r_2, respectively, we must have $r_2 = r_1 + h$. The latter relationship allows r_2 in the superinfection version to be sufficiently large to agree with data on the recovery of untreated cases resulting from a single infection.

Thus the Macdonald-Irwin model had an immediate temporary success in explaining real data in terms of credible estimates of the two basic parameters. Nevertheless, later investigators had difficulties in accepting (7.4.4) exactly as it stands. And it is now generally agreed that (7.4.4) does not adequately represent the implications of superinfection (Bailey, 1957, 1975a; Fine, 1975a, b).

In fact, in order to be clear about any model involving superinfection, it is really necessary to consider explicitly the proportions of infected individuals harbouring different numbers of broods. We shall, therefore, do just this for two alternative points of view, as previously referred to by Dietz (1970) and later discussed in more detail by Fine (1975b). See also the more recent analysis in depth by Nåsell (1980c).

Suppose, first, that the population contains a proportion of persons f_j with exactly j broods, $j \geqslant 1$. When $j = 0$, we have the proportion f_0 $(\equiv 1 - x)$ of susceptibles. Thus

$$\sum_{j=0}^{\infty} f_j = 1. \tag{7.4.6}$$

Now Irwin's assumption was that the recovery-rate r applied to all infected *individuals* in the population, no matter how many times they had been infected previously (see personal communication from Irwin, quoted in Fine, 1975b). In this case it is easy to see that in time Δt, the proportion f_j is increased by the contributions from adjacent classes given by $(hf_{j-1} + rf_{j+1})\,\Delta t$ and decreased by $(h + r) f_j \Delta t$. Note that these transition-rates do *not* contain any factors like j or $j + 1$. The set of basic differential equations describing this deterministic process are clearly

$$\left.\begin{aligned} \frac{df_0}{dt} &= -hf_0 + rf_1, \\ \frac{df_j}{dt} &= hf_{j-1} - (h + r) f_j + rf_{j+1}, \quad j \geqslant 1. \end{aligned}\right\} \tag{7.4.7}$$

These equations are formally equivalent to those for the non-equilibrium treatment of the length of a simple stochastic queueing process (see, e.g., Bailey, 1964, p. 149), where f_j would be the probability of a queue of length j.

In fact, Dietz (1970) pointed out that, to obtain the Macdonald-Irwin mathematical results in a satisfactory manner, one would have to assume that all broods present in the body could not be dealt with simultaneously but would have to form a queue and be eliminated one by one. It is easily seen that the foregoing model can be reinterpreted on a probability basis as follows. Consider a single individual. Let infections occur at random in Δt with probability $h\,\Delta t$ and form a queue. Only one brood is dealt with at a time, and its elimination occurs at random with probability $r\,\Delta t$. Equations (7.4.7) then result, with $f_j(t)$ being the chance that the individual harbours j broods $(j \geqslant 1)$, while $f_0(t)$ is the chance of being in an uninfected, susceptible state.

The general time-dependent solution of such a process is somewhat involved, and requires the use of Bessel functions with imaginary arguments (see Bailey, 1964). We shall therefore not pursue it here (but see Nåsell, 1980c, for details). An equilibrium solution is however easily obtained. Rearranging the right-hand side of (7.4.7), after putting $df_i/dt = 0$, gives immediately

$$0 = rf_1 - hf_0 = rf_2 - hf_1 = \ldots = rf_j - hf_{j-1} = \ldots. \tag{7.4.8}$$

It thus follows that

$$f_j = \left(\frac{h}{r}\right)^j f_0, \quad \text{all } j. \tag{7.4.9}$$

Using (7.4.6) then yields

$$\left.\begin{aligned} f_0 &= 1 - \frac{h}{r}, && h \leqslant r, \\ &= 0, && h \geqslant r. \end{aligned}\right\} \tag{7.4.10}$$

So the limiting prevalence-rate $x = 1 - f_0$ is

$$\left.\begin{aligned} x &= h/r, && h \leqslant r, \\ &= 1, && h \geqslant r, \end{aligned}\right\} \tag{7.4.11}$$

agreeing with the equilibrium conclusions obtained from the Macdonald-Irwin equations, as given in (7.4.5), when $t \to \infty$.

The alternative interpretation, as indicated by Dietz (1970) and conveniently labelled as the "Macdonald-Dietz model" by Nåsell (1980c), is

to follow what Macdonald appeared to imply in his *assumption* (b), given above, namely that the recovery-rate r should apply to individual broods, confusion having arisen because his *definition* (b) specified the recovery-rate r only in terms relevant to non-superinfected individuals. In any event, many would consider the concept of independently flourishing broods to be a closer approach to biological reality than notions involving some kind of queueing process.

We can now modify our previous arguments to give the increase in f_j in time Δt as $\{hf_{j-1} + r(j+1)f_{j+1}\}\Delta t$, and the decrease as $(h + rj)f_j \Delta t$. The new equations, analogous to (7.4.7), are evidently

$$\left.\begin{aligned} \frac{df_0}{dt} &= -hf_0 + rf_1, \\ \frac{df_j}{dt} &= hf_{j-1} - (h + rj)f_j + r(j+1)f_{j+1}, \quad j \geqslant 1. \end{aligned}\right\} \tag{7.4.12}$$

These equations are essentially the same as those derived by Bailey (1957, p. 158) for a slightly more general form of the superinfection model now under discussion, which included the size of the mosquito population explicitly. The limiting distribution for the number of different broods harboured was also given.

However, an exact time-dependent solution of the equations in (7.4.12) can be obtained without too much difficulty. (It may be noted that, assuming $f_{-1} \equiv 0$, the expression for df_0/dt is precisely that given by df_j/dt with $j = 0$. The corresponding statement is not, however, true of equation (7.4.7).) As pointed out by Dietz (1970), we can reach an immediate solution by drawing a close analogy between our deterministic model and a stochastic homogeneous immigration-death process with immigration-rate h and death-rate r, starting with zero population size. If, for a single individual, the transitions take place at random, and we write $f_j(t)$ for the chance that the individual harbours j broods, then the relevant differential equations for the $f_j(t)$ are precisely those in (7.4.12). In this case (see, e.g., Cox & Miller, 1965) the number of infections at any age t turns out to be a Poisson distribution with parameter

$$(h/r)(1 - e^{-rt}). \tag{7.4.13}$$

The resulting parasite-rate $x(t)$ is the proportion of individuals of age t with at least one infection, i.e.

$$x(t) = 1 - \exp\{-(h/r)(1 - e^{-rt})\}, \tag{7.4.14}$$

and we can use this formula either for the straightforward prevalence-rate in the deterministic model, or for the average prevalence-rate in the stochastic version.

A direct time-dependent solution of (7.4.12), due to Bailey (1975a, p. 319), is as follows. Define the generating function

$$F(w, t) = \sum_{j=0}^{\infty} w^j f_j. \tag{7.4.15}$$

Next, multiply the general equation of (7.4.12) by w^j and sum over all j. We easily obtain

$$\frac{\partial F}{\partial t} = h(w-1)F - r(w-1)\frac{\partial F}{\partial w}, \tag{7.4.16}$$

with the initial condition, suitable for studying a cohort of newborn susceptibles,

$$F(w, 0) = 1. \tag{7.4.17}$$

The subsidiary equations for (7.4.15) are

$$\frac{dt}{1} = \frac{dw}{r(w-1)} = \frac{dF}{h(w-1)}, \tag{7.4.18}$$

yielding the two intermediate integrals

$$(w-1)\,e^{-rt} = \text{const.}, \qquad Fe^{-hw/r} = \text{const.} \tag{7.4.19}$$

The general solution may therefore be written as

$$F = e^{hw/r}\Phi\{(w-1)\,e^{-rt}\}, \tag{7.4.20}$$

where the arbitrary function Φ can be determined from the initial conditions. Putting $t = 0$ in (7.4.20) and using (7.4.17) gives $e^{hw/r}\Phi(w-1) = 1$, or

$$\Phi(v) = e^{-h(v+1)/r}. \tag{7.4.21}$$

Applying this result to (7.4.20) gives the solution

$$F(w, t) = \exp\{(h/r)(w-1)(1 - e^{-rt})\} \tag{7.4.22}$$

The prevalence-rate is $x(t) = 1 - f_0(t) = 1 - F(0, t)$. We thus obtain the result already given in (7.4.14).

We therefore have three different specifications, which we have designated, in the order treated, as the Ross, Macdonald-Irwin and Macdonald-

Dietz models, respectively, leading to the age-specific prevalence-curves in humans given by (7.4.3), (7.4.5) and (7.4.14). It is important to realize that the models all differ in certain essential aspects of structure, though, as we have seen, the net result for the first two models is a change in only one of the parameters, at least in the context of the analysis undertaken. However, the recovery-rate r has a different meaning in each case.

It is convenient to refer to the three models, in the order introduced above, by the suffixes 1, 2 and 3 (we have already used 1 and 2), thus writing, for example, r_1, r_2 and r_3 for the respective recovery-rates. Direct comparisons of the different prevalence-curves $x_1(t)$, $x_2(t)$ and $x_3(t)$ at a given value of t for some common value of r_i (as done by Fine, 1975b) can be misleading, because the choice of any given r_i depends on i, i.e. on the model chosen. For a given body of data the best estimates of the three r_i will in general be different from one another. See Nåsell (1980c) for a detailed comparative review.

We can, however, easily compare behaviour at the end-points. Let us write $h/r_i = \theta_i$. All three prevalence-curves start of necessity at $x_i(0) = 0$, and the three initial rates of increase are all equal, i.e. $dx_i(0)/dt = h$. Finally, as $t \to \infty$, we have

$$\left.\begin{aligned} x_1(\infty) &= \theta_1/(1+\theta_1) && \text{(Ross);} \\ x_2(\infty) &= \theta_2,\ \theta_2 \leqslant 1 \\ &= 1,\ \theta_2 \geqslant 1 && \text{(Macdonald–Irwin);} \\ x_3(\infty) &= 1 - \mathrm{e}^{-\theta_3} && \text{(Macdonald–Dietz).} \end{aligned}\right\} \tag{7.4.23}$$

It is of interest to note that the limiting value of the prevalence-rate for the Macdonald–Dietz model, shown in (7.4.23), is virtually equivalent to a formula obtained much earlier by Walton (1947). The latter's work was in fact mentioned by Macdonald (1950a) but rejected by him in the form in which it was proposed. Although, suitably interpreted, Walton's formula does lead to what is now regarded as the correct implication of Macdonald's thinking, the justification appears to have been insufficiently established in terms of epidemiological detail (see also Fine, 1975b, for further comments).

The foregoing discussion shows how erroneous it can be to choose formulae from the literature (as some investigators have done) simply on the basis of algebraic convenience. Sometimes, it is true, the choice is not very critical, e.g. when h/r is small. But in other situations, e.g. h/r large and especially if greater than one, there may be appreciable differences. A wrong choice can then be very misleading. We must therefore make up our

minds in any given application what assumptions *vis-à-vis* superinfection seem to be most appropriate. Having made such a choice, we are then committed to that particular model and must accept the formulae and other results that flow from it.

One final point to be made in connexion with the determination of human prevalence-rates is that a consistent measure must be used. Thus we may choose to work with the gametocyte-rate, i.e. the proportion of the population with detectable levels of gametocytes in their bloodstreams. We may also determine the corresponding levels of trophozoites. Or we may work with the overall parasite-rate, which would include all observable forms of the parasite. In comparing results based on different assumptions, appropriate adjustments will have to be made.

It should also be emphasized that none of the previous discussions makes any allowance for the emergence of some degree of human immunity (see Section 2.5). This arises through repeated exposure to malarial infection, which in turn leads to a fall in parasite densities in the blood. Conversely, in the absence of continual inoculations immunity may eventually be lost. As the biological mechanisms are not yet well understood it is not easy to incorporate immunity phenomena into the mathematical modelling. However, this subject will be taken up more explicitly in Chapter 8.

7.5 Infection in the mosquito

The previous section dealt primarily with malarial infection in humans, assuming for the most part a constant threat of infection from the attendant mosquito population. It is now time to look more closely at certain aspects of the latter. Whatever conclusions we come to about the most appropriate way to handle the problems of infection and superinfection in humans, the entomological phenomena can be investigated more or less independently. Once again Macdonald (1952a and later) has made a series of major contributions, which are fortunately independent of the Macdonald–Irwin superinfection model.

A minor problem of notation now arises. The extensive mathematical work of Ross (1909 and later) and the detailed further investigations of Lotka (1923) introduced a plethora of symbols whose usage was not always maintained by later writers, e.g. Macdonald (1950a and later), Armitage (1953) and Macdonald, Cuellar & Foll (1968). Fortunately, Macdonald and his associates have been fairly consistent, and we shall adopt their notation here. The parameter specifications of Sections 7.2 and 7.3 will therefore be largely set aside, although some will survive. We

shall retain the general terminology of Section 7.4 so far as x, h and r are concerned, although, as previously noted, care may be needed with precise definitions in specific applications (particularly x and r).

Let us therefore distinguish the following:

- x: proportion of human population infectious to mosquitoes (gametocyte-rate);
- h: rate at which new infections take place (exact implication depends on model);
- r: recovery-rate (exact definition depends on model);
- m: mosquito density per human;
- a: average number of humans bitten per day by any one mosquito;
- s: proportion of mosquitoes with sporozoites in their salivary glands;
- b: proportion of mosquitoes, having sporozoites in their salivary glands, that are actually infectious;
- n: time taken for extrinsic parasite cycle in the mosquito;
- p: probability that a given mosquito survives over a given day;
- ν': mosquito death-rate.

As originally presented by Armitage (1953), we can adopt the following straightforward argument. We assume that the mosquito population is stationary, with births and deaths balancing each other. Suppose, in addition, that the survival curve is negative exponential. Then the proportion of insects surviving for more than t days is $e^{-\nu' t}$, and the proportion of the population between ages t and $t + \Delta t$ is $\nu' e^{-\nu' t} \Delta t$.

Next, we consider the implications of the extrinsic cycle, of length n. Mosquitoes of age less than n have not had time to become infectious. For those of any age $t > n$. there is an interval of length $t - n$ during which infection in the form of gametocytes could have been acquired by biting an infected human: such an infection would have had time to develop into the sporozoite phase, invade the salivary glands of the mosquito and make the latter infectious at least by age $t > n$.

We assume, as previously, that the recovery-rate in the mosquito is negligible compared with the death-rate. Thus, once infected, the mosquito remains so for the rest of its life. For any given age $t > n$, we can suppose that the distribution of adequate contacts with infected humans is Poisson with parameter $ax(t - n)$, where a is the man-biting rate per mosquito. The chance of at least one such contact, as required for the receipt of infection, is $1 - \exp\{-ax(t - n)\}$. The proportion of infected mosquitoes

in the population between the ages t and $t + \Delta t$ must therefore be

$$[1 - \exp\{-ax(t - n)\}]\, \nu' e^{-\nu' t} \Delta t,$$

and the overall proportion of infected mosquitoes is the sporozoite-rate

$$s = \int_n^\infty [1 - \exp\{-ax(t - n)\}]\, \nu' e^{-\nu' t}\, dt$$

$$= \frac{ax\, e^{-\nu' n}}{ax + \nu'}. \tag{7.5.1}$$

For a death-rate ν', the chance of a mosquito surviving for one day, i.e. not dying during one day, is just the tail of the survival curve for $t > 1$. Macdonald's parameter p is therefore given by

$$p = e^{-\nu'}. \tag{7.5.2}$$

From (7.5.2) we see that

$$-\log p = \nu'. \tag{7.5.3}$$

Substituting in (7.5.1) for both ν' and $e^{-\nu'}$ quickly yields Macdonald's formula for the steady-state value of the sporozoite-rate

$$s = \frac{ax\, p^n}{ax - \log p}. \tag{7.5.4}$$

This can of course be measured directly in practice by dissecting the salivary glands of a representative sample of mosquitoes in an infected area.

It also follows from general biological considerations that the inoculation-rate h is given by

$$h = masb, \tag{7.5.5}$$

where on the right-hand side we are simply compounding multiplicatively the relevant factors of mosquito density, mosquito man-biting rate, sporozoite-rate and the proportion of infected mosquitoes that are actually infectious. If we substitute the value of s given by (7.5.4) into (7.5.5) we obtain

$$h = \frac{ma^2 bx\, p^n}{ax - \log p}. \tag{7.5.6}$$

Note that in both (7.5.4) and (7.5.6) we are using x for the gametocyte-

rate. If we want to retain x as the overall parasite-rate then we must replace x by xg, say, where g is the proportion of infected humans who are actually infectious (e.g. Onori & Grab, 1980a).

Formulae (7.5.4)-(7.5.6) can be used to calculate certain parameters in terms of others whose values are supposed to be known. Thus we may be able to measure m, a and s more or less directly, and can estimate h using prevalence-rate data plotted against age as in equations (7.4.5). The relatively elusive parameter b can then be found from (7.5.5). Alternatively, given estimates of all the parameters in (7.5.4) for various bodies of data, we might test the validity of the formula. For a number of practical numerical applications, see Macdonald (1952a), though it would of course be preferable to fit the Macdonald-Dietz model given by (7.4.14).

Some time-dependent analyses have also been made, e.g. Armitage (1953) and the later summary in Macdonald (1957, Appendix 1). These works could be consulted by readers wishing to go into the details, but it should be realized that many of these extensions depend on the assumptions of the Macdonald-Irwin model that are no longer acceptable. For the moment, therefore, we shall confine attention to certain steady-state and threshold results that are more generally valid.

We have already seen in Section 7.3 how a quantity R arose, in the context of Ross's equations for both human and mosquito populations, which was in fact identical with Macdonald's (1952b, 1957) *basic reproduction-rate*. In the present context Macdonald's somewhat heuristic argument runs as follows. We start from the definition of the basic reproduction-rate as the average number of secondary cases of malaria arising from a single primary case in a very large population of susceptibles.

If the human recovery-rate is r, the primary human case would be infective for r^{-1} days on average. On each day the case would be bitten on average by ma mosquitoes. The probability that any one of these mosquitoes survives for the n days of the extrinsic cycle is p^n, after which the expected time of further survival is $(-\log p)^{-1}$. On each day of this survival a mosquito will make an average of a bites, of which a proportion b would be infective. The resultant number of secondary infections, which defines the basic reproduction-rate z_0, would thus be

$$z_0 = -\frac{ma^2bp^n}{r\log p}. \qquad (7.5.7)$$

In a deterministic formulation, therefore, the condition $z_0 (\equiv R) \leqslant 1$ means that any initial malarial infection will die out, while if $z_0 > 1$ the appropriate endemic level will be reached and maintained.

The simplest approach to the implementation of control measures therefore involves any actions that tend to reduce the basic reproduction-rate z_0. Although from the above deterministic theory we try to make $z_0 < 1$, Dietz (1970) has indicated that, from a wider standpoint, such a condition might be neither sufficient nor necessary to achieve eradication within a specified period of time. If we think of the spread of infection in terms of a stochastic branching process, there will still be a positive probability of extinction for $z_0 > 1$. Conversely, even if $z_0 \leqslant 1$, the time to certain extinction may be beyond the time horizon envisaged.

Mention should also be made at this point of further formulae found in the literature referring to "the ruling reproduction rate" (Macdonald, 1957), "a net rate, applicable in all conditions" (Macdonald, Cuellar & Foll, 1968), "the *net reproduction rate*, viz. the actual number of secondary infections" (Bruce-Chwatt, 1980), and so on. Unfortunately, these indices are mostly defined and interpreted somewhat ambiguously as well as being insufficiently justified, not to mention a number of mutual inconsistencies. In addition, they are liable to be tied to the Macdonald-Irwin model and generally speaking they are best avoided. Macdonald's *basic reproduction-rate* as defined in (7.5.7) above is, however, a well-tried index with both practical implications and theoretical justification.

An important modification of Macdonald's basic reproduction-rate, defined in (7.5.7), has been made by Garrett-Jones (1964a) as follows. Macdonald's index, the average number of secondary cases of malaria arising from a single primary case in a large population of susceptibles, refers to the potentiality for the spread of disease in the context of the dynamic interaction between human and mosquito populations. Garrett-Jones proposed concentrating on the entomological aspects, defining the *vectorial capacity*, C, as the number of potentially infective contacts made by the mosquito population per case per day. In other words, C is the average number of bites on humans that the mosquitoes which have bitten an individual on any given day will distribute after the extrinsic cycle during the remaining periods of their lives. When an infected human is bitten, potentially infective bites will be distributed. Garrett-Jones recommended dropping the parameter b as involving an unnecessary refinement; he considered it more prudent to assume all bites by mosquitoes with sporozoites in their salivary glands to be capable of transmitting infection. Such an assumption could however introduce a misleading bias when applied to the overall population dynamics, especially as Pull & Grab (1974) found an upper limit of 0·026 for b (see Section 9.3 below). It follows that C, as proposed by Garrett-Jones, will be given by

$$C = -\frac{ma^2p^n}{\log p}, \tag{7.5.8}$$

thus depending only on the three basic entomological parameters m, a and p, together with the parasite's extrinsic cycle n occurring within the mosquito. If several vector species or subpopulations are associated then the total vectorial capacity, relevant to a given *Plasmodium* species, is the sum of the individual vectorial capacities. Thus if the jth population has entomological parameters m_j, a_j and p_j, we can write the overall vectorial capacity as

$$C = \sum_j m_j a_j^2 p_j^n/(-\log p_j). \tag{7.5.9}$$

Whether two (or more) separate *Plasmodium* species can be considered to act independently has been questioned by Cohen & Singer (1979) in their use of a simple Markov model, but they did not allow for the probably substantial effect of false negatives in human parasitaemia (see Section 9.2).

The vectorial capacity C has often been used for making quantitative epidemiological assessments of a mosquito population's impact *vis-à-vis* malaria, e.g. Garrett-Jones (1964a), Garrett-Jones & Grab (1964), Garrett-Jones & Shidrawi (1969), Molineaux *et al.* (1979), etc. Another term used for C is *daily reproduction-rate* since, apart from the parameter b, it corresponds to Macdonald's basic reproduction-rate considered on a daily basis.

A similar concept to the basic reproduction-rate is Moshkovskii's (1950) ratio of the "communicability" to the "exhaustibility" of malaria. Following Martini (1921), Moshkovskii (1967) specified α as the "mean number of infective bites distributed in the population and due to the presence of a case in a unit of time", while τ "corresponds to the mean probability of the disappearance (through recovery, death or departure) of an emerged case in a unit of time". The endemic level M was then represented by

$$M = 1 - \frac{\tau}{\alpha}. \tag{7.5.10}$$

Moshkovskii observed that when M became zero, malaria should disappear, but, like Martini, he was also prepared to use (7.5.10) when $\tau > \alpha$ and $M < 0$, interpreting this to indicate a higher "epidemiological security".

However, all this is essentially another, perhaps over-simplified, way of looking at the threshold results already obtained by Ross (1909), indicated in the formula (7.2.1) or (7.2.2), or Macdonald (1952b), indicated in (7.5.7) and following discussion. It seems, on the whole, that the approaches of Ross and Macdonald are more readily subjected to constructive criticism and development. We shall see in the next chapter how some of these ideas can be adapted further to obtain more realistic models that can handle such aspects as changing immunity status in humans and annual variations in mosquito populations. In particular, we shall find it very useful to introduce the notion of a time-dependent vectorial capacity $C(t)$, designated as in (7.5.9) to take care of several entomological variables simultaneously.

8 Advances in the population dynamics of malaria

8.1 Introduction

While the discussion of Chapter 6 dealt with the mathematical generalities of host-vector modelling, Chapter 7 came to grips with the specific biological and epidemiological aspects of malaria itself. We have followed the main trend of developments from the early work of Ross (1909 and later), through the extensive studies of Macdonald (1950a and later), to a more precise reformulation of the superinfection problem by Dietz (1970). Although this body of knowledge has now achieved a considerable degree of recognition, there are still a number of serious over-simplifications. The principal aspects to be incorporated are (a) the enormous seasonal variations that frequently occur in mosquito populations associated with climatic changes, and (b) the development of human immunity in response to repeated infection, plus a gradual loss of immunity with the passage of time (see end of Section 2.5).

These aspects were of course well known to Ross (e.g. 1911) and Macdonald (e.g. 1957), as we should expect from highly experienced malariologists well versed in a wide variety of real-life field situations. But no satisfactory means was found of incorporating such variable factors in the quantitative studies they undertook. It was hoped however that, in spite of neglecting these and other undoubtedly substantial influences, approximately valid insights into major phenomena would be forthcoming. There seems to be little doubt that this expectation was justified. Nevertheless, as understanding increases, the potentialities for guiding the choice of strategies for malaria control become more apparent. And this demands better and more realistic models, and entails a more critical review of the available approaches. Questions of parameter estimation and model evaluation become more acute. The incorporation of additional factors of importance thus becomes unavoidable.

Of course, there is always the question of whether, and to what extent, we should include probabilistic elements into the basic modelling. Stochastic aspects were in fact discussed quite explicitly in Sections 6.3

and 6.4 of the chapter on general host-vector theory, with some lesser references in Section 6.6. Again, it is clear from Ross's (1911) early work that he was quite well aware of the relevance of probability considerations, e.g. the appeal to the Poisson distribution in connexion with the "Theory of Happenings" referred to above in Section 7.4, although the methodology for dealing in detail with such matters was not then available. In any case, it was supposed, with some justification, that straightforward deterministic treatments would provide good insights into the major mechanisms of transmission.

Similarly, most of Macdonald's work was carried out using relatively simple deterministic forms of analysis, as in Sections 7.4 and 7.5, although a Poisson distribution was used for the distribution of contacts between mosquitoes and infected humans. More importantly, the true status of the Macdonald-Irwin model of superinfection is most easily appreciated by reinterpreting the model in terms of a stochastic queueing process (as in Section 7.4), while the biologically more credible Macdonald-Dietz version can be readily couched in terms of a stochastic homogeneous immigration-death process (also as in Section 7.4). Again, a number of questions to do with equilibria and thresholds (see Section 7.5) are conveniently examined in terms of the distribution of the time to extinction of an appropriate stochastic branching process.

An important advance in stochastic modelling occurred with the publication of the paper of Macdonald, Cuellar & Foll (1968), shortly after Macdonald's death. The model was based on the Macdonald-Irwin version of superinfection in humans, but incorporated probabilistic transitions for the inoculation and recovery of humans. Other aspects of the model were left in deterministic form. Various sophistications were introduced so as to provide an adequate level of biological realism, as well as opportunities for including seasonal variations in basic reproduction rates and periodic curative mass treatment. It is not surprising that anything as complicated as this led to the use of a simulation approach. A considerable amount of computerized analysis was undertaken in relation to practical methods of control in specific field areas. One of these involved endemic malaria in the Kankiya District in Northern Nigeria, while another related to an epidemic outbreak in the neighbourhood of Damascus, Syria.

A problem that began to receive more serious attention than previously was how to fit a given model to a specific field situation. Earlier work by Ross and Macdonald had drawn on parameter estimates from whatever relevant data might be available. These sources were often heterogeneous

and therefore led to quantitative pictures that were at best typical, though unlikely to be specific for any actual population anywhere. Ronald Ross himself was constantly careful to emphasize the need to use mathematical arguments for clarifying thought and arriving at general conclusions, rather than achieving precise descriptions and predictions.

The new model of Macdonald, Cuellar & Foll (1968) required four principal epidemiological parameters, and the authors used the following approach. It was supposed that three parameters, the man-biting rate of the mosquito, the probability of the mosquito surviving through a day, and the human recovery-rate, were sufficiently well known. A large number of computer runs were then made for a range of values for the fourth parameter, the reproduction-rate, and a value chosen for which calculated parasitaemia curves best matched the observed ones. While this idea was a big step forward, more attention to genuine statistical fitting was obviously required.

Unfortunately, the discussion presented in the paper does not, in general, make sufficiently clear exactly what numerical conclusions follow from any specific set of assumptions. With a stochastic set-up we of course want to know about the stochastic means and variances of the different variables, the distribution of the time to extinction of the infection, statistical comparisons between intervention strategies, etc. In addition it is essential to be clear about the validation of the modelling, as discussed earlier in Section 5.4.

Apart from these methodological difficulties, a careful analysis of the Kankiya data by Nájera (1974) suggested there was insufficient epidemiological information available to specify model parameters with the accuracy required. There were, moreover, gross discrepancies between estimates of the reproduction rate obtained by different formulae. A thorough revision of the mathematical modelling was therefore required, especially in regard to the incorporation of immunity effects.

At the same time the paper of Macdonald, Cuellar & Foll (1968) is well worth studying by anyone who wants to understand the problems of developing a model that is eventually to be of practical value. The way was paved for the more realistic approach of Dietz (1971) and Dietz, Molineaux & Thomas (1974), described in Section 8.2, as well as the subsequent work of Dutertre (1976) in Section 8.3, and Nåsell (1980a, b, c) in Section 8.4.

Mention should also be made of the review of Macdonald's work by Rao, Vig & Agarwala (1974a, b). Although these discussions were described as stochastic, they were in fact deterministic so far as the

mathematical analysis went. One or two modifications and elaborations were introduced, such as a mosquito population growth-rate and a more explicit distinction between sickness and infectiveness in humans. A subsequent computer simulation by Rao *et al.* (1975) based on this model did incorporate probabilistic elements and was therefore of a stochastic nature. The effect of treatment was also included by supposing this to enhance the recovery-rate. Parametric values based on data from the Bombay city area were used. Tables were presented showing computed distributions of the time required for malaria eradication to occur under different treatment strategies. The extent to which the simulation model actually fits the real life situation is however unclear.

Finally, we must examine the model of Békéssy, Molineaux & Storey (1976) which is essentially a stochastic version of the model used by Ross (see beginning of Section 7.4) to describe the course of human infections. Although the new formulation has a number of limitations, e.g. with regard to superinfection, it is of considerable importance in the development of techniques for the statistical estimation of parameters. Accordingly, we introduce the basic aspects of the process envisaged in Section 8.5, reserving discussion of specific estimation problems until Section 9.2.

8.2 The Dietz–Molineaux–Thomas model

In conformity with the remarks made at the end of the previous section, many subsequent developments in malaria modelling have been directed towards the handling of specific field situations, especially with regard to parameter estimation and goodness-of-fit testing. This implies, of course, that such models must include major factors influencing the population dynamics of the disease, such as the development of immunity or seasonal variations in the mosquito population density. From the point of view of general *eradication* theory it may be unnecessary to introduce certain aspects like the effect of immunity on transmission. But for *control* theory such features may be crucial, since here we are concerned with the possibility of achieving a new balance between host and parasite populations within some specified period of time.

Thus the deterministic model described in Dietz (1971) was developed in close co-operation with epidemiologists, entomologists and immunologists in connection with a field project involving specific villages in the Garki District of Kano State, Northern Nigeria. Data were in process of collection, and were expected to provide a basis for testing the model.

The latter assumed that superinfection had no influence on the chance of recovery in the human population, but the notion of several immunity classes was included. Thus both infected and uninfected individuals were divided into K immunity classes. The infected classes differed from each other in respect of both infectivity and recovery-rates, while the susceptible classes varied with regard to susceptibility. Similarly, the mosquitoes were divided into K classes differing in infectivity, depending on the immunity class of humans from which they were infected. Allowance was also made for the presence of several different species of mosquitoes.

So long as an individual was infected, he could pass from one immunity class to the next class with a higher level of immunity. Similarly, a recovered individual could pass from one immunity class to the next with a lower level of immunity. Births and deaths were also included, newborns entering the class with lowest immunity.

A description of the mosquito population dynamics was also developed using several parameters of entomological importance.

Mathematical details (for these, see Dietz, 1971) are not given here as the model was later superseded (Dietz, Molineaux & Thomas, 1974). The latter authors noted that the earlier model gave a good fit to the yearly average age-distribution of malaria prevalence, but could not reproduce in a satisfactory manner the seasonal fluctuations later observed in two places chosen for testing. Appropriate modifications of the model were therefore indicated. Superinfection was reintroduced, and the immunity structure was made simpler and at the same time rather more specific.

For the human population, Fig. 3 shows the essence of the model adopted. The symbols x_i and y_i, defined in detail below, refer to the *proportions* of the human population in the relevant epidemiological classes. In this scheme individuals are called *positive* (y_1, y_2 and y_3) only if they have parasites in the blood. Those with parasites in the liver only (x_2 and x_4) are said to be *incubating*, while those with no parasites in the body at all are *negative* (x_1 and x_3). Of the positive persons, a distinction is made between those that are *infectious* (y_1) and those that are *non-infectious* (y_2 and y_3).

It is assumed that newborns enter the class of "non-immune negatives", having a proportion x_1, at constant rate δ. The death-rate from all classes is also δ. The inoculation-rate h defines the rate of transfer to an "incubating" class with proportion x_2, followed after a fixed incubation period of length N by conversion to the class of "infectious positives", having proportion y_1, at a rate Q to be defined below. Infectivity is then lost at

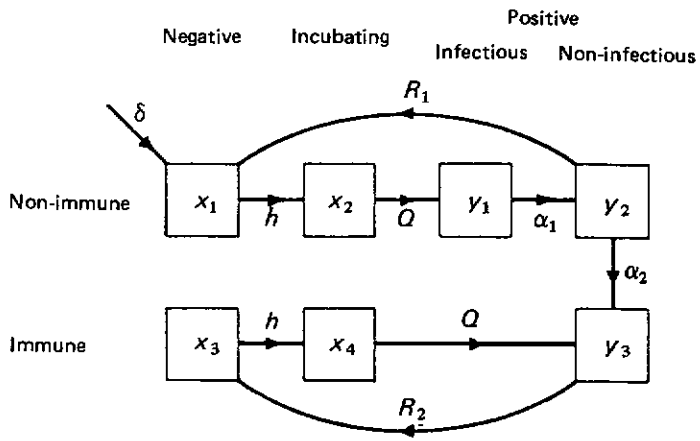

Fig. 3 Epidemiological model of malaria

Arrows show directions of transfer. Associated symbols are relevant transfer-rates (see text for definitions). Symbols in boxes indicate population proportions at a given time. All classes also have a death-rate δ. (Based on Dietz, Molineaux & Thomas, 1974.)

a constant rate α_1, and individuals transfer to the class of "non-infectious positives", having proportion y_2. From here they may lose their immunity, recover, and go back directly to the "non-immune negatives" at rate R_1 (defined below). Alternatively, they may pass to another class of "non-infectious positives", having proportion y_3, at a constant rate α_2. These individuals retain their immunity and transfer to the class of "immune negatives", with proportion x_3, at rate R_2 (defined below). They also keep a high recovery-rate from any reinfections that occur. The latter take place at rate h, and result in individuals moving to the class of "incubating immunes", with proportion x_4, at rate h. A return to the class of "non-infectious positives" then occurs after an incubation period of length N, at rate Q (defined below).

With the above definitions we can immediately write down deterministic equations specifying the dynamics of the human population. There are, however, two further points to be borne in mind at this stage. First, the assumption of a constant incubation period of length N leads to the inclusion of a time-lag in Q, and the associated variables x_1 and x_3. The time-variable t is omitted below, except where a lagged value of $t-N$ is required. Secondly, from a computational point of view it is more convenient to work with a difference equation format. We thus have

$$\left.\begin{aligned}
\Delta x_1 &= \delta + R_1 y_2 - (h+\delta)x_1,\\
\Delta x_2 &= hx_1 - Qx_1(t-N) - \delta x_2,\\
\Delta x_3 &= R_2 y_3 - (h+\delta)x_3,\\
\Delta x_4 &= hx_3 - Qx_3(t-N) - \delta x_4,\\
\Delta y_1 &= Qx_1(t-N) - (\alpha_1+\delta)y_1,\\
\Delta y_2 &= \alpha_1 y_1 - (\alpha_2 + R_1 + \delta)y_2,\\
\Delta y_3 &= \alpha_2 y_2 + Qx_3(t-N) - (R_2+\delta)y_3,
\end{aligned}\right\} \qquad (8.2.1)$$

where the quantities Q, R_1 and R_2 have to be defined more explicitly. Note also that all the variables x_i and y_i, and all parameters except for δ, α_1 and α_2, are taken to be time-dependent.

Consider now the transfers from the x_2 group to the y_1 group. These are individuals who were infected at time $t-N$ when in the "non-immune negative" group, with proportion $x_1(t-N)$ and inoculation-rate $h(t-N)$. With a death-rate of δ, the proportion of newly infected individuals who survive the incubation period of N days is approximately $(1-\delta)^N$. Hence the quantity transferring to the "non-infectious positive" class in unit time is

$$(1-\delta)^N h(t-N)x_1(t-N),$$

i.e.

$$Q = (1-\delta)^N h(t-N). \qquad (8.2.2)$$

Next we have to derive R_1 and R_2 on the basis of the occurrence of superinfection. In general, ignoring suffixes for the moment, we adopt the basic assumptions of the detailed discussion of the Macdonald–Dietz superinfection model in Section 7.4. Thus, h is the inoculation-rate, irrespective of how many broods an individual already has, and r is the elimination-rate for any given brood. Let us assume that h varies sufficiently slowly with time for the equilibrium results to apply. The proportion f_j of individuals harbouring exactly j broods then follows a Poisson distribution with parameter h/r, obtained from (7.4.13) when $t \to \infty$. For present purposes we want an expression for the average rate at which positive individuals become negative. In fact, transfers of this type occur only from the class with one brood to the class of susceptibles. In time Δt the proportion of the population so transferred is $rf_1\Delta t$, but the total proportion of positives is $1-f_0$. Hence the required transfer-rate R from positive to negative

status is given by

$$R = \frac{rf_1}{1-f_0} = \frac{h}{e^{h/r}-1}, \qquad (8.2.3)$$

since $f_0 = e^{-h/r}$ and $f_1 = (h/r)\,e^{-h/r}$. Thus for R_1 and R_2 we can write

$$R_1 = \frac{h}{e^{h/r_1}-1}, \qquad R_2 = \frac{h}{e^{h/r_2}-1}. \qquad (8.2.4)$$

The parameters r_1 and r_2 may be regarded as fixed, though they will need to be estimated. On the other hand, in the context of equations (8.2.1) we must consider the inoculation-rate to be a function of time, i.e. $h(t)$. This will depend to a very marked degree on the behaviour of the mosquito population, which may easily undergo drastic seasonal fluctuations. Fortunately, most of the relevant information about the mosquito population can be combined into a single variable $C(t)$ called the *vectorial capacity* (Garrett-Jones, 1964a; Garrett-Jones & Shidrawi, 1969), previously defined above in (7.5.8). Thus C depends on entomological parameters and the length of the extrinsic cycle, but not on the parasite-rate or sporozoite-rate.

Next we derive an approximate expression for the inoculation-rate. It is clear from the definition of $C(t)$ that the average number of potentially successful contacts, which the infectious positives existing on day $t-n$ make on day t and thereafter, is $C(t-n)y_1(t-n)$. For a fairly short-lived vector it is not unreasonable to assume that all the latter bites occur on day t itself. If we further suppose that these bites have a Poisson distribution, the probability that any given susceptible receives at least one infectious bite is

$$1 - \exp\{-C(t-n)y_1(t-n)\}.$$

We now postulate a susceptibility parameter g, so that

$$h(t) = g[1 - \exp\{-C(t-n)y_1(t-n)\}]. \qquad (8.2.5)$$

This formulation entails a strongly density-dependent aspect. While $h(t)$ is linearly related to C for small values of the latter, high vectorial capacities cause the inoculation-rate to reach a saturation level.

Thus, given the average annual pattern of the vectorial capacity $C(t)$ (which may in fact vary appreciably from year to year), we can use

equations (8.2.1) to simulate the transmission dynamics of malaria in a given population.

As a refinement, Dietz, Molineaux & Thomas (*loc. cit.*) introduce the notion of the detectability of parasites in the blood of a positive individual, according to the class to which he belongs. Probabilities q_i are associated with y_i, $i = 1, 2, 3$, where it can be assumed that $q_1 = q_2 > q_3$.

After an initial period of operation to allow the attainment of equilibrium, the computer program used will produce a representation of the seasonal variations in any variable of interest. Of special importance are the daily inoculation-rate $h(t)$, the observed proportion of positives,

$$z = \sum_{i=1}^{3} q_i y_i,$$

the true proportion of positives,

$$y = \sum_{i=1}^{3} y_i,$$

the proportion of positives that are actually infectious, i.e. y_1/y; etc.

Dietz, Molineaux & Thomas also use the following ingenious device to compute the age-specific values of any variables investigated, assuming that the chosen pattern of vectorial capacity goes on repeating itself periodically. After equilibrium is achieved for the whole population as described above, the yearly pattern for the inoculation-rate is applied to a cohort of newborns, conveniently represented by an initial $x_1 = 1$. The time parameter can now be interpreted as the age of the cohort. Moreover, the inoculation-rate in this application is given, so the vectorial capacity is not involved as an input parameter. Finally, the parameter δ is put equal to zero, since the cohort is by definition not augmented by birth, and age-specific rates of the type mentioned above are unaffected by death.

The above theoretical approach was applied to field data collected from a total of 16 villages (totalling approximately 5000 individuals) involved in a WHO malarial research project in the African savannah in the Garki District of Kano State, Northern Nigeria. The most complete entomological, parasitological and epidemiological observations were collected from eight of the villages in five surveys at 10-week intervals during 1971, and extensive studies of this baseline material have been made. However, the paper of Dietz, Molineaux & Thomas (*loc. cit.*) is only an interim report dealing with the possibility of fitting a model to the baseline data in order to try to predict the prevalence and incidence of the parasite (primarily

Plasmodium falciparum in this case) from the entomological observations. Subsequent investigations deal with the control phase of the project, involving intervention measures such as the use of larvicides, adulticides or drugs, alone or in combination.

For a detailed statistical discussion the reader is referred to the paper quoted. In the present account we shall merely give a broad indication of the results obtained. Two villages were selected on the basis of showing the most extreme values of vectorial densities. They were used to examine whether the two epidemiological situations could be adequately simulated assuming the same model and the same parameters, the only difference being the two widely different levels of vectorial capacity.

In order to make the estimation problem more manageable, at least to begin with, the number of parameters to be estimated was reduced by assuming reasonable values for some of them, namely:

δ = 0·0001 per day per person. This corresponds to annual birth- and death-rates of 36·5 per thousand, with an expectation of life of about 27 years.

r_2/r_1 = 10. This ratio of elimination was regarded as reasonable, leaving only one of the parameters to be estimated, r, say.

q_1, q_2 = 1; q_3 = 0·7. Only the class of non-infectious positives retaining their immunity was considered to have a probability of detection less than unity.

n = 10 days. This is the approximate length of the extrinsic cycle, i.e. the incubation period in the mosquito.

N = 15 days. The approximate incubation period in humans.

α_1 = 0·002 per day. The rate at which infectivity is considered to be lost in the group of infectious positives.

These assumptions leave α_2, g and r_1, to be estimated from the parasitological data, given the values of $C(t)$ drawn from the entomological observations. A minimum-χ^2 technique was adopted, using the CERN computer program MINROS. Because of the large-sample equivalence of the minimum-χ^2 and maximum-likelihood approaches, this computerized method also yields the relevant standard errors. The estimates obtained were

$$\left.\begin{aligned} \hat{\alpha}_2 &= 0{\cdot}000\,19 \pm 0{\cdot}000\,01, \\ \hat{g} &= 0{\cdot}097 \pm 0{\cdot}017, \\ \hat{r}_1 &= 0{\cdot}0023 \pm 0{\cdot}0005. \end{aligned}\right\} \qquad (8.2.6)$$

These results were derived from a total of 52 data points, the actual minimum value of χ^2 being 53·5. Since three parameters have been estimated, the number of degrees of freedom is 49. Thus, so far as the investigation went, the goodness-of-fit was entirely satisfactory. However, more searching analyses by the authors in question were planned to be carried out later and are mentioned below.

An important means of evaluating the general malaria situation in any given community is to determine the critical vectorial capacity C^*, below which the disease cannot maintain itself at an endemic level. This quantity C^* can be calculated from the relationship that holds when the basic reproduction rate $z_0 = 1$. In the present context we can obtain the average number of secondary cases generated from a single primary infectious case as follows. The average period during which a case is infectious is $(\alpha_1 + \delta)^{-1}$. During this time, for small vectorial capacities, (8.2.5) gives the approximate number of successful contacts per unit of time as gC, of which $gC(1-\delta)^N$ survive to the end of the incubation period. Hence

$$C^* = (\alpha_1 + \delta)/\{g(1-\delta)^N\}. \tag{8.2.7}$$

For the parametric values assumed or estimated in the above application, we obtain 0·021 contacts per day as the critical vectorial capacity.

In order to obtain a broad qualitative impression of the intensity of effort required to achieve some degree of malaria control, the model can be used to compute the yearly average crude parasite-rate as a function of the yearly average vectorial capacity. The village with the highest vectorial capacities showed average values of about 0·2 during the dry season, and anything up to 30 during the wet season. The yearly average was 8. It could be shown that the latter value would have to be reduced by a factor of more than 170 in order to lower the yearly average crude parasite-rate to half its original value.

Subsequent work by the same authors (Molineaux, Dietz & Thomas, 1978) has addressed itself to further evaluation of the foregoing Garki model in two specific field areas. First, the model, previously fitted to one year of baseline data, was tested against data collected in the same area over a three-year period. For one-half of this period the insides of certain houses in a part of the area were sprayed with propoxur. Secondly, the model was tested against data from Kisumu, in Nyanza Province, Kenya. The period of time was again three years, but this included 20 months when the insides of houses in part of the area were sprayed with fenitrothion.

The evaluation involved calculating the vectorial capacity from entomological data collected under the new conditions, and using this as

input to the Garki model while retaining all the original baseline estimates of the other parameters. Comparisons were then made between the predicted and observed values of the prevalence of *P. falciparum* parasitaemia.

Agreement turned out to be relatively good. A simple form of sensitivity analysis (see Chapter 11) was also carried out to test dependence of the results on the vectorial capacity, since estimates of the latter were subject to large errors. Simulations were repeated for vectorial capacities 10 times larger and 10 times smaller. Generally speaking, this had relatively little effect on the results.

8.3 The model of Dutertre

It is obvious that further developments of the type of modelling described in the previous section are urgently needed. To be convincing, models must be sufficiently realistic with regard to the salient epidemiological, parasitological, immunological and entomological factors. At the same time excessive complexity must be avoided if really efficient instruments are to be constructed for assisting the practical choice of intervention strategies.

Comparatively little new work along the lines required has appeared in recent years, but some useful suggestions occur in the deterministic simulation model of Dutertre (1976). This is particularly important because of its treatment of immunity. The basic model envisages four main compartments for the human population: susceptibles, infectious individuals with gametocytes in their blood-streams, non-infectious individuals infected only with non-gametocytic parasites, and immunes. Two forms of immunity are recognized. The first is a natural immunity considered to be heterogeneously spread in the population and represented by a negative-binomial distribution of new infections (as opposed to a Poisson distribution). The second is an acquired immunity which can be gradually lost. The first form is accounted for by suitably modifying the expression for the infection-rate, while the second is incorporated by appropriate transfers to and from the compartment containing immunes. As presented, the model does not involve the human incubation period explicitly, this being justified on the grounds of a comparatively long iteration-step in the time scale (one month). However, this parameter as well as other neglected aspects could be included explicitly for greater realism if it were thought necessary.

The mosquito population consists of just two compartments: susceptibles and infectives. In the model described, the simplifying

assumption of mosquito equilibrium is used so as to employ Macdonald's formula, as in (7.5.4) above, for the proportion of infectious mosquitoes with sporozoites in their salivary glands. This restriction could also be removed by a suitably extended model.

Dutertre (*loc. cit.*) gives some numerical results with a view to illuminating certain field data from Bobodioulasso, Haute Volta. Most parametric values were assigned in advance on the basis of general epidemiological knowledge, but two key parameters in the negative binomial distribution mentioned above were chosen so as to minimize the differences between observed and calculated incidence-rates.

Far more work is needed to validate the above model, either in the simplified form actually published or in some more realistic modified version. In particular, an extensive mathematical investigation is required of the quantitative and qualitative behaviour of the model. For this purpose it is highly desirable to distinguish the structure of the model, using purely algebraic formulations and derivations, from the implications of specific numerical parameter choices such as occur in the paper. The novel use of a negative binomial distribution of incidence to account for a heterogeneous distribution of natural immunity is also worth special study. If it turns out, on closer inquiry, to be really justified then it would be interesting to try incorporating it in the Dietz-Molineaux-Thomas model of Section 8.2.

8.4 The hybrid models of Nåsell

As we have already seen, most work prior to the simulation modelling of Macdonald, Cuellar & Foll (1968) was purely deterministic in treatment, although occasional interpretations were made in probabilistic terms. The first mathematical discussion of a genuine stochastic formulation appeared in the critical review of Dietz (1970), particularly in connexion with the Macdonald-Dietz death-immigration model of superinfection. Subsequent work by Dietz (1971) and Dietz, Molineaux & Thomas (1974) returned to purely deterministic specifications, albeit in an improved form. The simulation model of Dutertre (1976) was also deterministic, though the modifications required to provide stochastic transitions could be made very easily.

Recent studies by Nåsell (1980a, b) have revived the stochastic approach. The employment of deterministic models, as a first approximation to the handling of mean values in more sophisticated stochastic models, is always open to doubt, especially when the processes concerned

are nonlinear. This can still be true even when population sizes are quite large. See, for example, the work of Bartlett (1956) on the occurrence of fade-out in measles, mentioned at the end of Section 6.6, where stochastic modelling could still be important in towns with as many as 250 000 inhabitants.

There is, therefore, good reason to take the stochastic modelling of malaria seriously. Nåsell's (1980a) first investigation was confined to a re-examination of Ross's original model, but reformulating this with stochastic transitions for infection and recovery in humans, and infection and death in mosquitoes. The infection-rates, as usually defined in the malaria literature for both humans and mosquitoes, are then stochastic variables. Nåsell replaces these by their mean values, thus giving rise to the "hybrid" character of the model by including both deterministic and stochastic ingredients. While certain results obtained are essentially the same as those of Ross, e.g. the age-prevalence curve, interesting developments can be made in respect of the so-called incidence and recovery probabilities, which also depend on age.

There is a close analogy with Ross's model, which was introduced near the beginning of Section 7.3, and discussed again in Section 7.4 in the context of the "Theory of Happenings". The possibility of superinfection is for the time being ignored. We shall use Nåsell's formulation and notation as follows. First, it is typographically more convenient in the present discussion to employ a subscript "1" to refer to the human population, and a subscript "2" for the mosquito population (instead of the presence and absence of primes, as earlier).

If there are N_1 humans and N_2 mosquitoes, the model consists essentially of the collection of $N_1 + N_2$ Markov chains given by

$$\{X_1^{(1)}(t), X_1^{(2)}(t), \ldots, X_1^{(N_1)}(t); X_2^{(1)}(t), X_2^{(2)}(t), \ldots, X_2^{(N_2)}(t)\}, \quad t \geqslant 0. \tag{8.4.1}$$

Each $X_i^{(k)}(t)$ can take the value "0" to represent a susceptible state and the value "1" to represent an infected state. We also define the transition probabilities

$$P_{mn,i}(s, t) = P\{X_i^{(k)}(t) = n \,|\, X_i^{(k)}(s) = m\}, \text{ all } k;$$
$$i = 1, 2; m, n = 0, 1; 0 \leqslant s \leqslant t. \tag{8.4.2}$$

The form indicated in (8.4.2) shows the transition probabilities as depending on the host or vector and state involved, but not on the individual within a given category.

We now write $h_1(t)$ and $h_2(t)$ for the average human and mosquito infection-rates, respectively, at time t; while r_1 and r_2 are the human recovery-rate and mosquito death-rate, respectively. Human deaths and mosquito recoveries are ignored. Note that r_i is independent of t. It is supposed, for convenience, that on the death of any mosquito, the latter is replaced by the birth of another uninfected mosquito. Thus the death of an uninfected mosquito results in no change of state in the corresponding state variable, while the death of an infected mosquito causes a transition from "1" to "0". Replacing s by t and t by $t + \Delta t$ in (8.4.2) then gives the basic transition probabilities, to first order in Δt, as

$$P_{01,i}(t, t+\Delta t) = h_i(t)\,\Delta t,\ P_{10,i}(t, t+\Delta t) = r_i \Delta t; \quad i = 1, 2. \tag{8.4.3}$$

In particular, we must also have

$$\left.\begin{aligned} P_{01,i}(s, t) + P_{00,i}(s, t) &= 1, \\ P_{10,i}(s, t) + P_{11,i}(s, t) &= 1. \end{aligned}\right\} \tag{8.4.4}$$

By considering transition probabilities over the time-interval $(s, t+\Delta t)$ in terms of the two steps (s, t) and $(t + \Delta t)$, we easily obtain

$$P_{01,i}(s, t + \Delta t) = P_{01,i}(s, t)(1 - r_i \Delta t) + P_{00,i}(s, t)\, h_i(t)\, \Delta t,$$

with an analogous expression for $P_{10,i}(s, t)$. Using (8.4.4) and taking $\Delta t \to 0$ gives

$$\left.\begin{aligned} \frac{\partial P_{01,i}(s, t)}{\partial t} &= h_i(t) - \{r_i + h_i(t)\}\, P_{01,i}(s, t), \\ \frac{\partial P_{10,i}(s, t)}{\partial t} &= r_i - \{r_i + h_i(t)\}\, P_{10,i}(s, t), \end{aligned}\right\} \tag{8.4.5}$$

with initial values

$$P_{01,i}(s, s) = 0 = P_{10,i}(s, s). \tag{8.4.6}$$

An explicit solution of (8.4.5) and (8.4.6) is easily found by standard methods. Let us define functions $\tau_i(t)$ and $\sigma_i(t)$, given by

$$\left.\begin{aligned} \tau_i(t) &= r_i t + \int_0^t h_i(u)\, du, \\ \sigma_i(t) &= \int_0^t h_i(u) \exp\{\tau_i(u)\}\, du. \end{aligned}\right\} \tag{8.4.7}$$

It readily follows that

$$P_{01,i}(s, t) = \{\sigma_i(t) - \sigma_i(s)\} \exp\{-\tau_i(t)\}, \tag{8.4.8}$$

and

$$P_{10,i}(s, t) = 1 - \exp\{\tau_i(s) - \tau_i(t)\} - P_{01,i}(s, t). \tag{8.4.9}$$

We are also interested in the state probabilities at any given time. The probabilities of infection of humans and mosquitoes can be written as $p_1(t)$ and $p_2(t)$, respectively, where

$$p_i(t) = P\{X_i^{(k)}(t) = 1\},\ \text{all } k; \qquad i = 1, 2; \quad t \geqslant 0. \tag{8.4.10}$$

Now it is easily seen that

$$p_i(t) = \{1 - p_i(0)\}\, P_{01,i}(0, t) + p_i(0)\{1 - P_{10,i}(0, t)\},$$

by summing the probabilities of reaching an infected state at time t, given that the initial state is uninfected with probability $1 - p_i(0)$ and infected with probability $p_i(0)$. Substitution from (8.4.8) and (8.4.9) then gives

$$p_i(t) = \{p_i(0) + \sigma_i(t)\} \exp\{-\tau_i(t)\}. \tag{8.4.11}$$

It follows from (8.4.11) and the definitions of τ_i and σ_i in (8.4.7) that the infection probabilities $p_i(t)$ must satisfy the differential equations

$$\frac{dp_i(t)}{dt} = h_i(t) - \{r_i + h_i(t)\}\, p_i(t); \quad i = 1, 2. \tag{8.4.12}$$

The equation in (8.4.12) for $i = 1$ is clearly analogous to (7.4.1), which gives a differential equation for the proportion of human infectives x in the simplest deterministic Ross model where h is assumed to be constant.

Let us now return to the parameter definitions used at the beginning of Section 7.3 in the Ross model covering both human and mosquito populations, the latter now represented in Nåsell's notation by the subscripts "1" and "2". The population sizes are of course N_1 and N_2. The proportions of infected individuals who are infectious are f_1 and f_2. The human recovery-rate γ is now r_1, while the joint birth- and death-rates in mosquitoes, μ' and ν', are r_2. The man-biting rate in mosquitoes is b_2. As before, we can ignore the human birth- and death-rates and set the mosquito recovery-rate to zero.

Using arguments similar to those in Section 7.3, we can reason as follows. The rate at which a given mosquito bites a given human is b_2/N_1. The rate $h_1(t)$ at which a given human susceptible becomes infected is

$f_2 b_2/N_1$ times the number of infected mosquitoes. The latter is a stochastic variable with mean value $N_2 p_2(t)$, the use of which entails the "hybrid" aspect of the model. We thus have

$$h_1(t) \doteq f_2 b_2 N_2 p_2(t)/N_1 = r_1 T_1 p_2(t), \tag{8.4.13}$$

where the transmission factor T_1 defined in (7.3.9) is now

$$T_1 = \frac{f_2 b_2 N_2}{r_1 N_1}. \tag{8.4.14}$$

Similarly, the mosquito infection-rate $h_2(t)$ is approximately $f_1 b_2/N_1$ multiplied by $N_1 p_1(t)$, giving

$$h_2(t) \doteq f_1 b_2 p_1(t) = r_2 T_2 p_1(t), \tag{8.4.15}$$

where the transmission factor T_2 defined in (7.3.9) is now

$$T_2 = \frac{f_1 b_2}{r_2}. \tag{8.4.16}$$

Using (8.4.13) and (8.4.15) allows us to rewrite the equations (8.4.12) in the form

$$\left.\begin{aligned} \frac{dp_1}{dt} &= r_1 T_1 p_2(1-p_1) - r_1 p_1, \\ \frac{dp_2}{dt} &= r_2 T_2 p_1(1-p_2) - r_2 p_2. \end{aligned}\right\} \tag{8.4.17}$$

These equations are of course the counterparts of the deterministic Ross model equations described by (7.3.3), at least when the latter are written in terms of proportions y/n and y'/n'.

As before, in Section 7.3, a steady state is given by

$$\left.\begin{aligned} P_1(T_1, T_2) &= \frac{T_1 T_2 - 1}{T_1 T_2 + T_2}, \\ P_2(T_1, T_2) &= \frac{T_1 T_2 - 1}{T_1 T_2 + T_1}, \end{aligned}\right\} \tag{8.4.18}$$

corresponding to (7.3.10). When $T_1 T_2 \leqslant 1$ any infection introduced into the community will ultimately die out. But if $T_1 T_2 > 1$ even a small amount of infection will cause the prevalence to rise to the stable endemic level (P_1, P_2) indicated by (8.4.18).

In terms of directly estimated values of the prevalences given by $\hat{P}_1$ and $\hat{P}_2$, we can easily rearrange (8.4.18) to give estimated values of $\hat{T}_1$ and $\hat{T}_2$ as

$$\hat{T}_1 = \frac{\hat{P}_1}{\hat{P}_2(1-\hat{P}_1)}, \qquad \hat{T}_2 = \frac{\hat{P}_2}{\hat{P}_1(1-\hat{P}_2)}, \tag{8.4.19}$$

corresponding to the previous result in (7.3.11).

Arguments similar to those of Section 7.4 can be used to find the steady-state prevalence of infection in the human population. We put $p_2 = P_2$ in the first equation of (8.4.17), namely

$$\frac{dp_1}{dt} = r_1 T_1 P_2(1-p_1) - r_1 p_1, \tag{8.4.20}$$

and solve for initial condition $p_1(0) = 0$. The solution $Q(s)$ in terms of a time variable s can be written as

$$Q(s) = P_1[1 - \exp\{-(r_1 + r_1 T_1 P_2)\, s\}], \tag{8.4.21}$$

$$= \frac{H_1}{r_1 + H_1}[1 - \exp\{-(r_1 + H_1)\, s\}], \tag{8.4.22}$$

where H_1, called the *steady-state public health factor* by Nåsell, is the steady-state value of $h_1(t)$ given by

$$H_1 = r_1 T_1 P_2. \tag{8.4.23}$$

The form (8.4.22) is virtually identical with the age-prevalence distribution obtained by Ross in (7.4.3), but is now established in relation to a more explicitly defined underlying biological process.

Nåsell (*loc. cit.*) also defines the *incidence probability* given by

$$I(s, T) = P_{01,1}(s, s+T), \tag{8.4.24}$$

which is the probability that an individual who is uninfected at age s will be infected at age $s + T$; together with the *recovery probability* given by

$$R(s, T) = P_{10,1}(s, s+T), \tag{8.4.25}$$

which is the probability that an individual who is infected at age s will be free of infection at age $s + T$.

Under steady-state conditions it is easily shown that

$$I(s, T) = Q(T), \tag{8.4.26}$$

where $Q(T)$ is given by (8.4.21) or (8.4.22), and

$$R(s, T) = (P_1^{-1} - 1)\, Q(T). \tag{8.4.27}$$

Thus the incidence and recovery probabilities are independent of the age of the human host in this model, and increase monotonically with T to the values P_1 and $1 - P_1$, respectively.

Nåsell also discusses various aspects of malaria control in the context of the above model, using in particular the notion of *control efficiencies*. This subject will be pursued further in Chapters 10 and 11.

It should be noted that much of the above analysis, described in more detail by Nåsell (1980a), can be short-cut by using standard methods (e.g. Bailey, 1964) of deriving a partial differential equation for the joint moment-generating function of the random variables Y_1 and Y_2, say, representing the total number of infected humans and infected mosquitoes, respectively. The use of the "hybrid" assumption means that the infinitesimal transition probabilities for infection, e.g. $h_1(t)(N_1 - Y_1)\,\Delta t$, are only *linear* functions of either Y_1 or Y_2, no terms in $Y_1 Y_2$ appearing as in the full stochastic model. This linearity leads in the usual way (by picking out coefficients in moment-generating function expansion) to equations for the stochastic means which are identical with the equations for the corresponding deterministic values.

Thus if we write the stochastic means as

$$m_1(t) = E(Y_1), \qquad m_2(t) = E(Y_2), \tag{8.4.28}$$

we can immediately form the equations

$$\left.\begin{aligned} \frac{dm_1}{dt} &= h_1(t)(N_1 - m_1) - r_1 m_1, \\ \frac{dm_2}{dt} &= h_2(t)(N_2 - m_2) - r_2 m_2. \end{aligned}\right\} \tag{8.4.29}$$

Since

$$m_1(t) = N_1 p_1(t), \qquad m_2(t) = N_2 p_2(t), \tag{8.4.30}$$

we can substitute (8.4.30) in (8.4.29), and use the definitions (8.4.13) and (8.4.15) for $h_1(t)$ and $h_2(t)$, to yield equation (8.4.17) directly. The steady-state solutions (8.4.18) and steady-state age prevalence distribution (8.4.22) then follow as before.

Nåsell's (1980b) second investigation extended his ideas outlined above to the superinfection situation, using the basic concepts of the Macdonald-Dietz model for the human population, but treating the mosquito population as before.

This work is somewhat involved, and interested readers should consult the original paper. We can, however, use the short-cut approach via a moment-generating function to obtain at least some of Nåsell's results very quickly. Let us again write Y_1 and Y_2 for the random variables representing the total numbers of infected individuals in populations of N_1 humans and N_2 mosquitoes, respectively. As in Section 7.4 we now distinguish the numbers of humans harbouring different numbers of broods, e.g. $Y_{1j}, j \geqslant 1$ for those with exactly j broods. Thus

$$\sum_{j=1} Y_{1j} = Y_1.$$

It is then convenient to put $j = 0$, with Y_{10} to refer to the number of susceptibles. We then have $Y_1 + Y_{10} = N_1$. Let the stochastic means be

$$m_{1j}(t) = E(Y_{1j}),\ j \geqslant 0; \quad m_1(t) = E(Y_1); \quad m_2(t) = E(Y_2). \tag{8.4.31}$$

The mosquito equation is unaffected by the superinfection assumption, and will therefore be the same as the second line of equation (8.4.29). The human equation, on the other hand, will turn out to be closely analogous to (7.4.12). This is because all infinitesimal transition probabilities will be no more than linear in any of the stochastic variables. For example, the probability of the transition $Y_{1j} \to Y_{1j} + 1$ in the interval Δt is simply $h_1(t)\, Y_{1,j-1}\Delta t + r_1(j+1)\, Y_{1,j+1}\Delta t$, the two terms corresponding to one new infection occurring in the group Y_{j-1} or one recovery occurring in the group Y_{j+1}, respectively. Picking out linear coefficients in the joint moment-generating function for the variables Y_{1j} and Y_2 will therefore lead, as before, to equations for stochastic means which are identical with those for deterministic values, namely

$$\left.\begin{aligned} \frac{dm_{1j}}{dt} &= h_1 m_{1,j-1} - (h_1 + r_1 j)\, m_{1j} + r_1(j+1)\, m_{1,j+1}, \quad j \geqslant 0; \\ \frac{dm_2}{dt} &= h_2(N_2 - m_2) - r_2 m_2, \end{aligned}\right\} \tag{8.4.32}$$

where for convenience we assume that $m_{1,-1} \equiv 0$, and h_1 and h_2 are time-dependent functions given, as before, by (8.4.13) and (8.4.15). Let us now write

$$m_{1j}(t) = N_1 p_{1j}(t),\ j \geqslant 0; \quad m_1(t) = N_1 p_1(t); \quad m_2(t) = N_2 p_2(t). \tag{8.4.33}$$

Equations (8.4.32) then take the form

$$\left.\begin{aligned} \frac{dp_{1j}}{dt} &= h_1 p_{1,j-1} - (h_1 + r_1 j)\, p_{1j} + r_1 (j+1)\, p_{1,j+1}, \quad j \geqslant 0; \\ \frac{dp_2}{dt} &= h_2 (1 - p_2) - r_2 p_2, \end{aligned}\right\} \tag{8.4.34}$$

where $p_{1,-1} \equiv 0$. To find the steady-state solutions, P_{1j} and P_2, we put the differential coefficients in (8.4.34) equal to zero. The steady-state equations are therefore

$$\left.\begin{aligned} & H_1 P_{1,j-1} - (H_1 + r_1 j)\, P_{1j} + r_1 (j+1)\, P_{1,j+1}, = 0, \quad j \geqslant 0; \\ & H_2 (1 - P_2) - r_2 P_2 = 0, \end{aligned}\right\} \tag{8.4.35}$$

where

$$H_1 = r_1 T_1 P_2, \qquad H_2 = r_2 T_2 P_1. \tag{8.4.36}$$

Rearranging the first line of (8.4.35), and combining the equations for values of j decreasing to zero, gives

$$\begin{aligned} r_1 (j+1) P_{1,j+1} - H_1 P_{1j} &= r_1 j P_{1j} - H_1 P_{1,j-1} \\ &= r_1 (j-1) P_{1,j-1} - H_1 P_{1,j-2} \\ &= \ldots \\ &= r_1 P_{11} - H_1 P_{10} \\ &= \qquad - H_1 P_{1,-1} \\ &= 0. \end{aligned}$$

We now obtain, almost immediately,

$$P_{1j} = (H_1/r_1)^j P_{10}/j!, \quad j \geqslant 0. \tag{8.4.37}$$

Since

$$\sum_{j=0} P_{1j} = 1,$$

it follows that $P_{10} = \mathrm{e}^{-H_1/r_1}$, and thus

$$P_{1j} = \mathrm{e}^{-H_1/r_1} (H_1/r_1)^j/j!, \quad j \geqslant 0, \tag{8.4.38}$$

namely, a Poisson distribution with parameter H_1/r_1.

In particular, we have

$$P_1 = 1 - P_{10} = 1 - \mathrm{e}^{-H_1/r_1} = 1 - \mathrm{e}^{-T_1 P_2}.$$

Combining this with the second line of (8.4.35), where in the latter we substitute the value of H_2 given in (8.4.36), then yields a pair of equations exhibiting the transmission factors T_1 and T_2 as functions of P_1 and P_2, namely

$$T_1 = -\frac{\log (1 - P_1)}{P_2}, \qquad T_2 = -\frac{P_2}{P_1 (1 - P_2)}. \tag{8.4.39}$$

This result is identical with that given by Nåsell (1980b, equations (5.3) and (5.4)), and allows T_1 and T_2 to be estimated from observed values of the parasite- and sporozoite-rates P_1 and P_2, respectively.

The steady-state age-distribution of prevalence in the human population can now be found rather easily, following the same approach as before. We put H_1 for h_1 in the first line of (8.4.34), to give

$$\frac{dp_{1j}}{dt} = H_1 p_{1,j-1} - (H_1 + r_1 j)\, p_{1j} + r_1 (j+1)\, p_{1,j+1}, \quad j \geqslant 0, \tag{8.4.40}$$

and solve with the initial conditions

$$p_{10}(0) = 1; \qquad p_{1j}(0) = 0, \quad j \geqslant 1. \tag{8.4.41}$$

Since (8.4.40) and (8.4.41) are identical in form with the previously discussed equations (7.4.12), we have the immediate result that the parasite-rate $Q(s)$ in terms of a time variable s can be written as

$$Q(s) = 1 - \exp\{-(H_1/r_1)(1 - \mathrm{e}^{-r_1 s})\}, \tag{8.4.42}$$

thus providing a neat extension of the expression given earlier in (7.4.14) for the Macdonald-Dietz model. However, the quantity H_1 is a function of P_2, and the latter can be obtained from (8.4.39) as a function of T_1 and T_2, though not in an explicit form.

If we put $U = H_1/r_1 = T_1 P_2$, we can write $P_2 = U/T_1$ in (8.4.39) and then eliminate P_1. We find that U must satisfy

$$T_2 (T_1 - U)(1 - \mathrm{e}^{-U}) = U. \tag{8.4.43}$$

Results (8.4.42) and (8.4.43) are identical with those of Nåsell (1980b, equations (5.21), (5.19), (4.2) and the definition in the paragraph following (4.12)). By writing (8.4.43) as $T_2 T_1 - T_2 U = U/(1 - \mathrm{e}^{-U})$ and graphing both sides, it is easily shown that for $U > 0$ a solution exists only for $T_1 T_2 > 1$.

8.5 The stochastic model of Békéssy, Molineaux & Storey

Mention should be made at this point of a model used by Békéssy, Molineaux & Storey (1976), called by these authors a "reversible two-state catalytic model" with reference to the work of Muench (1959). However, it is in fact a fully stochastic version of the Ross model described at the beginning of Section 7.4, and is relatively simple in its conception. This work seems to have attracted little attention, partly because the model is not presented in a very explicit manner, and partly because there are *no* references to any part of the available epidemiological literature. The main relevance of the discussion is to the estimation of human incidence- and recovery-rates. This aspect will be taken up again in Section 9.2.

Essentially, all we have to do is rework the paragraph preceding equation (7.4.1), using the approach introduced at the beginning of Section 6.3 on stochastic epidemics. For example, we can represent the actual number of individuals infected at time t by the random variable $X(t)$. Assume that the total population is of size n, so that the number of susceptibles is $n - X(t)$. There are just two types of transition to be considered, corresponding to infection and recovery. We take the probability of one new infection, with $X(t)$ increasing by one unit, occurring in time-interval Δt to be $h(n - X)\,\Delta t$, and the probability of one recovery, with $X(t)$ decreasing by one unit, to be $rX\Delta t$.

Let us write the probability that $X(t)$ takes the value j at time t as

$$\Pr\{X(t) = j\} = p_j(t), \quad j = 0, 1, \ldots, n. \tag{8.5.1}$$

The probability-generating function for $X(t)$ is defined as

$$P(z, t) = \sum_{j=0}^{n} p_j(t)\, z^j. \tag{8.5.2}$$

It immediately follows, by a standard method (e.g. Bailey, 1964, sect. 7.4) as in Section 6.3, that $P(z, t)$ satisfies the partial differential equation

$$\frac{\partial P}{\partial t} = (1 - z)(r + hz)\frac{\partial P}{\partial z} + nh(z - 1)P. \tag{8.5.3}$$

If we further assume the process to start at $t = 0$ with $X(0) = a$ infected individuals, the initial condition is

$$P(z, 0) = z^a. \tag{8.5.4}$$

These equations are readily solved directly (or as a special case of the equation describing a general birth-death-immigration process, e.g.

Bailey, 1964, sect. 8.7, with $\lambda = -h$, $\mu = r$, $\nu = nh$, and $e^{\theta} = z$), the subsidiary equations being

$$\frac{dt}{1} = \frac{dz}{(z - 1)(r + hz)} = \frac{dP}{nh(z - 1)P}. \tag{8.5.5}$$

Two easily found intermediate integrals are

$$\frac{z - 1}{r + hz}\, e^{-(r+h)t} = \text{const.}, \qquad P(r + hz)^{-n} = \text{const.} \tag{8.5.6}$$

The general solution may therefore be written as

$$P(z, t) = (r + hz)^n \Phi\left(\frac{z - 1}{r + hz}\, e^{-(r+h)t}\right), \tag{8.5.7}$$

where Φ is an arbitrary function to be determined from the initial conditions.

Putting $t = 0$ in (8.5.7) and using (8.5.4) gives

$$z^a = (r + hz)^n \Phi\left(\frac{z - 1}{r + hz}\right). \tag{8.5.8}$$

Hence, with $w = (z - 1)/(r + hz)$, we can identify the function Φ by

$$\Phi(w) = (r + h)^{-n}(1 + rw)^a(1 - hw)^{n-a}. \tag{8.5.9}$$

Applying this to (8.5.7) gives the required solution

$$P(z, t) = (r + h)^{-n}\{(r + hz) + r(z - 1)\, e^{-(r+h)t}\}^a \times \{(r + hz) - h(z - 1)\, e^{-(r+h)t}\}^{n-a}. \tag{8.5.10}$$

Such expressions can usually be manipulated to yield valuable information about the process involved. For example, if we want the endemic prevalence-rate of infection in terms of age, this will be given by

$$x(t) = m(t)/n, \tag{8.5.11}$$

where $m(t)$ is the mean value of $X(t)$, assuming an initial value of $a = 0$. We have

$$m(t) = \left.\frac{\partial P}{\partial z}\right|_{z=1} = \frac{nh}{r + h}(1 - e^{-(r+h)t}),$$

or

$$x(t) = \frac{h}{r+h}(1 - e^{-(r+h)t}), \tag{8.5.12}$$

which corresponds exactly to the deterministic value for Ross's model in (7.4.3), as we should expect for a stochastic process in which the transition probabilities are only linear functions of the random variables involved.

Of course it will be realized that the use of a stochastic death-immigration process, as discussed in Section 7.4 for handling the Macdonald-Dietz model, is probably preferable since it allows the incorporation of a biologically more credible superinfection mechanism.

It may therefore not be worth discussing the present stochastic model in detail since it is only an extension of the earlier non-superinfection Ross version. However, since it has been used by Békéssy *et al.* (1976) for approximate estimation purposes this aspect will be looked at explicitly in Section 9.2. In addition, it should be noted that these latter authors give, in the Appendix to their paper, relatively simple methods of analysing the transition probabilities of the underlying Markov process which is suited to their particular application. The brief account given above presents a rather more general view of the process as a whole.

9 Statistical estimation problems

9.1 Introduction

The approach of this book, especially as expounded in Chapter 5 on the theory and practice of modelling, is that models must be validated, not only through the clarification of theory but also by their power to explain observed phenomena. While Chapter 6 dealt deliberately with the general theory of host-vector diseases, the following Chapters 7 and 8 were explicitly devoted to the elucidation of malaria itself. Emphasis was laid on understanding basic ideas, and on the attempts that have been made to improve realism and clear up various sources of confusion. Practical data were referred to wherever these were available to disprove or confirm existing themes, e.g. Macdonald's (1950a) introduction of the superinfection concept. For detailed analysis of numerical data the reader should refer to the technical epidemiological literature cited.

A major problem here is that much of the data analysis that has been carried out has not been clearly based on sound statistical principles, or at least it cannot be *seen* that such principles have been followed. Moreover, it is often not stated precisely what epidemiological models have been assumed. And it is not infrequent for authors to seize on a formula merely because it is convenient to use, without considering the implicit epidemiological assumptions. The result is that it is difficult to know when numerical conclusions are well-founded, when they are only speculative, and when they are unlikely to be of value.

Whether sophisticated or simple approaches are used, the basic assumptions should be clearly stated, parameters and their standard errors should be estimated by methods of proven value, the goodness-of-fit between hypotheses or models and observational material should be tested where possible, major conclusions should be examined for their sensitivity to possible changes in parametric values (or even model structure), etc. While all this is of course a counsel of perfection, it is highly desirable to regard it as a standard to aim at. The work of Dietz, Molineaux & Thomas (1974) is a good example of the application of such principles.

On the whole, the works of Ross and Macdonald are well documented, and careful reading allows one to identify points of difficulty and disagreement. In this way improvements and new advances are promoted. But, unfortunately, many authors of otherwise excellent and useful papers dealing with the analysis of field data fail to achieve their full impact, partly because the underlying assumptions and the principles followed are not visible, and partly because little reference is made to the publications of other research workers. Indeed, a good deal of the relevant literature is quite chaotic in these matters.

The fitting of a relatively complex model, like that of Dietz, Molineaux & Thomas, is often not possible in full detail, but valuable partial information can frequently be obtained by simpler methods. The credence to be given to the latter depends both on the statistical accuracy achieved, e.g. in terms of standard errors, and on the acceptability of the assumptions. Evidence may arise from entomological data, human data, or a combination of the two. We shall examine in the following sections some of the methods that have been tried with malariological data, emphasizing the aspects that are specific to malaria. It is not of course appropriate in the present context to go deeply into either biostatistical details or laboratory and field techniques. For routine biostatistical methods the reader should refer, if necessary, to standard textbooks. But in specific applications the biological and malariometric aspects will usually have to be discussed *ad hoc* with the investigators involved. The following treatment is accordingly intended only as a general guide to a range of important statistical problems that constantly arise in practical applications.

9.2 Human parameters

Although in any detailed analysis it is usually necessary to take account of the dynamic interaction between the human and mosquito populations, within certain limits we can consider these separately. In Section 7.4 we examined at some length different models purporting to describe the processes of infection and superinfection in humans. It is probably best, for the reasons given, not to attempt using the Macdonald–Irwin model. Thus in any fairly elementary analysis we can appeal to the Ross model if superinfection is excluded, or the Macdonald–Dietz model if it is considered significant.

One of the commonest ways of presenting epidemiological data is in the form of age-dependent prevalence curves for human parasitaemia, such as those defined by the formula in (7.4.3) or (7.4.14). In the simplest

treatments these arise in a deterministic context. But if the stochastic version has infinitesimal transition probabilities which are only linear functions of the variables involved then we obtain the same curves for the stochastic means. This is certainly the case for the original deterministic model of Ross and for the stochastic version of the latter used by Békéssy *et al.*, as described in Section 8.5. It is also true for the Macdonald–Dietz model, as already pointed out in Section 7.4, as well as for the hybrid models of Nåsell examined in Section 8.4. Generally speaking, therefore, we might as well use the stochastic formulations as being rather more realistic from a biological point of view, especially when many observed samples of data are not very large.

In some situations it may also happen that certain individuals can be followed longitudinally. If we could determine the exact time of infection, e.g. to the nearest day, we would simply have a set of observations drawn from a negative-exponential distribution with parameter h. If the mean time to infection is $\bar{t}$, then the standard maximum-likelihood estimate of h is $(\bar{t})^{-1}$ with large-sample variance h^2/n, i.e.

$$\left.\begin{aligned} \hat{h} &= (\bar{t})^{-1}, \\ \operatorname{var}(\hat{h}) &= h^2/n. \end{aligned}\right\} \tag{9.2.1}$$

More frequently, however, we are not sure of the precise day of infection, when it occurs, but can say at the end of a given period, e.g. one month, that an individual has become infected for the first time at some point of time during the month. This means that clinical and other circumstantial evidence are used to exclude the possibility that an infected person recovers and returns unrecognized to the susceptible condition. The probability of infection occurring for the first time in some fixed period of length τ is

$$p(\tau) = 1 - \mathrm{e}^{-h\tau}. \tag{9.2.2}$$

If we observe a sample of n negative individuals and find that a of them have become positive for the first time during a fixed interval τ, application of the standard maximum-likelihood procedure quickly gives

$$\left.\begin{aligned} \hat{h} &= -\frac{1}{\tau}\log\left(1-\frac{a}{n}\right), \\ \operatorname{var}(\hat{h}) &= \frac{a}{n(n-a)\tau^2}. \end{aligned}\right\} \tag{9.2.3}$$

The results in (9.2.1) or (9.2.3) obviously hold, irrespective of whether we choose the Ross or the Macdonald-Dietz model.

As already noted in Section 7.4, Macdonald (1950a, b) fitted the Ross age-dependent prevalence curve shown in (7.4.3) to actual data on observed parasite-rates in children. He stated that "The mathematical working of Ross (1916) was used and though theoretical curves could be produced which corresponded well with the observed, their production demanded in every case the derivation of constants which were clearly inapplicable" (Macdonald, 1950b). As previously indicated, values of r were obtained as low as 0·001 per day, or less, implying an average duration of parasitaemia of around three years. The exact numerical work was not shown, although there is little reason to doubt its essential validity, at least in approximate terms.

A considerable amount of parasitological and entomological data, drawn from a WHO field trial to test a new insecticide in Kisumu, Nyanza Province, Kenya, has been presented and analysed in some detail by Pull & Grab (1974). For full particulars, the reader should consult the original paper. But for our present purposes it is sufficient to record that blood specimens from about 650 infants were collected once every three months, and 200 thick-blood fields on each slide were examined to detect the presence of malaria parasites. Tables were given showing quarterly results by age, as well as consolidated prevalence-rates by age. Cohorts of about 45 infants were also examined early in each month and followed up longitudinally every month until found to be positive for the presence of parasites. Aggregated results for just over 500 infants enabled observed age-specific incidence-rates to be presented.

The method adopted for analysing the age-specific incidence and prevalence curves was to fit an appropriately chosen reversible catalytic model in the manner of Muench (1959). (For further examples of Muench-type modelling applications, see Payne, Grab, Fontaine & Hempel, 1976; and Onori & Grab, 1980b.) Although Pull & Grab (*loc. cit.*) give references to the work of Ross, Macdonald and Dietz *et al.*, the exact status of the Muench catalytic model is not explicitly discussed: as applied, it is in fact equivalent to the Ross model we have previously analysed in Section 7.4 (although allowing for the occurrence of false negatives by including an additional factor k equal to the proportion of infected infants actually detected in any prevalence survey).

It is therefore, at first sight, not surprising that Pull & Grab obtain estimates $h = 0{\cdot}0084$ and $r = 0{\cdot}0007$ (with $k = 0{\cdot}70$), where r is very much lower than the value 0·005 previously accepted. The estimated recovery-

rate 0·0007 corresponds to an average duration of parasitaemia of nearly four years, a point only referred to by the authors in remarking "As is to be expected, the recovery rate is practically negligible in the first year of life." The appearance of such a low recovery-rate was of course the point of departure adopted by Macdonald when he rejected the Ross formulation in favour of a superinfection model. The latter concept is, however, not discussed or even mentioned by Pull & Grab.

Obviously, it is preferable to use an explicit statistical method of estimating parameters, like maximum-likelihood, that is not only statistically efficient but also provides standard errors, as well as allowing goodness-of-fit tests when there is a sufficient number of degrees of freedom. In their Table 2, Pull & Grab (1974) give consolidated monthly incidence data of the kind that can be analysed by formulae (9.2.2) and (9.2.3). Since the incidence of infection occurs as a random process, all the observations made on individuals during periods of $\tau = 30$ days (i.e. one month) are independently distributed, irrespective of whether these observations are made on different individuals or different periods for the same individual. Ignoring data for the first month of life, which may well be affected by the length of the incubation period and the possibility of partial immunity due to maternal antibodies, we find a total of 1282 observations on infants who were negative at the beginning of an interval. Of those 1282 a total of 300 were considered to have become infected for the first time during the period in question. (It would incidentally be worth undertaking further investigations of how to make a proper allowance for the effect of the incubation period on observed data, instead of merely ignoring data where the effect is probably greatest.)

We therefore put $n = 1282$, $a = 300$ and $\tau = 30$ in (9.2.3), giving the estimate

$$\hat{h} = 0{\cdot}00889 \pm 0{\cdot}00051, \tag{9.2.4}$$

which may be compared with the Pull & Grab figure of 0·0084, from which it does not differ significantly.

In Table 1 of Pull & Grab (1974) results are also given for each of four prevalence surveys carried out at three-monthly intervals, and analysed by age at monthly intervals up to one year of age. Because the same population was being repeatedly surveyed, an obvious cohort effect is involved and the separate surveys are not entirely independent. For example, children in the first month of life at the first survey could reappear in the fourth month of life at the second survey, and so on. However, for illustrative purposes we shall use data aggregated over all four

surveys (as done by Pull & Grab). The relevant figures are shown in Table 9.1.

TABLE 9.1 Aggregated age-specific prevalence-rates for malaria in infants under one year of age (based on data from Pull & Grab, 1974)

Age interval (i)	Centre-point of interval in days (t_i)	Number of infants examined (n_i)	Number found to be positive (a_i)	Observed prevalence rate as a percentage	Expected prevalence (see text) (p_i)
1	45	202	26	12·9	21·7
2	75	235	76	32·3	32·0
3	105	210	91	43·3	39·9
4	135	210	93	44·3	46·0
5	165	243	140	57·6	50·6
6	195	219	121	55·3	54·2
7	225	219	128	58·4	57·0
8	255	241	135	56·0	59·1
9	285	199	121	60·8	60·7
10	315	204	118	57·8	61·9
11	345	205	132	64·4	62·9

For any specific age t the probability of an initially susceptible newborn being found positive for malaria parasites is given by equation (7.4.3) for the Ross model and (7.4.14) for the corresponding Macdonald-Dietz model incorporating superinfection, namely

$$\left.\begin{aligned} x(t) &= \frac{h}{r+h}(1-\mathrm{e}^{-(r+h)t}) && \text{(Ross)},\\ x(t) &= 1-\exp\{-(h/r)(1-\mathrm{e}^{-rt})\} && \text{(Macdonald-Dietz).}\end{aligned}\right\} \quad (9.2.5)$$

It was noted in Section 7.4 that the latter expression for the Macdonald-Dietz model was equally valid for a suitable defined stochastic version, while a similar conclusion was established for the Ross formula in Section 8.4.

As indicated above, Pull & Grab thought it advisable to consider the possibility of infected infants failing to be recognized, writing k for the actual chance of detecting a true positive. If this additional factor is included we can replace (9.2.5) by

$$\left.\begin{aligned} p(t) &= \frac{hk}{r+h}(1-\mathrm{e}^{-(r+h)t}) && \text{(Ross)},\\ p(t) &= k[1-\exp\{-(h/r)(1-\mathrm{e}^{-rt})\}] && \text{(Macdonald-Dietz),}\end{aligned}\right\} \quad (9.2.6)$$

where $p(t)$ is the probability of actually recording a positive result.

The easiest way to proceed, if a sufficiently powerful computer maximization/minimization package is available, is to maximize the joint likelihood of the data directly.

Let us suppose that the incidence data, as already discussed above, consist of a_0 positive individuals out of a total of n_0. The probability of being positive at the end of an interval τ is

$$p_0(\tau) = 1-\mathrm{e}^{-h\tau}, \quad (9.2.7)$$

using (9.2.2).

Similarly, let the prevalence data recorded at time t_i consist of a_i positives out of a total n_i individuals examined, where $i = 1, 2, \ldots, v$. The probability of being positive at time t_i is p_i, where the latter is obtained from (9.2.6) according to the model chosen with $p_i \equiv p(t_i)$.

It is easily seen that the overall log likelihood L is, except for a constant independent of the parameters,

$$L = \sum_{i=0}^{v} \{a_i \log p_i + (n_i - a_i)\log(1-p_i)\}. \quad (9.2.8)$$

Application of a standard package like MINUIT or the CERN program MINROS will give the required results.

This approach has been applied to the incidence data previously quoted and the prevalence data in Table 9.1. Data for the first month of life were ignored because of the effect of the incubation period and possible immunity due to maternal antibodies. Thus $\tau = 30$ and each $t_i = 30i + 15$ for $i = 1, \ldots, 11$. If a large computer is not available a suitably adapted maximum-likelihood scoring procedure could be used. In this case it would be worth observing that, in the case of the Ross model, it can be shown that all the information on h comes from the incidence data, so that these can be conveniently analysed independently. Given the relevant value of $\hat{h}$, we can then analyse the prevalence material to yield $\hat{r}$ and $\hat{k}$.

Analysis of the data presented above, using MINROS and carried out by Mr A. S. Thomas, gives

$$\left.\begin{aligned} \hat{h} &= 0{\cdot}00889 \pm 0{\cdot}00059,\\ \hat{r} &= -0{\cdot}00050 \pm 0{\cdot}00148,\\ \hat{k} &= 0{\cdot}637 \pm 0{\cdot}078,\end{aligned}\right\} \quad (9.2.9)$$

where the first line of (9.2.9) is virtually the same as (9.2.4). (The standard errors of $\hat{h}$ should agree exactly in both cases. The difference is probably due to the computations in MINROS using numerical differentiation to

arrive at the covariance matrix.) It is interesting to observe that $\hat{r}$ is not significantly different from zero. This fits in with certain lines of epidemiological thinking, in which true recovery in infants under one year is considered negligible, although the possibility of false negatives is admitted due to misclassification of slides or purely parasitological latency, e.g. the remark of Pull & Grab, quoted on p. 146 above.

If r can safely be assumed zero over the infant age-range considered, we must repeat the maximum-likelihood calculations with $r = 0$. The estimates turn out to be

$$\left.\begin{aligned} \hat{h} &= 0{\cdot}00881 \pm 0{\cdot}00054, \\ \hat{k} &= 0{\cdot}662 \quad \pm 0{\cdot}024, \end{aligned}\right\} \tag{9.2.10}$$

where of course $\hat{h}$ has changed only slightly.

It is now important to examine the goodness-of-fit. The required function, relevant to the $v+1$ groups involved in (9.2.8), is

$$\chi^2_{v-1} = \sum_{i=0}^{v} \frac{(n_i\hat{p}_i - a_i)^2}{n_i\hat{p}_i(1-\hat{p}_i)}, \tag{9.2.11}$$

but this is supplied automatically by MINROS. In our example there are 12 independent data points (one from the incidence data and 11 taken from the prevalence data of Table 9.1, where the figure for age 0–1 month was omitted). We thus have 12 degrees of freedom in all, but have estimated the two parameters $\hat{h}$ and $\hat{k}$, leaving 10 degrees of freedom. We find $\chi^2_{10} = 11{\cdot}4$, an entirely satisfactory value. It should, however, be noted that the main contributions come from the classes centred on 45 days and 165 days in the prevalence data.

In this case there is no advantage in examining the Macdonald-Dietz version in the second line of (9.2.6), since for small r it is indistinguishable from the Ross model.

So far then we may conclude that, using a stochastic interpretation of the basic Ross model, satisfactory results can be obtained from the joint analysis of incidence and prevalence data over the first year of life. There is no evidence of any appreciable recovery-rate, but we do have to allow for the clearly non-zero probability of false negatives.

Another, slightly more recent, approach is that of Békéssy, Molineaux & Storey (1976). This has already been introduced in Section 8.5, where a general discussion is given of the model used. The latter is in fact a stochastic version of the Ross model we have previously examined in

Section 7.4, although Békéssy *et al.* simply refer to a "reversible two-state catalytic model". It is assumed that repeated surveys of a population can be carried out, and that individuals examined in one survey can be identified and re-examined in the next. The data collected can be used to provide efficient statistical estimates of both h and r.

Specifically, we can suppose that in the first of a pair of consecutive surveys there are N_0 susceptible individuals, who are negative for the parasite, and N_1 infected individuals, who are positive for the parasite. In the subsequent survey the individuals are re-examined. Of the original N_0 susceptibles, N_{00} are recorded as negative and N_{01} are positive. Similarly, of the original N_1 infected persons, N_{10} are now negative and N_{11} are positive. Let us write α and β for the observed transition frequencies, namely

$$\alpha = N_{01}/N_0, \qquad \beta = N_{10}/N_1, \tag{9.2.12}$$

which are simply the proportions of negatives and positives observed to have changed their status at the time of the second survey.

Since the stochastic model, as discussed in Section 8.5, is clearly multiplicative in that the epidemiological history of each individual can be followed independently, we can simplify matters by putting $n = 1$ in the probability-generating function given in (8.5.10). Consider now an individual who is initially susceptible, but found after an interval t to be infected. The probability of this event can be written, in an obvious notation, as $p_{01}(t)$. The converse event of an initially infected person being subsequently uninfected has probability $p_{10}(t)$. These transition probabilities can now be found immediately from (8.5.10) with $n = 1$, first putting $a = 0$ and picking out the coefficient of z, and secondly putting $a = 1$ and picking out the constant term. Thus

$$\left.\begin{aligned} p_{01}(t) &= \frac{h}{r+h}(1 - e^{-(r+h)t}), \\ p_{10}(t) &= \frac{r}{r+h}(1 - e^{-(r+h)t}). \end{aligned}\right\} \tag{9.2.13}$$

It is evident that the random variables N_{01} and N_{10} follow independent binomial distributions for samples of sizes N_0 and N_1, with parameters p_{01} and p_{10}, respectively. Maximum-likelihood estimates of the latter are therefore given directly by α and β in (9.2.12). Using (9.2.13), we can then find maximum-likelihood estimates of h and r by solving

$$\left.\begin{aligned} \frac{h}{r+h}(1-e^{-(r+h)t}) &= \alpha, \\ \frac{r}{r+h}(1-e^{-(r+h)t}) &= \beta. \end{aligned}\right\} \tag{9.2.14}$$

That those simple equations actually provide maximum-likelihood estimates of h and r follows from the fact that we have just two degrees of freedom amongst the observational classes and two parameters to estimate. By a slight extension of a result in Bailey (1951), maximum-likelihood solutions for the parameters h and r are given by equating the observed numbers to their expectations, i.e. $p_{01} = \alpha$ and $p_{10} = \beta$. Dividing corresponding sides of the two equations in (9.2.14), and also adding corresponding sides, then leads almost straightaway to the required solutions

$$\hat{h} = -\frac{\alpha}{(\alpha+\beta)t}\log(1-\alpha-\beta), \qquad \hat{r} = -\frac{\beta}{(\alpha+\beta)t}\log(1-\alpha-\beta). \tag{9.2.15}$$

Alternatively, we can write down the joint likelihood for the two observed samples and maximize in the usual way. This again leads to the result (9.2.15), as given previously in the reference quoted.

Standard errors or variances can then also be found by the standard procedure. The original authors calculated the relevant information matrix and inverted it. Actually, it is much quicker to form increments of both sides of the equations in (9.2.15), square, and take expectations. The quantities α and β are independently distributed with variances

$$\text{var}(\alpha) = \alpha(1-\alpha)/N_0, \qquad \text{var}(\beta) = \beta(1-\beta)/N_1. \tag{9.2.16}$$

The large-sample variances of $\hat{h}$ and $\hat{r}$ are then quickly obtained, and can be written in the form

$$\left.\begin{aligned} \text{var}(\hat{h}) &= \gamma^{-4}\{S_1(\alpha V+\beta U)^2 + S_2(\alpha U-\alpha V)^2\}, \\ \text{var}(\hat{r}) &= \gamma^{-4}\{S_2(\beta V+\alpha U)^2 + S_1(\beta U-\beta V)^2\}, \end{aligned}\right\} \tag{9.2.17}$$

where

$$\left.\begin{aligned} &\alpha = N_{01}/N_0, \quad \beta = N_{10}/N_1, \quad \gamma = \alpha+\beta, \\ &U = -t^{-1}\log(1-\gamma), \qquad V = \gamma/\{t(1-\gamma)\}, \\ &S_1 = \alpha(1-\alpha)/N_0, \qquad S_2 = \beta(1-\beta)/N_1. \end{aligned}\right\} \tag{9.2.18}$$

Békéssy *et al.* (1976) gave a suitable computer program and instructions for use with a Hewlett-Packard HP65 pocket calculator, enabling the formulae in (9.2.15), (9.2.17) and (9.2.18) to be handled very expeditiously, starting from the given observations N_{01}, N_0, N_{10} and N_1.

Extensive applications of this method were made to field data obtained from surveys carried out in Garki, Kano State, Nigeria, by a malaria research project conducted over the years 1971–73 by the Government of the Federal Republic of Nigeria and the World Health Organization. Detailed results are presented for individual pairs of successive surveys, according to different age-groups, and distinguishing between wet and dry seasons. Comparisons could be made between a baseline period using a residual insecticide, propoxur, covering about 80 weeks. Obviously, for a full analysis and discussion, reference must be made to the original paper. A typical result, quoted for infants under one year in age during the baseline period, was: $\hat{h} = 0{\cdot}0134 \pm 0{\cdot}0042$, $\hat{r} = 0{\cdot}0045 \pm 0{\cdot}0016$, these figures being obtained from averages over five pairs of surveys.

An immediate question arises as to whether these results are in conflict with those of Pull & Grab reviewed above. However, Békéssy *et al.* observe that the "apparent incidence rate includes the incidence both of new infections and parasitological relapses". This appears to mean, as before, that false negatives may appear. When the latter again become positive it is no more than a recognition of their true state, although superficially it looks like a combination of recovery and reinfection. In a similar manner, Békéssy *et al.* explain that the "apparent recovery rate includes true recovery and latency".

Given such flexibility of definition and interpretation, the Békéssy model has certain obvious advantages. The transition probabilities are easy to calculate on the basis of data which can be collected without too much difficulty in many field situations. And the assumption of constant rates over relatively short periods of time is not an unreasonable approximation. However, there is the problem that in the type of analysis formulated there is no provision for any goodness-of-fit testing. Moreover, *if false negatives are admitted*, then, in calculating the transition probabilities, proper allowance should be made for the fact that observed negatives are of two kinds: those that are genuine and those that are not. We are then committed to an additional parameter and a more intricate form of analysis, but the extra degree of freedom required for estimation is not available from the kind of data envisaged by Békéssy *et al.*

The solution to this problem is to have sufficient data on incidence, age-specific prevalence, and observed change of parasitological state, so

that a minimum of three parameters, namely h, r and k, can be properly estimated and the relevant goodness-of-fit tested.

Summary and recommendations

(i) If incidence data are collected with sufficient care from the longitudinal study of selected susceptible individuals, so that the *actual times* of the first infection can be determined with a fair degree of accuracy, then a direct estimate of h can be made from (9.2.1).

(ii) Alternatively, if the selected susceptible individuals can be studied over a fixed period, and we can be practically certain which of them have suffered a first infection *during this period*, then h can be estimated from (9.2.3).

(iii) The collection of age-specific prevalence data, of the type shown in Table 9.1, in principle admits by itself the estimation of no more than two parameters. If the chance k of correctly recognizing a true positive is unity, then we can find both h and r from the prevalence data alone. But if $k < 1$ we shall need to use additional information, such as that supplied by (i) or (ii), in order to obtain joint maximum-likelihood estimates.

(iv) Again, if $k = 1$, the method of Békéssy *et al.* permits the easy estimation of h and r. But goodness-of-fit is not testable without a more detailed analysis, and results could be misleading if k is in fact less than unity.

(v) If, as occasionally happens, we have data on the recovery of untreated cases resulting from a single infection (see Section 7.4), then the recovery-rate r can be directly estimated from the analysis of recovery times using a negative-exponential curve with parameter r.

(vi) Since published data make it likely that in general the parameters h, r and k all need to be seriously considered, it appears essential to design the collection of data so as to provide estimates of all three. The basic models are fairly simple in form, whether we use the modified Ross version where superinfection is ignored or the Macdonald–Dietz version where it is included. Specification of the relevant joint likelihood is easy, and the computation of a maximum a routine operation. Standard errors can be found and goodness-of-fit assessed.

(vii) Unless the approach in (vi), or something analogous, is followed, the analysis of parasitological data in malaria will continue to be bedevilled by major uncertainties of interpretation. This is especially serious when one wants in addition to carry out an *approximately valid* analysis to compare the results for different age-groups, different seasons of the year (e.g. wet and dry), or different control strategies (e.g. use of drugs or insecticides) versus baseline conditions.

(viii) Consideration should also be given to the recent elaborate investigations of Singer & Cohen (1980). Though they discuss the effects of false negatives, this source of error is not fully allowed for in their main analysis.

9.3 Entomological parameters

Just as the human aspects of malaria infection can be investigated to some extent in their own right, so can the entomological factors be examined independently, as already outlined in the elementary discussion of Section 7.5. It is fortunate that Macdonald's many contributions in this area can be pursued separately from the controversy that has surrounded the Macdonald-Irwin treatment of superinfection in humans. However, the multiplicity of parameters relating to the mosquito itself as well as to the mosquito phase of the parasite's development leads to several statistical difficulties. And there are in particular a large number of unsolved practical problems to do with the statistical design of actual data collection.

Let us consider first the man-biting propensity of the mosquito. Following the notation introduced at the beginning of Section 7.5, we write m for the mosquito density per human and a for the average number of humans bitten per day by any one mosquito. The average number of bites received per human per day is thus the overall human biting-rate d, say, given by

$$d = ma. \tag{9.3.1}$$

There are two principal methods of estimating d more or less directly from field data. First, a record is kept, for selected human bait in given indoor stations, of the numbers of mosquitoes actually biting. Basic observations may be considered as Poisson variables, and suitably constructed weighted averages for different times and places can be calculated with appropriate standard errors. (Statistical details are, however, rarely given in the literature.) Secondly, mosquitoes can be collected from specially selected huts after pyrethrum spraying, or from exit-traps designed to retain mosquitoes attempting their normal passage to the outside world. The female mosquitoes can then be examined to find the

numbers that have taken human-blood meals. Again it is quite possible to assign appropriate standard errors to average estimates.

In practice there are great difficulties in avoiding both uncontrollable heterogeneity and systematic errors. The use of human bait may make the huts involved more attractive to mosquitoes, and some observers seem to obtain consistently higher scores than others. Again, many mosquitoes killed by pyrethrum spraying may simply be undetected after falling to the ground. It is not surprising therefore that the alternative estimates of d may vary considerably: Pull & Grab (1974) give annual averages for daily human biting-rates of 0·573 obtained from human bait data and 0·319 obtained from mosquito collections.

The sporozoite-rate s is the proportion of female mosquitoes with sporozoites in their salivary glands. This can thus be obtained by dissection of the mosquitoes collected, and is basically a binomial variable. Unfortunately s is often relatively small: in Pull & Grab's data most values were 10 percent or less and several were 1-2 percent only. The number of dissections required to achieve an adequate coefficient of variation is thus liable to be considerable. In fact, Table 3 in Pull & Grab shows a total of over 60 000 dissections performed over 12 months for two species of mosquito. To gain a rough idea of the accuracy involved let us look at the following example.

In January 1973 some 2624 specimens of *Anopheles gambiae* were dissected to give a sporozoite-rate of 0·1079. Assuming a binomial distribution to be approximately valid, the relevant standard error is

$$\{(0{\cdot}1079)(0{\cdot}8921)/2624\}^{1/2} = 0{\cdot}0061.$$

The human bait observations totalled 253 which we might assume to be a Poisson variable with standard deviation $(253)^{1/2} = 15{\cdot}91$. The biting-rate quoted is 5·27, presumably obtained from observations on 48 individuals since $253/48 = 5{\cdot}27$. Allowing for the multiplying factor of 1/48, the standard error of the biting-rate is $15{\cdot}91/48 = 0{\cdot}331$. For the pyrethrum-spray and exit-trap collections a total of 1611 fed female mosquitoes was observed. For a Poisson variable the standard deviation is $(1611)^{1/2} = 40{\cdot}14$. This time a biting-rate is quoted as 2·17, presumably averaged over 742 human individuals at risk, since $1611/742 = 2{\cdot}17$. The standard error of the latter estimate is then $40{\cdot}14/742 = 0{\cdot}054$. A further adjustment must be made to allow for the proportion of blood-meals actually taken from humans. A total of 2903 independent blood-meals in *A. gambiae* were examined by precipitin testing and showed 95 percent as being derived from a human source, i.e. a *human blood index* of $0{\cdot}95 \pm 0{\cdot}004$ (for an extensive discussion of this index, see Garrett-Jones, 1964a). Ignoring this very small standard error, the final estimate of the biting-rate is $2{\cdot}06 \pm 0{\cdot}051$. Similar calculations must also be performed on the data given for a second abundant mosquito species, *A. funestus*. The results for January 1973 may thus be summarized as in Table 9.2, where the parameter estimates agree exactly with those in Table 3 of Pull & Grab (1974), but we have added approximate standard errors.

TABLE 9.2 Parameters estimated from entomological observations for January 1973 (based on data from Pull & Grab, 1974), with attached standard errors calculated as indicated in the text

	A. gambiae	*A. funestus*
Sporozoite-rate (s)	$0{\cdot}1079 \pm 0{\cdot}0061$	$0{\cdot}0538 \pm 0{\cdot}0046$
Biting-rate from human bait ($ma = d$)	$5{\cdot}27 \pm 0{\cdot}331$	$9{\cdot}63 \pm 0{\cdot}448$
Biting-rate from mosquito collection ($ma = d$)	$2{\cdot}06 \pm 0{\cdot}052$	$2{\cdot}19 \pm 0{\cdot}055$

If we now write h' for the so-called entomological inoculation-rate, i.e. the rate at which humans are bitten by mosquitoes carrying sporozoites in their salivary glands, we have

$$h' = mas = ds. \tag{9.3.2}$$

Thus the results in Table 9.2 are easily combined to give those appearing in Table 9.3. The standard errors of h' are found from the usual formulae for the variances of the products of random variables (see (9.3.5) below). It is evident that when more than one species of mosquito is present, but assuming a single type of malarial parasite, the several inoculation-rates must be added as shown in the third column of Table 9.3. The large discrepancy between the estimates based on different methods of assessing the biting-rate is quite evident. It is obviously highly desirable to improve the measurement of man-vector contacts, and further practical research is very necessary.

If we could, in addition, make some independent estimate of b, the proportion of mosquitoes with sporozoites in their salivary glands that are actually infectious, then we could use formula (7.5.5), namely

$$h = masb = bh', \tag{9.3.3}$$

to obtain an estimate of h based on entomological observations that we could compare with the corresponding quantity estimated from human

TABLE 9.3 Entomological inoculation-rates, human infection-rate and infectiousness of mosquitoes with sporozoites in their salivary glands, for January 1973 (based on data from Pull & Grab, 1974), with attached standard errors calculated as in text

	Entomological inoculation-rate (h')			Human infection-rate (h)	Infectiousness of mosquitoes (b)
	A. gambiae	*A. funestus*	Both species combined		
Biting-rate from human bait	0·569 ± 0·048	0·518 ± 0·050	1·087 ± 0·069	0·00881 ± 0·00054	0·00810 ± 0·00071
Biting-rate from mosquito collection	0·222 ± 0·014	0·118 ± 0·011	0·340 ± 0·017		0·0259 ± 0·0020

observations (as in Section 9.2). Unfortunately, b, called the "factor of proportionality" by Macdonald, is a somewhat elusive quantity, though of great epidemiological importance. If we accept the value of h from human data, we can estimate b from (9.3.3), i.e.

$$b = h/h'. \tag{9.3.4}$$

In the present case we could use the figure for h given in (9.2.10), quoted in the fourth column of Table 9.3. This then leads to the estimates of b shown in the final column of the table, for the two different measurements of biting-rate, the standard error being calculated from (9.3.7) below. Of course, this makes an obvious over-simplification in assuming h to be approximately constant throughout the year. Ideally, we should like to have estimates of h, d and s month by month, and then calculate the corresponding values of b. Tests could then be made to see if there was significant variation in any parameter between months. In any case, it is obvious that values of b as small as the two different estimates of 0·008 or 0·026, which we have calculated for January 1973, cannot be ignored or replaced by unity (as suggested by Garrett-Jones, 1964a). Pull & Grab (1974), working with annual averages, give a possible range for b as 0·015 to 0·026. However, very different values might well be found in other situations.

Clearly, one could pursue such statistical discussions at considerable length. In practice far too little attention is paid in the literature to achieving a minimum statistical review of data presented. Improvements in this matter would also be likely to result in better data being collected: data collection should be designed to meet the requirements of valid estimation and hypothesis testing.

We have simply outlined above one kind of analysis that may be required using particular data by way of illustration. It is not our purpose here to try and codify details. These will vary according to circumstances, and appropriate standard methods should be used. However, it should be emphasized that, wherever possible, efficient methods of parameter estimation should be used, such as maximizing the likelihood with provision of standard errors and goodness-of-fit. The methods employed above, of deriving and combining the various estimates obtained originally from binomially or Poisson distributed observations, have mostly relied on simple applications of such formulae as

$$\left.\begin{aligned} &\operatorname{var}(\lambda X) = \lambda^2 \operatorname{var}(X), \\ &\operatorname{var}\Big(\sum_i X_i\Big) = \sum_i \operatorname{var}(X_i), \\ &\operatorname{var}(XY) = \operatorname{var}(X)\operatorname{var}(Y) + \bar{X}^2 \operatorname{var}(Y) + \bar{Y}^2 \operatorname{var}(X), \end{aligned}\right\} \tag{9.3.5}$$

where the random variables X and Y, or members of the set X_i, are taken to be distributed independently.

In sufficiently large samples of data where the coefficients of variation are small we can use the following approximation. Suppose Z is a multiplicative function of the X_i given by

$$Z = \prod_i X_i^{\pm 1}, \tag{9.3.6}$$

then, representing a coefficient of variation by the function $C(\,.\,)$, we have the additive formula

$$C^2(Z) \doteq \sum_i C^2(X_i). \tag{9.3.7}$$

This approximation was used above to find the standard error of $b = h/h'$.

Needless to say, in more searching analyses of observational data more powerful and accurate methods of investigation may have to be used than those employed in the relatively heuristic approach adopted in the foregoing discussion.

Further details of infection in the mosquito were considered in Section 7.5. In addition to the parameters discussed so far in the present section, there is the length n of the extrinsic part of the parasite cycle taking place in the mosquito, the mosquito death-rate ν', and the probability of survival over a given day p. The parameters n, ν' and p have usually been taken as constant, so that $p = e^{-\nu'}$ as in (7.5.2). Experimental studies of sporozoite development (see, e.g., Macdonald, 1952a, for review and discussion) have shown that the value of n depends not only on the species of parasite involved but also markedly on temperature. Again, experimental studies on mosquito survival (e.g. Macdonald, 1952a) suggest that the negative-exponential assumption is not unreasonable, but great variations in p can be expected between species. Moreover, natural conditions may be very different from those produced experimentally. There is accordingly considerable difficulty in obtaining broadly acceptable field estimates of p (see Garrett-Jones & Shidrawi, 1969, for detailed discussion), let alone developing a reliable form of statistical analysis.

A satisfactory basis for a sound statistical approach to the estimation of parameters such as n or p, suited to specific applications, appears to be lacking. Further research is clearly needed, especially in view of the importance attached to the vectorial capacity C, given by (7.5.8) in

Section 7.5, namely

$$C = -\frac{ma^2p^n}{\log p}, \tag{9.3.8}$$

which depends entirely on entomological variables, e.g. m, a and p, plus the length of the extrinsic parasite cycle in the mosquito n.

We shall therefore not pursue this discussion here, though the status of C will come up again in the following section in connexion with its use in the Molineaux-Dietz-Thomas model previously reviewed in Section 8.2.

9.4 Combined human/mosquito modelling

So far in this chapter we have concentrated on estimating parameters for, first, the human population and, secondly, the mosquito population, assuming in each case that the infection risk from the alternative population was more or less constant. This parallels the previous discussions in Sections 7.4 and 7.5, respectively. Although this allows certain convenient simplifications, it ignores the dynamic interaction, varying with time, between humans and mosquitoes already explicitly introduced in the early model of Ross described near the beginning of Section 7.3 (e.g. in equation (7.3.3)).

The steady-state solution of (7.3.3) was indicated in (7.3.6). The latter could be given the simpler form of (7.3.10) by use of Nåsell's two transmission factors T_1 and T_2 defined in (7.3.9). Thus if $\hat{P}_1$ and $\hat{P}_2$ were observed estimates of the steady-state parasite- and sporozoite-rates, then T_1 and T_2 could be estimated as in (7.3.11). The estimates $\hat{P}_1$ and $\hat{P}_2$ would have binomial distributions, and for relatively large samples at least, the standard errors of $\hat{T}_1$ and $\hat{T}_2$ could be found by standard methods.

Formulae analogous to (7.3.10) were given later in (8.4.18), this time in the context of Nåsell's hybrid modelling. Again T_1 and T_2 could be estimated from observed steady-state values of $\hat{P}_1$ and $\hat{P}_2$. Further investigation is required to see whether such estimates could actually be used in practice as approximations in the absence of true steady states. For instance, given substantial annual variations, to what extent is it legitimate to regard the situation as approximately stable at any particular point of time?

Again, it might be worth considering some modification of the Ross equations (7.3.1) and (7.3.2). For example, suppose that the human population and its basic epidemiological and demographic parameters are held constant, while the same is assumed true for the mosquito popula-

tion, except that the birth-rate and population-size are allowed to fluctuate over the year in a predetermined manner. Steps would then be taken to reparametrize the model in terms of a smaller number of parameters, in such a way that awkward and elusive items would not appear explicitly but would be subsumed under some more general indices (like Nåsell's transmission factors) that *could* be estimated from relatively accurate observations. The population dynamics of such a model would be more easily investigated, and calculation of the consequences of changing even an implicit parameter would be facilitated.

The more sophisticated model of Dietz, Molineaux & Thomas (see Section 8.2) does in fact involve such an improved degree of biological realism, including the variable interaction between humans and mosquitoes: both human immunity and seasonal changes in mosquito density are explicitly allowed for. Although some eight parameters were given assumed values, three of them, considered to be of major importance, were jointly estimated from field data. We saw in Section 9.3 how special difficulties arose in estimating parameters from entomological data because b, the proportion of mosquitoes with sporozoites in their salivary glands that are actually infectious (see equation (9.3.3)), could be obtained only indirectly. Dietz *et al.* dealt with this difficulty by introducing a "susceptibility parameter" g, which relates the human infection-rate $h(t)$ to the probability of a susceptible receiving at least one potentially successful bite on a given day, as in equation (8.2.5). The parameter g is then one of the three items estimated jointly from the data, in fact taking the value $0{\cdot}097 \pm 0{\cdot}017$. Although, strictly speaking, g refers to human susceptibility rather than mosquito infectiousness, at least some allowance is made for the possible failure of potentially infectious bites.

It seems a useful approach therefore to try to develop further models in which as much biologically realistic behaviour as possible is represented by the smallest feasible number of parameters. However, we may have to allow for such complications as seasonal fluctuations in mosquito population, the acquisition and loss of immunity in humans, the presence of more than one species of malaria parasite, possible interactions with other coexisting parasitic diseases such as schistosomiasis, etc. The real situation may easily become extremely complex, even at the broad level of population dynamics. Any valid model will have to mirror a good deal of this complexity, although workable approximations may be feasible. And if public health decision-makers are to be assisted directly, e.g. by predictions that are sufficiently accurate to discriminate between alternative intervention strategies, then the models must properly reflect

local conditions. This implies that many key parameters must be estimated from local data. To develop really effective ways of doing this presents a considerable challenge to mathematical statisticians.

9.5 New developments in statistical estimation

At the time of writing, the main content of this section can provide little more than a strong plea for the development of new methods. The importance of parameter estimation in the context of combined human/mosquito modelling was emphasized at the close of Section 9.4, but similar remarks would apply equally to the subject-matter of the two previous Sections 9.2 and 9.3, dealing with humans and mosquitoes separately. Existing methods of model building, parameter estimation and goodness-of-fit testing, usually become progressively more cumbersome as the structure of a model grows in complexity and the essential parameters increase in number. Whether there is any escape from this difficulty remains to be seen, though there are certain indications of a possible line of advance.

Thus, in the context of infectious diseases transmissible from person to person, Becker (1976; 1977a, b; 1979a, b) has strongly advocated, and indeed developed in some detail, the application of branching processes to disease modelling. Although a start can be made with a general branching process model, any attempt to introduce the usual realistic feature of allowing the infection intensity at any time to depend on the number of susceptibles remaining in the population leads to a more complex formulation in terms of a compound regular point process. Becker then uses the approach of Aalen (1975, 1977, 1978) to construct martingales associated with the infection process and to which a method of moments can be applied.

Specifically, appropriately chosen zero-mean martingales are equated to zero and manipulated to provide estimates and standard errors of important parameters. These methods not only allow the use of various generalizations, such as assuming arbitrary distributions of infectious periods and latent periods, but also avoid the complexities and unproved conjectures involved in attempts to carry out some of the more familiar maximum-likelihood estimation of parameters (see e.g. Bailey, 1975a).

Whether such methods can be fruitfully applied to parasitic disease problems remains to be seen. We shall, therefore, not pursue details of the general approach here. However, Dr Niels Becker (personal discussions and communications) has indicated that there is a strong possibility of such

developments, and we look forward to hearing more about these in due course. As indicated at the end of the previous section, we need to be able to handle not only single parasitic diseases, with both human and vector populations, but situations involving more than one species of parasite as well as the not uncommon occurrence of coexisting and interacting parasitic diseases such as malaria and schistosomiasis.

In addition to more readily applicable methodological techniques of handling parasitic disease data, there is also an urgent felt requirement for the establishment of readily available data sets to enable mathematicians and statisticians to validate their work in real-life contexts. This point has been strongly made by Becker (1980). The essence of the matter is that most modelling work is carried out in academic institutions where contacts with epidemiologists and field data are minimal. Too much emphasis is therefore placed on purely theoretical problems, and not enough attention is paid to the needs of decision-makers dealing with actual disease-control programmes. Conversely, those who are in a position to collect objective field observations are often unaware of what data are needed to support or reject hypotheses of crucial practical importance. The World Health Organization is in a unique position for promoting a more dynamic interaction between academic methodological research and high-priority disease control in the field, and it is hoped that substantial progress will be made before long.

10 Control theory

10.1 Introduction

The principal theme underlying this book is that scientific knowledge and insight into the population dynamics of malaria can be translated into practical measures leading to the public health control of the disease, including in certain circumstances actual eradication. We have already seen in earlier chapters, especially Chapters 6, 7 and 8, how malaria modelling is being gradually improved to provide both descriptions that are sufficiently realistic in biological terms and prediction procedures that can be validated against field data. We are thus moving into a phase where forecasts of the likely consequences of different available intervention strategies will be able to assist decision-makers. Applications to specific communities may depend crucially on obtaining adequate estimates of the local values of the relevant parameters, as already discussed in Chapter 9.

In this sense, therefore, we have already been concerned with major aspects of the theory of malaria control, e.g. reducing the basic reproduction-rate or vectorial capacity to a point where the infection is held down to an acceptably low endemic level or perhaps eliminated altogether. In such cases we are adopting the view that, given alternative intervention choices, we simply try to predict the consequences of each and favour that which provides the qualitatively best scenario.

The main reason for approaching such problems with some caution is that, as with the control of communicable diseases generally, there are many factors other than immediate medical treatment and preventive measures to be taken into account (see Bailey, 1975a, b, c, d). These include nutrition, health education, human behaviour, sanitation, water supplies, economic standards, etc. This means that public health decision-making may well have to take place within the wider context of an interactive system of components, of which the malaria subsystem is only one component. Care must therefore be taken to leave the relevant decision-makers in a position to make the final choice of strategy or strategies, taking into account all the available technical guidance.

Questions then arise as to the feasibility of using the mathematical methods of control theory, typically including an appeal to Pontryagin's

Principle. The application of such ideas has in fact been developing since about 1958 (see Bailey, 1975a, b, c, d for references up to that time; and the more recent excellent survey by Wickwire, 1977). These treatments are, however, liable to be highly theoretical. It seems likely that the main contribution of control-theoretic types of analysis is in clarifying ideas and in providing general insight into the decision-making process.

There is good reason to suppose that many of the problems of malaria control should be more thoroughly investigated through the multi-disciplinary approaches of operational research, system dynamics and systems analysis, where the overall objectives are to increase understanding of a wide range of interlocking phenomena, and to make recommendations of direct practical value to planners and decision-makers. These latter topics will accordingly be taken up in more detail in Chapter 12.

In the subsequent Section 10.2 we shall outline the broad principles of control theory, and in Section 10.3 give a relatively simple example of their application by Dietz (1975) to the problem of the optimal allocation of resources in malaria control.

10.2 Principles of control theory

In order to set the scene for the outline presentation of Dietz's malaria control model in the next section, it may be useful to give a brief description of the main aspects of modern control theory in defining the components of an optimal control problem and looking for its solution. For a more extensive discussion, including a wide range of applications to the control of pests and infectious diseases, see the authoritative review by Wickwire (1977).

To begin with, the dynamic system under investigation must be specified by a *model* which identifies the *transition mechanisms* involved. It is assumed that the operation of the system generates certain measurable *costs* or *rewards*. Initially the system is not deliberately controlled, although subject to the constraints inherent in its structure. We then specify various interventions or *control actions* which may change the operation of the system as well as entailing additional costs. Rules prescribing what control action is to be chosen at any time constitute the *control policy*. If this policy uses only the *current state* of the system it is referred to as a *closed-loop* or *feedback* control. If the current state cannot be observed, *open-loop* control policies may be used which depend only on the current *time* variable. In order to quantify the cost aspects a *cost function* must be specified to assign a *total cost* to the operation of the system, seen as proceeding from any starting point under a given control policy.

At this point we shall have defined the basic dynamic mechanisms of the system, together with various constraints and alternative modes of operation, every feasible realization having an appropriate total cost. The problem facing the investigator is then to determine an *optimal policy* that prescribes the control actions for which the total cost is as small as possible.

In many biologically oriented systems the control policies concern the *rates* at which events occur, like birth, death, infection, or recovery. The control functions may appear *linearly* in the dynamic and cost equations, and in optimal solutions will then take only extreme values. They are said in this case to be of *bang-bang* form, when it is supposed that the relevant rates can be switched from one bound to another virtually instantaneously.

Generally speaking, the discovery of an optimal solution, if indeed one exists, may involve a considerable amount of complicated mathematical work as well as computerized calculation. The following steps usually have to be taken in one form or another:

(1) Show that an optimal policy exists.
(2) Find and apply necessary conditions for optimality, e.g. Pontryagin principle in deterministic formulations, or the principle of dynamic programming in stochastic problems.
(3) Investigate properties of policies satisfying the necessary conditions in (2), and try to restrict the class of policies to be considered.
(4) Try to find an explicit optimal policy from this class of policies, and demonstrate optimality by applying sufficient conditions. The latter may be very difficult in practice, especially in nonlinear problems.
(5) Use computerized numerical methods if necessary: both computer time and storage may turn out to be prohibitive.
(6) When the discovery of an optimal solution is too difficult or costly, it may be possible to find *suboptimal* policies more easily, but their properties must be carefully investigated.

The above verbal description should give some idea of the mathematical complexities that are liable to be encountered in practice in the application of standard control theory. It should not of course be taken as a rigid format, but rather as a guide to the kind of approach frequently adopted. Although there is now a growing literature on control investigations in the infectious disease field (see Wickwire, 1977 for references), very little has

yet been done for parasitic diseases. And the main paper dealing specifically with malaria control is the work of Dietz (1975), which is sketched out in the next section.

10.3 The malaria control model of Dietz

In order to investigate the implications of using a control-theoretic approach to the problem of optimally allocating resources in parasitic disease control in general, but with special reference to malaria, Dietz (1975) started by adapting a simplified model originally introduced by Ross (1911). We have already looked at this briefly at the end of Section 7.3. In fact, with a slight change of notation, we can define the basic transition mechanisms by formula (7.3.14) rewritten as

$$\frac{dy}{dt} = \beta y(1-y) - \gamma y, \tag{10.3.1}$$

where y is the proportion of individuals affected in the human population, i.e. the parasite-rate; β is the effective contact-rate between individuals, mediated via contact with the mosquito population; and γ is the recovery-rate. As in the previous discussion of Section 7.3, we can denote the basic reproduction-rate R by

$$R = \beta/\gamma. \tag{10.3.2}$$

This simplified model was chosen for convenience, but Dietz noted that it had the same qualitative behaviour as more realistic developments.

The effective contact-rate β can easily be expressed as a function of various more fundamental entomological and parasitological parameters. As in the discussion of Section 7.5, we can distinguish the following parameters:

- m: mosquito density per human;
- a: man-biting rate of mosquitoes;
- b: proportion of infected mosquitoes that are actually infectious;
- g: proportion of infected humans that are actually infectious;
- n: length of extrinsic parasite cycle in the mosquito;
- ν': mosquito death-rate;

where we have here included g (following Onori & Grab, 1980a) to represent the proportion of infected humans carrying gametocytes. This is

needed for accuracy, since we have defined y above simply as the overall parasite-rate.

The basic ideas of malaria transmission, as presented in Section 7.5, can be applied directly to the contact-rate β defined above. Thus an infected human is bitten by ma mosquitoes in unit time, of which a proportion g will become infected. Of these infected mosquitoes, a fraction $e^{-n\nu'}$ will survive the extrinsic cycle. The latter will live on average for a further period $(\nu')^{-1}$ distributing a/ν' bites of which a proportion b will be infectious. The contact-rate is then the product of these individual items, i.e.

$$\beta = ma^2gb\, e^{-n\nu'}/\nu'. \tag{10.3.3}$$

Dietz concentrates first on the effects of a control action consisting of drug administrations which are applied to the population at certain intervals of length T. The drugs are assumed to make an affected person immediately unaffected. They may also be given to unaffected people. Individuals are temporarily protected against new infections after receiving drugs, but lose their protection at a rate δ. It is also supposed that only a proportion p of the population is accessible to drug administration, and that the drugs are distributed to a random proportion c of the accessible population.

Let $z(t)$ be the proportion of the population actually protected at time t, with y_1 the proportion affected in the accessible population, and y_2 the proportion affected in the inaccessible population. We must then replace (10.3.1) by

$$\left.\begin{aligned} \frac{dy_1}{dt} &= \beta y(p - z - y_1) - \gamma y_1, \\ \frac{dy_2}{dt} &= \beta y(1 - p - y_2) - \gamma y_2, \\ \frac{dz}{dt} &= -\delta z, \end{aligned}\right\} \tag{10.3.4}$$

where

$$y_1 + y_2 = y. \tag{10.3.5}$$

Let us use the arguments $(iT-)$ and $(iT+)$ to denote the instants of time just before and just after the ith drug application, respectively. Then

at the ith application we have

$$\left.\begin{aligned} y_1(iT+) &= (1-c)\,y_1(iT-), \\ y_2(iT+) &= y_2(iT-), \\ z(iT+) &= z(iT-) + c\{p - z(iT-)\}. \end{aligned}\right\} \tag{10.3.6}$$

In addition the proportion of protected individuals $z(t)$ is independent of y_1 and y_2. So

$$z(t) = z(iT+)\,e^{-\delta t}, \quad iT < t < (i+1)\,T. \tag{10.3.7}$$

Now introduce the new variable

$$x = y_2/y. \tag{10.3.8}$$

Equations (10.3.4) then yield

$$\left.\begin{aligned} \frac{dx}{dt} &= \beta(1-p) - \beta(1-z)\,x, \\ \frac{dy}{dt} &= -\beta y^2 + (\beta - \gamma - \beta z)\,y, \end{aligned}\right\} \tag{10.3.9}$$

where

$$\left.\begin{aligned} x(iT+) &= x(iT-)/[1 - c\{1 - x(iT-)\}], \\ y(iT+) &= y(iT-)\ [1 - c\{1 - x(iT-)\}]. \end{aligned}\right\} \tag{10.3.10}$$

The transformation has thus reduced the problem to solving a non-homogeneous linear differential equation in x and a non-homogeneous Riccati equation in y. Explicit solutions are possible, though involving a certain degree of algebraic and computational complexity. For fuller details Dietz (1975) should be consulted.

In order to evaluate a given control action it is convenient to consider the long-term average of the parasite-rate that would be achieved. As x and y change between drug applications, it is of interest to look for periodic solutions with period T to which x and y will converge. In particular, we shall want to evaluate

$$\bar{Y} = \lim_{i\to\infty}\left\{T^{-1}\int_{iT}^{(i+1)T} y(t)\,dt\right\}. \tag{10.3.11}$$

Generally speaking, we shall be looking for policies to minimize $\bar{Y}$, but

questions of cost may exert a substantial influence on what appears to be best.

The work may simplify considerably in certain special cases. Thus, with a completely accessible population and the drugs affording no protection to the unaffected, we have $p = 1$, $\delta \to \infty$. It turns out that

$$\bar{Y} = \max\,[0, R^{-1}\{R - 1 + D\log(1-c)\}], \tag{10.3.12}$$

where the quantity

$$D \equiv (\gamma T)^{-1} \tag{10.3.13}$$

can be viewed as the number of drug administrations during one average infectious period.

If $\bar{Y}$ is to be zero, then D, c and R have to satisfy

$$\left.\begin{aligned} D &\geqslant (1-R)/\log(1-c), \\ c &\geqslant 1 - \exp\{(1-R)/D\}, \\ R &\leqslant 1 - D\log(1-c). \end{aligned}\right\} \tag{10.3.14}$$

Dietz quotes figures from the simulations of Macdonald, Cuellar & Foll (1968), whose work we have already briefly commented on in Section 8.1. In one study the latter authors took $\gamma = 0{\cdot}005$ and $T = 60$ days, giving $D = 3\frac{1}{3}$, with $c = 0{\cdot}8$. The last inequality in (10.3.14) yields $R \leqslant 6{\cdot}36$. It is interesting to see that Macdonald *et al.* considered that $R = 4$ would lead to eradication, though it appeared from the simulations carried out that this was getting near the limiting value.

An alternative scheme would be to give the same amount of drugs but always to the same sector of the population. In this case we have "fixed coverage" with $c = 1$, rather than the "random coverage" just discussed, and consider $p < 1$. Again letting $\delta \to \infty$, we can obtain an inequality for p, corresponding to the second line of (10.3.14), namely

$$p \geqslant [1 - \exp\{(1-R)/D\}]/(1 - e^{-\beta T}), \tag{10.3.15}$$

which must be satisfied if $\bar{Y}$ is to be zero. This shows that the proportion of the population to be covered is higher for fixed than for random coverage, and the discrepancy increases as the number of contacts, βT, between two drug administrations decreases.

Generally speaking, when the population is completely accessible with $p = 1$ it is always possible to find a drug administration programme for any R that will make $\bar{Y}$ zero. But when $p < 1$ this is no longer true. The question then arises whether a positive endemic level could maintain itself

in the inaccessible part of the population even if the accessible part were kept protected all the time. It can be shown that this will be so if and only if

$$p < 1 - (\gamma/\beta). \tag{10.3.16}$$

Intuitively, we can argue that a proportion p of contacts are "wasted" on the accessible, and protected, part of the population. Hence the reproduction-rate is effectively $\beta(1-p)/\gamma$. If the latter quantity is to be greater than one, the inequality in (10.3.16) follows immediately.

Dietz then goes on to discuss the more general and realistic problem of choosing between four policies where reducing the contact-rate, by spraying for example, is considered in addition to drug administration. We thus have to choose amongst the policies:

$$\left.\begin{aligned} N &\equiv \text{do nothing;} \\ D &\equiv \text{give drugs alone;} \\ S &\equiv \text{reduce contact-rate;} \\ B &\equiv \text{apply both } D \text{ and } S \text{ combined.} \end{aligned}\right\} \tag{10.3.17}$$

Let σ belong to the class $\{N, D, S, B\}$. The problem is then to minimize the total cost of the programme of intervention given by the cost function

$$K_\sigma = \bar{Y}_\sigma + W_\sigma, \tag{10.3.18}$$

where $\bar{Y}_\sigma$ is the average parasite-rate for any policy σ, and W_σ is the yearly per capita cost of implementing the control programme *in units of the cost of the disease for one person-year.*

Two different decision rules can be contemplated:

(1) Adopt the σ^* which minimizes K_σ; if the minimum is attained by more than one strategy, adopt that which also minimizes $\bar{Y}_\sigma$.
(2) Adopt the σ^* which minimizes $\bar{Y}_\sigma$, subject to the constraint $K_\sigma \leqslant K_N = \bar{Y}_N$.

The first of these two rules is concerned primarily with the reduction of costs, irrespective of the parasite-rate. The second rule, on the other hand, seems more in keeping with public health practice, in that the parasite-rate is reduced to the lowest level possible, within the constraint that the total cost after implementing the control programme should be no greater than the cost before this was done.

As an additional simplification, it is assumed that $W_N = 0$ and $W_B = W_D + W_S$. We also have $\bar{Y}_N = Y_0 = 1 - R_0^{-1} = 1 - (\gamma/\beta)$, where Y_0 and R_0 are the initial prevalence-rate and reproduction-rate, respectively.

Dietz (*loc. cit.*) discusses in some detail a variety of numerical results, to which the reader should refer for a full account. General conclusions may, however, be summarized as follows.

A drug administration programme can be specified by p, c, δ and T, while a programme aimed at reducing the contact-rate β can be characterized by the new reproduction-rate R, down to which the initial R_0 is to be lowered. Hence we can use the drug administration theory above to calculate the $\bar{Y}_\sigma$ for any given programme involving specified values of p, c, δ, T, γ, R_0 and R. Next, values must be chosen for the basic costs W_D and W_S. We can then calculate the cost K_σ and investigate the consequences of the two alternative decision rules, (1) and (2), given above.

Dietz pays particular attention to special situations with high and low initial reproduction-rates, e.g. $R_0 = 100$ and $R_0 = 2$, respectively. He points out that for high initial reproduction-rates drug administration alone may be relatively ineffective, but if the rate can first be substantially reduced by spraying, etc., eradication of the parasite may be achieved. The optimal policy in given circumstances depends of course on the actual numerical values of the parameters involved.

To apply the decision rules in practice, a public health decision-maker would have to follow the scheme:

(1) First determine the initial reproduction-rate R_0. As the endemic parasite-rate $1 - R_0^{-1}$ is nearly constant for large R_0, it would be highly desirable in such cases to try and estimate $R_0 = \beta/\gamma$ directly from observations on the contact-rate β and the recovery-rate γ.
(2) Decide which decision rule to adopt.
(3) Determine the costs of implementing the available policies in units of the costs of the disease for one man-year.
(4) Determine for policy D the coverage that can be attained in terms of c and p; and for policy S the reproduction-rate R which is expected to be achieved.
(5) Use the theory outlined above to find the optimal policy.

A recent general extension of Dietz's model has been made by Gonzalez-Guzman (1980). This paper provides an extensive theoretical study of the problem of controlling a parasitic disease using a permanent, time-continuous mixed programme of vector reduction and drug application. It is shown that both the insecticide and drug applications should be as intensive as possible, and that the mixed control programme can have greater long-term effects than the sum of the effects for each programme

taken separately. A procedure is specified for finding the optimal balance between the two control aspects in order to maintain the proportion of affected humans below some given level. The detailed implications for malaria, however, still remain to be worked out.

Many practical problems arise in connexion with the estimation of essential parameters. For a specific locality where malaria control is to be implemented, parameters may have to be estimated from data from that area. If these do not exist in the required form, a pilot project will be required to obtain estimates of sufficient accuracy. Of course, if preliminary estimates are available indicating ranges of possible values, and if it can be shown that policy choice is unaffected by variations within these ranges, then a pilot project would be unnecessary for this purpose. On the other hand, if different policies were optimal according to different assumptions that might be made then a pilot project would have to be implemented in such a way that the results would enable an optimal choice to be found.

In addition, the whole process of decision-making is apt to involve many wider issues, as indicated in Section 10.1, including interactions with several sectors other than health. The further exploration of control-theoretic methods should probably be carried out therefore within this wider context, adopting the approaches of operational research and systems analysis which are discussed more fully in Chapter 12.

11 Sensitivity theory

11.1 Introduction

The importance of sensitivity analysis has already been referred to in Section 5.4, dealing with the principles of validating models. It is undeniable that the value of modelling is closely dependent upon proven usefulness and acceptability. But even when a well-established model is applied to a new situation there may be many uncertainties as to the local values of the relevant parameters. Field data may be initially too incomplete to provide reliable statistical estimates. Some parameters may have been estimated from independent *ad hoc* inquiries, while others are subjectively based on general epidemiological knowledge. And in some cases there may be doubts about the soundness or applicability of the structure of the model as well. Whatever the circumstances, it is a very common occurrence to find that one's assumptions are fraught with a considerable degree of uncertainty. The question immediately arises as to what corresponding level of uncertainty should be attached to the resultant conclusions. Even general results are likely to be of practical value only if the margin of error involved can be held within certain bounds. Accordingly, the subject of sensitivity analysis is primarily concerned with how sensitive the properties or behaviour of a model are to changes that can be envisaged in the supposed parametric values, or even in the assumed model structure itself.

A theoretical basis for general sensitivity theory has been well developed in the context of engineering control, but relatively few applications have been made in the field of epidemiology and public health. We shall therefore introduce some of the elementary notions in Section 11.2 and give a more general account in Section 11.3. Suggestions for a systematic application to malaria modelling are then presented in Section 11.4. An analogous, though slightly different, approach recommended by Nåsell (1980a, b) is discussed in Section 11.5.

11.2 Elementary ideas of sensitivity

First of all it is clear that elementary ideas of sensitivity are rooted in everyday common sense. People constantly weigh up the pros and cons of

different choices of action and make judgements involving intuitive assessments of various probabilities. Without going into detailed logical analysis, people appreciate that assumptions are inevitably liable to varying degrees of error and uncertainty, and that this leads to corresponding uncertainty in conclusions. It is also understood that conclusions are more influenced by possible errors in some assumptions than they are in others. The difficulty of course is to formalize all this so that a complex situation can be analysed in such a way that the relative effects of different factors on the final conclusions can be expressed quantitatively.

Some idea of sensitivity is usually introduced at an early stage in the development of arithmetic skills in relation to the influence of rounding errors. Although only elementary school-level mathematics is involved, the difficulty is a persistent one and is often ignored, to their cost, by higher-level practitioners of numerical computing. Consider, for example, the simple calculation:

$$\frac{5{\cdot}0}{5{\cdot}1 - 4{\cdot}9} = 25{\cdot}0. \tag{11.2.1}$$

An error of half a unit in the decimal place of the numerator, e.g. 5·05 instead of 5·0, means an overall error of no more than *one percent* in the answer. But similar errors in the denominator could lead to the answer being wrong by a multiplying factor of *two*, e.g.

$$\frac{5{\cdot}00}{5{\cdot}05 - 4{\cdot}95} = 50{\cdot}0. \tag{11.2.2}$$

In practice, with a random distribution of rounding errors, the error in the answer would in general not be so extreme, but would still have its own statistical distribution. Although these points are elementary they have far-reaching implications. Rounding-error problems can to some extent be overcome by the retention of a large number of significant figures. But even though electronic computing will handle the increased numerical work very rapidly, intermediate stages of complex calculations often disappear from view and a substantial loss of accuracy may pass unnoticed.

To come to the more immediately relevant field of infectious disease control, it is interesting to note that Daniel Bernoulli (see Bailey, 1975a, p. 360) gave as early as 1760 a simple illustration of the dependence of a model on possible variations in parametric values. He was investigating the public health consequences of variolation (i.e. inoculation, as opposed to the then unknown vaccination) in combating the effects of smallpox.

Bernoulli demonstrated that the calculated gain of three years in life expectancy, if inoculation were completely effective, would be reduced by only two months if the risk of dying from the inoculation itself were at what he regarded as the upper limit (namely, 1 in 200).

An example from recent times is in the work of Feldstein, Piot & Sundaresan (1973) on resource allocation problems in tuberculosis control. It was shown that the consequences for decision-making appeared to be very insensitive to variations in a number of demographic and epidemiological parameters.

The trouble is that when a model has a large number of parameters it is impossible to examine all their simultaneously interacting effects by means of computer simulations. Moreover, the selection of a relatively small number of supposedly important parameters for special investigation could be influenced by unconscious bias. When the latter can be discounted it may be possible to use the following device. Suppose there is a set of k components involved in the mechanism of some piece of equipment, where each component is characterized by a single index and believed on strong *a priori* grounds to have little effect on overall performance. Approximate upper and lower bounds are set for the specification of each component, giving a total of 2^k combinations. The relevant computations, carried out directly or by some form of simulation, are completed for all 2^k combinations. If one is lucky and the performance figures do not differ appreciably amongst the different combinations, then it can be stated that the performance is relatively insensitive to variations in the k components examined. If k is small this may be feasible. But if k is large, and especially if it is necessary to work with a distribution of values for each specification, not just high and low values, then the work required may become astronomical in time and cost.

A simple application of this approach to malaria has already been mentioned in Section 8.2, where Molineaux, Dietz & Thomas (1978) validated their model in circumstances in which the major entomological factor of vectorial capacity was recognized as being subject to appreciable errors of estimation. It turned out that results were relatively unaffected when the simulations were carried out for vectorial capacities set at values that were 10 times larger and 10 times smaller than the preferred set of estimates. In this case, of course, $k = 1$.

Generally speaking, when k is large, one requires a method that will allow one to take into account possible variations in *all* parameters, and to identify by actual investigation, rather than intuitive guesswork, the contribution of each parameter to the overall variation in results. In this

way indications can be given more objectively about the parameters to which the model is most sensitive. This subject will be taken up in more detail in the following section.

Let us conclude the present section by noting three different kinds of situation in which sensitivity considerations are likely to be important. First, there is the case already outlined above, where we have a model in which, given a set of parametric values, certain results logically follow. We then envisage possible variations in our assumptions about the parameters, due perhaps to error or ignorance, and ask about the magnitude of the corresponding variations in the results.

Secondly, as an alternative to this formulation, we may suppose that the parameters are subject, not merely to prior uncertainties, but to actual uncontrollable variations in the real world. Thus, a good deal of work has been done in recent years in mathematical demography to examine the sensitivity of population forecasts to changes in fertility, mortality and migration. Similar considerations obviously apply to a wide range of physiological and epidemiological processes.

Thirdly, we may be interested in the consequences of deliberate changes in a given system, as, for example, when various public health interventions are contemplated like spraying mosquitoes or administering drugs to humans in order to control malaria. Not only do we want to know how to choose an optimal policy from a set of feasible alternatives (see Chapter 10), but we want to know the relative consequences on the overall results of deliberate changes that might be made in different parameters like contact-rates or recovery-rates. In this way we can concentrate practical measures on influencing those parameters to which certain end-results, e.g. predicted parasite-rates, are most sensitive.

11.3 General sensitivity analysis

The previous section has set the general scene for considering sensitivity investigations, using only simple intuitive ideas. In the present section we shall introduce the basic elements of a more systematic and mathematical account. There is in fact an extensive literature relevant to engineering applications, and readers may wish to refer to the standard theoretical text by Tomović & Vukobratović (1972) or the book edited by Cruz (1973), while for a general discussion on the related subject of system identification, Mehra & Lainiotis (1976) should be consulted. Another useful reference is Dickinson & Gelinas (1976), which emphasizes the systematic determination of the sensitivity of solutions of ordinary

nonlinear differential equations to uncertainties in both rate-parameters and initial conditions.

At the very simplest level we are interested in a single function $x(\theta)$ depending on the single parameter θ, and wish to relate changes in x to changes in θ. If we confine ourselves to small variations, we can write in the usual way

$$\delta x = \frac{dx}{d\theta}\,\delta\theta, \tag{11.3.1}$$

where $dx/d\theta$ is the *sensitivity function.*

This immediately generalizes of course to the situation where there are k functions, given by the column vector $\mathbf{x} \equiv \{x_i\}$, $i = 1, \ldots, k$; and p essential parameters indicated by $\boldsymbol{\theta} \equiv \{\theta_j\}, j = 1, \ldots, p$. We can then extend (11.3.1) to

$$\delta\mathbf{x} = \mathbf{H}\,\delta\boldsymbol{\theta}, \tag{11.3.2}$$

where

$$\mathbf{H} \equiv \{h_{ij}\} \equiv \left\{\frac{\partial x_i}{\partial \theta_j}\right\} \equiv \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}}. \tag{11.3.3}$$

The matrix $\mathbf{H}$ is in fact what is usually called the *sensitivity matrix*, whose individual elements h_{ij} are the sensitivity functions $\partial x_i/\partial\theta_j$. It thus comes about quite naturally that the calculation of these partial differential coefficients is fundamental to sensitivity analysis. The situation may, however, become more complicated if various kinds of discontinuity occur in the processes under investigation.

Let us suppose for the moment that such difficulties are not important. The question immediately arises as to the implications and interpretation of equation (11.3.2). In sophisticated contexts, such as control engineering, there are plenty of opportunities for developing and applying the appropriate mathematical instruments. But in epidemiology and public health it is important to couch conclusions in terms that are readily assimilated to decision-making practice. Since elementary statistical ideas of natural variation, frequency distributions, standard errors, coefficients of variation, etc. are fairly widely understood, it is very convenient to consider putting results into this form if it is appropriate. Some situations will, however, need a more searching form of analysis.

For example, in the context of physics and chemical kinetics, Dickinson & Gelinas (1976, footnote to p. 129 and in Section V) consider that a

statistical approach may often be unsuitable since the modeller is more concerned with the extreme values that could occur in the worst cases, and the latter might be obscured by the use of means and variances. The authors also suggest that "by the time the statistical properties of rate data become available some of the most pressing needs for sensitivity analysis may have diminished". But this suggestion has less force in biology and epidemiology where uncertainty and natural variation are far more pervasive.

Suppose, therefore, that the vector $\boldsymbol{\theta}$ is subject to a multidimensional prior distribution, for which the covariance matrix is $\mathbf{V}(\boldsymbol{\theta})$. It then easily follows that the corresponding covariance matrix $\mathbf{V}(\mathbf{x})$ for the multidimensional distribution of the vector x is given by

$$\begin{aligned}\mathbf{V}(\mathbf{x}) &= E(\delta\mathbf{x})(\delta\mathbf{x})' \\ &= E(\mathbf{H}\,\delta\boldsymbol{\theta})(\mathbf{H}\,\delta\boldsymbol{\theta})' \\ &= E(\mathbf{H}\,\delta\boldsymbol{\theta}\;\delta\boldsymbol{\theta}'\mathbf{H}') \\ &= \mathbf{H}E(\delta\boldsymbol{\theta}\;\delta\boldsymbol{\theta}')\,\mathbf{H}' \\ &= \mathbf{H}\mathbf{V}(\boldsymbol{\theta})\,\mathbf{H}', \end{aligned} \tag{11.3.4}$$

where the primes indicate transposition.

Generally speaking, the elements of $\mathbf{V}(\boldsymbol{\theta})$ may have been derived from various sources, e.g. (i) joint maximum-likelihood estimation using appropriate data, (ii) independent *ad hoc* estimation of the different parameters, (iii) general expert knowledge summarizing existing epidemiological opinion, or (iv) more intuitive *prior* estimates of subjective uncertainty.

In many cases it will not be unreasonable to suppose that all the θ_j are uncorrelated, especially when the parameters are subjective prior estimates. When this happens the covariance matrix $\mathbf{V}(\boldsymbol{\theta})$ is of diagonal form, namely

$$\mathbf{V}(\boldsymbol{\theta}) = \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_p^2 \end{bmatrix}, \tag{11.3.5}$$

where

$$\sigma_j^2 = \text{var}\,(\theta_j), \tag{11.3.6}$$

and all nondiagonal elements are zero.

Substituting (11.3.5) into (11.3.4) then gives

$$\text{var}\,(x_i) = \sum_{j=1}^{p} h_{ij}^2 \sigma_j^2. \tag{11.3.7}$$

This result enables us quickly to identify the parameters to which any given x_i is most sensitive. Improved information on those parameters would lead to a more reliable estimate of x_i. If there are several x_i of special interest the situation may be a little more complicated, but we may well be able to avoid the collection of a lot of data that would make little contribution to improved accuracy.

Most studies on dynamical systems involve sets of differential equations whose solutions usually supply the main source of information about system behaviour. This is certainly true of the population dynamics of infectious disease in general, and malaria in particular. Let us suppose, for convenience, and by a slight extension of the previous notation, that our basic model consists of k compartments whose state variables are now functions of time indicated by $\mathbf{x}(t) \equiv \{x_i(t)\}$, $i = 1, \ldots, k$; with p parameters given as before by $\boldsymbol{\theta} \equiv \{\theta_j\}, j = 1, \ldots, p$.

The rate of change in the ith compartment can usually be written as

$$\frac{\partial x_i}{\partial t} = f_i(\mathbf{x}, \boldsymbol{\theta}, t), \quad i = 1, \ldots, k, \tag{11.3.8}$$

where the functions f_i reflect the model structure. This can be put more compactly in the matrix form

$$\frac{\partial \mathbf{x}}{\partial t} = \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}, t), \tag{11.3.9}$$

where $\mathbf{f} \equiv \{f_i\}$, $i = 1, \ldots, k$. We can if we wish make the parameters time-dependent, i.e. $\boldsymbol{\theta} \equiv \boldsymbol{\theta}(t)$. It may also be convenient to include the initial conditions along with the parameters.

We have already defined the sensitivity matrix $\mathbf{H}$ in (11.3.3) as $\partial\mathbf{x}/\partial\boldsymbol{\theta}$. Two additional matrices are now required, namely

$$\mathbf{F} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \equiv \left\{\frac{\partial f_i}{\partial x_j}\right\}, \qquad \mathbf{G} = \frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}} = \left\{\frac{\partial f_i}{\partial \theta_j}\right\}. \tag{11.3.10}$$

Equation (11.3.9) can be differentiated with respect to $\boldsymbol{\theta}$ to give

$$\frac{\partial}{\partial \boldsymbol{\theta}}\left(\frac{\partial \mathbf{x}}{\partial t}\right) = \frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}} + \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\,\frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}}. \tag{11.3.11}$$

Reversing the order of differentiation on the left, and using the definitions of **F**, **G** and **H** in (11.3.10) and (11.3.3), then yields

$$\frac{\partial \mathbf{H}}{\partial t} = \mathbf{G} + \mathbf{FH}. \tag{11.3.12}$$

If we now combine (11.3.9) and (11.3.12), we have an extended set of differential equations that can be solved simultaneously to yield both $\mathbf{x}(t)$ and $\mathbf{H}(t)$. Since (11.3.9) involves k equations and (11.3.12) involves kp equations, the total number of simultaneous equations is $k(p+1)$. Explicit numerical solutions can normally be obtained by using appropriate computer methods, but careful attention to the algorithmic detail is required (see Dickinson & Gelinas, 1976).

In the special case of steady-state solutions the work may simplify considerably. Thus (11.3.9) will take the form

$$\mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) = 0, \tag{11.3.13}$$

while (11.3.12) yields

$$\mathbf{H} = -\mathbf{F}^{-1}\mathbf{G}. \tag{11.3.14}$$

If (11.3.13) is linear in the x_i the work of solving both (11.3.13) and (11.3.14) reduces to straightforward matrix manipulations. Most infectious disease models are liable to contain nonlinear elements, and special methods may be required unless the models are fairly simple. Thus, in applying sensitivity analysis to a typhoid fever model, Bailey & Duppenthaler (1979, 1980) used a device due to Békéssy (1971) that allowed the nonlinear simultaneous equations corresponding to (11.3.13) to be reduced to a linear set of equations using nothing more complicated than the numerical evaluation of certain determinants. By taking advantage of a fair amount of mathematical derivation, it turned out to be possible to deal with the numerical work arising from the 10-compartment typhoid model, involving a total of some 25 parameters, using only an independent Hewlett-Packard HP 9830A mini-computer.

Substantive applications of comprehensive sensitivity analysis have not yet been undertaken for malaria, nor indeed for any other parasitic disease. However, a number of elementary illustrations are given in the following section, together with suggestions for more systematic investigations.

11.4 Elementary applications to malaria

We have already seen in Section 6.6 how certain general endemic models, that might have some broad relevance to malaria, do in fact have explicitly soluble equilibrium solutions in spite of the nonlinearities involved. In such cases the relevant sensitivity functions can be calculated directly as described at the beginning of the previous section, using the formulae in (11.3.1)-(11.3.3). And if suitable assumptions can be made about a multi-dimensional prior distribution of parameters we can use the results in (11.3.4)-(11.3.7). Of course, none of this would be worth doing in practice unless we had sufficient confidence in a model to use it for interpreting actual data, or perhaps making tentative forecasts in relation to possible public health decision-making.

We might, however, use one of the Ross models introduced in Section 7.3, or the version based on the "Theory of Happenings" described in Section 7.4. In the latter case, we can write the endemic level x as

$$x = \frac{h}{r+h}, \tag{11.4.1}$$

whence

$$\begin{aligned} \delta x &= \frac{\partial x}{\partial h}\,\delta h + \frac{\partial x}{\partial r}\,\delta r \\ &= \frac{r}{(r+h)^2}\,\delta h - \frac{h}{(r+h)^2}\,\delta r. \end{aligned} \tag{11.4.2}$$

Suppose we now assume the existence of the variances var (h), var (r) and var (x), and write the corresponding coefficients of variation as C_h, C_r and C_x. Let us also assume for convenience that cov $(h, r) \equiv 0$. Squaring both sides of (11.4.2) and taking expectations then quickly leads to

$$C_x^2 = (1-x)^2(C_h^2 + C_r^2), \tag{11.4.3}$$

where of course $C_h^2 = \text{var}\,(h)/h^2$, etc.

A more elaborate analysis of a similar kind could also be made for finite time, using the explicit time-dependent solution in (7.4.2). Again, we could examine the sensitivity of the age-dependent prevalence curve in (7.4.3) to uncertainty in the parameters h and r.

A further possibility would be to examine the more realistic model of Ross given in (7.3.3), where variations in both human and mosquito

populations are allowed for. This also has an explicit steady-state solution, as given in (7.3.6). The slightly modified version in (7.3.7), relating to the human malaria-rate and the density of infected mosquitoes per human, would be more convenient in practice. There are now six basic parameters, namely a, b', f, f', γ and μ', whose values could all be estimated or provisionally guessed, leading to an appropriate sensitivity investigation. It should be noted that the simplification proposed by Nåsell (1980a), using the two transmission factors T_1 and T_2 defined in (7.3.9), could be useful when data are available to estimate T_1 and T_2 directly, thus simplifying the analysis of the behaviour of a given endemic situation. But in a new situation we should probably find that discussion prior to the collection of field data involved the use of educated guesses at all six parameters, in which case the preliminary conclusions reached should be tested for their sensitivity to the explicit assumptions made about these parameters.

Again, we could point to the use made by Dietz (1975) of a simplified model of Ross, virtually equivalent to the approximate form in (7.3.14), for the purpose of expounding a control-theoretic approach – as discussed in Section 10.3 above. It could easily turn out that the choice of an over-simplified model on the one hand, and the existence of considerable uncertainty as to many of the intrinsic parameters in practice, on the other hand, could have extensive implications for the resultant choice of intervention strategies. In any real-life application careful attention would have to be paid to the sensitivity aspects, possibly in relation to a more sophisticated model as well.

Finally, it would be of great practical value to investigate the full sensitivity of the model of Dietz, Molineaux & Thomas (1974), reviewed in Section 8.2. Since this model has an improved degree of epidemiological reality, and has passed statistical tests of fit for three parameters which clearly had to be estimated from local data, it would be a priority candidate for detailed sensitivity analysis. It must not be forgotten, however, that the attempt to validate the model by application to a new body of data by Molineaux, Dietz & Thomas (1978), was extremely encouraging as described at the end of Section 8.2. In addition, the authors were aware of the importance of sensitivity considerations which they applied specifically to the highly uncertain vectorial capacity $C(t)$, using repeated simulations with extreme values of $10\,C(t)$ and $0{\cdot}1\,C(t)$.

It must be confessed that we have done no more in this section than point to a number of possibilities for applying sensitivity analysis to certain available models. Obviously, there is a big field for further research here. It would be easy simply to undertake the purely mathematical

derivations indicated. But real conviction of the utility of the exercise is likely to be forthcoming only if this can be done in the context of specific epidemiological situations whose analysis, both initially in the absence of data and subsequently using data from field surveys, is geared to local decision-oriented problems of strategy choice.

11.5 Nåsell's control efficiency functions

In this development of a hybrid version of an early Ross model for malaria, Nåsell (1980a) employs the notions of *control efficiencies* and *control efficiency functions*, previously introduced in the context of schistosomiasis modelling (Nåsell & Hirsch, 1973; Nåsell, 1977). The hybrid model itself has already been described above in Section 8.4, where the basic mathematical properties were discussed. The use of control theory in determining an optimal choice of control policies was outlined in Chapter 10. While we might have discussed Nåsell's ideas on control at that point, it is preferable to include them here since they are more closely related to sensitivity analysis, especially the third type of application mentioned at the end of Section 11.2.

Nåsell defines the efficiency of control of any quantity of epidemiological interest, x say, through modification of a given parameter θ, in terms of the control efficiency function given by

$$\left.\begin{aligned} C_{x\theta} &= \left|\theta\,\frac{\partial x}{\partial \theta}\right|, \quad \text{if } x \text{ is dimensionless,} \\ &= \left|\frac{\theta}{x}\,\frac{\partial x}{\partial \theta}\right|, \quad \text{otherwise.} \end{aligned}\right\} \tag{11.5.1}$$

Control efficiency functions are thus essentially positive. If we consider variations in x consequent on variations in a single parameter (though there may in fact be several parameters), we can write

$$\delta x = \frac{\partial x}{\partial \theta}\,\delta\theta, \tag{11.5.2}$$

slightly extending formula (11.3.1).

If we now envisage, as in Section 11.4, a statistical distribution for variations in θ and therefore in x as well, we can square both sides of

(11.5.2) and take expectations to give

$$\text{var}\,(x) = \left(\frac{\partial x}{\partial \theta}\right)^2 \text{var}\,(\theta). \tag{11.5.3}$$

Let us now write the coefficients of variation of x and θ as C_x and C_θ. Thus (11.5.3) easily yields

$$\begin{aligned} C_x^2 &= \frac{\text{var}\,(x)}{x^2} \\ &= \frac{\theta^2}{x^2}\left(\frac{\partial x}{\partial \theta}\right)^2 \frac{\text{var}\,(\theta)}{\theta^2} \\ &= \frac{\theta^2}{x^2}\left(\frac{\partial x}{\partial \theta}\right)^2 C_\theta^2, \end{aligned} \tag{11.5.4}$$

or

$$C_x = \left|\frac{\theta}{x}\,\frac{\partial x}{\partial \theta}\right| C_\theta. \tag{11.5.5}$$

Thus Nåsell's control efficiency function $C_{x\theta}$, as defined in the second line of (11.5.1), would simply be

$$C_{x\theta} = C_x / C_\theta. \tag{11.5.6}$$

This ratio of coefficients of variation has been explicitly used by Bailey & Duppenthaler (1979, 1980) in their discussion of sensitivity analysis applied to a typhoid fever model, already mentioned in Section 11.3.

If one adopts the alternative definition of $C_{x\theta}$ in the first line of (11.5.1), as proposed by Nåsell for dimensionless x, then the interpretations work out slightly differently.

In his discussion of the steady-state endemic situation, Nåsell (1980a) identifies five epidemiological quantities as indicators of the health condition of the community, relative to the disease in question. These are (a) infection probability $P_1(T_1, T_2)$, (b) public health factor $H_1(z_1, r_1, T_2)$, (c) prevalence $Q(s; z_1, r_1, T_2)$ at age s, (d) incidence $I(s, T; z_1, r_1, T_2)$ at age s over time-interval T, and (e) recovery probability $R(s, T; z_1, r_1, T_2)$ at age s over time-interval T, where r_1, T_1 and T_2 have already been defined in Section 8.4, and z_1 is given by

$$z_1 = f_2 b_2 N_2 / N_1. \tag{11.5.7}$$

In particular, we have

$$T_1 = z_1 / r_1. \tag{11.5.8}$$

For many purposes, the behaviour of the system can be described in terms of two parameters only, namely T_1 and T_2. But for the public health factor and the age-dependent epidemiological quantities mentioned above, three basic parameters are required, e.g. z_1, r_1 and T_2.

Nåsell gives a fairly detailed discussion of the 14 control efficiency functions related to z_1, r_1 and T_2: two for item (a) above, and three for each of items (b)–(e). He also examines a corresponding series of functions related directly to basic parameters such as r_1, r_2, b_2, etc. A large variety of inequalities is established for many of the control efficiency functions. For example, using the definition in the first line of (11.5.1), it can be shown that

$$\frac{C_{P_1 T_1}}{C_{P_1 T_2}} = \frac{1}{1 - P_2}.$$

Hence

$$C_{P_1 T_1} > C_{P_1 T_2}, \quad \text{for } P_2 > 0. \tag{11.5.9}$$

Nåsell comments that "in any community above threshold a given proportionate reduction of T_1 leads to a larger reduction of P_1 than an equally large proportionate reduction of T_2. The advantage of T_1 over T_2 increases as P_2 becomes larger. Near the threshold P_2 is small and the advantage is small." Similar arguments also arise when working with basic parameters. For full details the reader should consult the paper by Nåsell (1980a).

It is, however, not clear how far such comparisons can be taken. In practice we may be more involved in trying to decide the optimal mix of interventions to adopt, given the cost of individual interventions, than in comparing the relative effects of given proportionate changes in different parameters.

Nevertheless, the use of some modified form of sensitivity approach to the problem of choosing an optimal intervention strategy is clearly worth investigating further. The mathematical work involved is less recondite than that of the classical control theory of Chapter 10, and the principles of the approach should accordingly be more readily accessible to those whose principal experience lies in epidemiology and public health.

References

An attempt has been made to include all relevant references that are primarily mathematical or statistical in character. Most, but not all, are explicitly mentioned in the text. There are inevitably a certain number of borderline references about whose inclusion or exclusion arbitrary decisions have had to be taken. Several works of a more clinical, epidemiological or public health character have also been added if they exemplify specific points in the text, or if they provide comparatively non-technical introductions to the basic concepts involved in understanding the broad background of infectious disease dynamics and the spread of malaria in particular. These latter references, which are particularly suited for general reading, are marked with an asterisk.

AALEN, O. O. (1975). Statistical inference for a family of counting processes. (Ph.D. Thesis: Univ. of California, Berkeley.)

AALEN, O. O. (1977). Weak convergence of stochastic integrals related to counting processes. *Z. Wahr. Verw. Geb.*, **38**, 261–77.

AALEN, O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.*, **6**, 701–26.

ADLER, G. (1976). A deterministic model of vector-borne epidemics using partial differential equations. *Math. Biosci.*, **28**, 301–20.

ANDERSON, R. M. (1976). Some simple models of the population dynamics of eucaryotic parasites. *Lecture Notes in Biomathematics*, **11**, 16–57. New York: Springer.

ANDERSON, R. M. & MAY, R. M. (1979). Population biology of infectious diseases: Part I. *Nature*, **280**, 361–7.

ARMITAGE, P. (1953). A note on the epidemiology of malaria. *Trop. Dis. Bull.*, **50**, 890–2.

BAILEY, N. T. J. (1951). Testing the solubility of maximum likelihood equations in the routine application of scoring methods. *Biometrics*, **7**, 268–74.

BAILEY, N. T. J. (1957). *The Mathematical Theory of Epidemics* (1st edn). London: Griffin.

BAILEY, N. T. J. (1964). *The Elements of Stochastic Processes with Applications to the Natural Sciences.* New York: Wiley.

BAILEY, N. T. J. (1975a). *The Mathematical Theory of Infectious Diseases and its Applications.* London: Griffin.

BAILEY, N. T. J. (1975b). Current trends in the modelling of infectious disease. *Proc. 8th International Biometric Conf.* (Constanţa, 25–30 August, 1974), 113–21. Bucharest.

BAILEY, N. T. J. (1975c). Approaches to the control of infectious diseases. *Perspectives in Probability and Statistics* (ed. J. Gani), 357–72. (Papers in honour of M. S. Bartlett on his 65th birthday.) London: Academic Press.

BAILEY, N. T. J. (1975d). Communicable disease and public health control. *Bull. I.S.I.*, **46**(1), 265–78.

*BAILEY, N. T. J. (1977a). *Mathematics, Statistics and Systems for Health.* Chichester, U.K.: Wiley.

BAILEY, N. T. J. (1977b). The role of validation in health care systems modelling. (IIASA Conference on "Modeling Health Care Systems", Laxenburg, Austria, 22–24 November 1977.)

BAILEY, N. T. J. (1978). The utilization and validation of mathematical models in medicine and public health. *Lecture Notes in Medical Informatics*, **1**, 393–402. Heidelberg: Springer.

BAILEY, N. T. J. (1979). Introduction to the modelling of venereal disease. *J. Math. Biol.* **8**, 301–22.

BAILEY, N. T. J. (1980a). Spatial models in the epidemiology of infectious diseases. *Lecture Notes in Biomathematics*, **38**, 233–61. New York: Springer.

BAILEY, N. T. J. (1980b). Modelling and informatics in health care management. *Proc. 3rd World Conf. Med. Inform.*, 404–8. Amsterdam, New York, Oxford: North-Holland.

BAILEY, N. T. J. & DUPPENTHALER, J. L. (1979). Sensitivity analysis in infectious disease control. *Lecture Notes in Medical Informatics*, **5**, 1–8. Heidelberg: Springer.

BAILEY, N. T. J. & DUPPENTHALER, J. L. (1980). Sensitivity analysis in the modelling of infectious disease dynamics. *J. Math. Biol.*, **10**, 113–31.

BAILEY, N. T. J. & THOMPSON, M. (eds) (1975). *Systems Aspects of Health Planning.* Amsterdam: North-Holland.

BARTLETT, M. S. (1955). *An Introduction to Stochastic Processes* (1st edn). Cambridge Univ. Press.

BARTLETT, M. S. (1956). Deterministic and stochastic models for recurrent epidemics. *Proc. 3rd Berkeley Symp. Math. Statist. & Prob.*, **4**, 81–109. Berkeley and Los Angeles: Univ. of California.

BARTLETT, M. S. (1957). Measles periodicity and community size. *J. R. Statist. Soc.*, Ser. A, **120**, 48–70.

BARTLETT, M. S. (1964). The relevance of stochastic models for large-scale epidemiological phenomena. *Appl. Statist.*, **13**, 2–8.

BARTLETT, M. S. (1966). Some notes on epidemiological theory. *Research Papers in Statistics: Festschrift for J. Neyman* (ed. F. N. David), 25–36. New York: Wiley.

BECKER, N. (1976). Estimation for an epidemic model. *Biometrics*, **32**, 769–77.

BECKER, N. (1977a). Estimation for discrete time branching processes with application to epidemics. *Biometrics*, **33**, 515–22.

BECKER, N. (1977b). On a general stochastic epidemic model. *Theor. Pop. Biol.*, **11**, 23–36. [Correction: *ibid.*, **14**, 498.]

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

BECKER, N. (1979b). An estimation procedure for household disease data. *Biometrika*, **66**, 271–7.

BECKER, N. (1980). (Unpublished Temporary Advisor's Report to the Health Statistical Methodology Unit, World Health Organization.)

BECKER, N. (1981). The infectiousness of a disease within households. *Biometrika*, **68**, 133–41.

BÉKÉSSY, A. (1971). Remarks relevant to the paper: Epidemiological model of typhoid fever, etc. (Internal WHO memo. 24/9/71.)

BÉKÉSSY, A., MOLINEAUX, L. & STOREY, J. (1976). Estimation of incidence and recovery rates of *Plasmodium falciparum* parasitaemia from longitudinal data. *Bull. Wld Hlth Org.*, **54**, 685–93.

BRADLEY, D. J. (1979). Prevention of disease in the tropics: Questions in health economics. *Epidemiol. Comm. Hlth*, **33**, 66–73. (Including *Comment* by N. M. Prescott.)

BRAUER, F. & NOHEL, J. A. (1969). *The Qualitative Theory of Ordinary Differential Equations*. New York: Benjamin.

BRUCE-CHWATT, L. J. (1976). Mathematical models in the epidemiology of malaria. *Trop. Geogr. Med.*, **28**, 1–8.

BRUCE-CHWATT, L. J. (1979). Man against malaria: conquest or defeat. *Trans. Roy. Soc. Trop. Med. Hyg.*, **73**, 605–17.

*BRUCE-CHWATT, L. J. (1980). *Essential Malariology*. London: Heinemann.

BUCK, A. A., ANDERSON, R. I., MacRAE, A. A. & FAIN, A. (1978a). Epidemiology of polyparasitism. I. Occurrence, frequency and distribution of multiple infections in rural communities in Chad, Peru, Afghanistan and Zaire. *Tropenmed. Parasit.*, **29**, 61–70.

BUCK, A. A., ANDERSON, R. I., MacRAE, A. A. & FAIN, A. (1978b). Epidemiology of polyparasitism. II. Types of combinations, relative frequency and associations of multiple infections. *Tropenmed. Parasit.*, **29**, 137–44.

BUCK, A. A., ANDERSON, R. I. & MacRAE, A. A. (1978c). Epidemiology of polyparasitism. III. Effects on the diagnostic capacity of immunological tests. *Tropenmed. Parasit.*, **29**, 145–55.

BUCK, A. A., ANDERSON, R. I. & MacRAE, A. A. (1978b). Epidemiology of polyparasitism. IV. Combined effects on the state of health. *Tropenmed. Parasit.*, **29**, 253–68.

*BURNET, F. M. & WHITE, D. O. (1972). *The Natural History of Infectious Disease* (4th edn). Cambridge Univ. Press.

CHURCHMAN, C. W., ACKOFF, R. L. & ARNOFF, E. L. (1957). *Introduction to Operations Research*. New York: Wiley.

COHEN, J. E. & SINGER, B. (1979). Malaria in Nigeria: Constrained continuous-time Markov models for discrete-time longitudinal data on human mixed-species infections. *Lectures on Mathematics in the Life Sciences*, **12**, 69–133.

COOKE, K. L. (1979). Stability analysis for a vector disease model. *Rocky Mountain J. Math.*, **9**, 31–42.

COX, D. R. & MILLER, H. D. (1965). *The Theory of Stochastic Processes*. New York: Wiley.

CRUZ, Jr., J. B. (ed.) (1973). *System Sensitivity Analysis*. Stroudsberg, Penn.: Dowden, Hutchinson & Ross.

CVJETANOVIĆ, B., GRAB. & UEMURA, K. (1978). Dynamics of acute bacterial diseases. *Bull. Wld Hlth Org.*, **56**, Supp., 11–143.

DICKINSON, R. P. & GELINAS, R. S. (1976). Sensitivity analysis of ordinary differential equations – a direct method. *J. Comp. Phys.*, **21**, 123–43.

DIETZ, K. (1970). Mathematical models for malaria in different ecological zones. (Presented to the 7th International Biometric Conference, Hannover, 16–21 August 1970.)

DIETZ, K. (1971). Malaria models. *Adv. Appl. Prob.*, **3**, 208–10.

DIETZ, K. (1975). Models for parasitic disease control. *Bull. I.S.I.*, **46** (1), 531–44.

DIETZ, K. (1976). The incidence of infectious disease under the influence of seasonal fluctuations. *Lecture Notes in Biomathematics*, **11**, 1–15. New York: Springer.

DIETZ, K., MOLINEAUX, L. & THOMAS, A. (1974). A malaria model tested in the African Savannah. *Bull. Wld Hlth Org.*, **50**, 347–57.

DUNN, F. L. (1979). Behavioural aspects of the control of parasitic diseases. *Bull. Wld Hlth Org.*, **57**, 499–512.

DUTERTRE, J. (1976). Etude d'un modèle épidémiologique appliqué au paludisme. *Ann. Soc. Belge Méd. Trop.*, **56**, 127–41.

ELDERKIN, R. H., BERKOWITZ, D. P., FARRIS, F. A., GUNN, C. F., HICKERNELL, F. J., KASS, S. N., MANSFIELD, F. I. & TARANTO, R. G. (1977). On the steady state of an age dependent model for malaria. *Nonlinear Systems and Applications* (ed. V. Lakshmikantham), 491–512. New York & London: Academic Press.

FARID, M. A. (1979). Malaria: reflections of a veteran malariologist. *World Health Forum*, **1**, 9–22.

FELDSTEIN, M. S., PIOT, M. A. & SUNDARESAN, T. K. (1973). Resource allocation model for public health planning. *Bull. Wld Hlth Org.*, **48**, Supp., 3–108.

FINE, P. E. M. (1975a). Ross's *a priori* pathometry – a perspective. *Proc. Roy. Soc. Med.*, **68**, 547–51.

FINE, P. E. M. (1975b). Superinfection – a problem in formulating a problem. *Trop. Dis. Bull.*, **72**, 475–88.

FORRESTER, J. W. (1961). *Industrial Dynamics*. Cambridge, Mass.: MIT.

FORRESTER, J. W. (1969). *Urban Dynamics*. Cambridge, Mass.: MIT.

FORRESTER, J. W. (1971). *World Dynamics*. Cambridge, Mass.: Wright-Allen.

GARRETT-JONES, C. (1964a). The human blood index of malaria vectors in relation to epidemiological assessment. *Bull. Wld Hlth Org.*, **30**, 241–61.

GARRETT-JONES, C. (1964b). Prognosis for interruption of malaria transmission through assessment of the mosquito's vectorial capacity. *Nature*, **204**, 1173–5.

GARRETT-JONES, C. & GRAB, B. (1964). The assessment of insecticidal impact on the malaria mosquito's vectorial capacity, from data on the proportion of parous females. *Bull. Wld Hlth Org.*, **31**, 71–86.

GARRETT-JONES, C. & SHIDRAWI, G. R. (1969). Malaria vectorial capacity of a population of *Anopheles gambiae* – an exercise in epidemiological entomology. *Bull. Wld Hlth Org.*, **40**, 531–45.

GEDDES SMITH, *see* SMITH, GEDDES.

GOFFMAN, W. & NEWILL, V. A. (1964). Generalization of epidemic theory. An application to the transmission of ideas. *Nature*, **204**, 225–8.

GONZALEZ-GUZMAN, J. (1980). A mixed program for parasitic disease control. *J. Math. Biol.*, **10**, 53–64.

*GREENWOOD, M. (1935). *Epidemics and Crowd Diseases*. London: Williams & Norgate.

GRIFFITHS, D. A. (1971). Epidemics models. (Ph.D. Thesis, University of Oxford.)

GRIFFITHS, D. A. (1972). A bivariate birth-death process which approximates to the spread of a disease involving a vector. *J. Appl. Prob.*, **9**, 65–75.

GRIFFITHS, D. A. (1973). Multivariate birth-and-death processes as approximations to epidemic processes. *J. Appl. Prob.*, **10**, 15–26.

*HARE, R. (1954). *Pomp and Pestilence*. London: Gollancz.

*HARRISON, G. (1978). *Mosquitoes, Malaria & Man: A History of the Hostilities since 1880*. New York: Dutton.

HETHCOTE, H. W. (1974). Asymptotic behavior and stability in epidemic models. *Lecture Notes in Biomathematics*, **2**, 83–92. New York: Springer.

HETHCOTE, H. W. (1975). Mathematical models for the spread of infectious diseases. *Proc. SIMS Conf. Epidemiol.* (Alta, Utah, 8–12 July, 1974), 122–31. Philadelphia: SIAM.

HETHCOTE, H. W. (1976). Qualitative analysis of communicable disease models. *Math. Biosci.*, **28**, 335–56.

HUBBERT, W. T., McCULLOCH, W. F. & SCHNURRENBERGER, P. R. (1975). *Diseases Transmitted from Animals to Man* (6th edn). Springfield, Illinois: Thomas.

KERMACK, W. O. & McKENDRICK, A. G. (1927). Contributions to the mathematical theory of epidemics (Part I). *Proc. Roy. Soc.*, A, **115**, 700–21.

KOSTITZIN, V. A. (1934). Symbiose, parasitisme et évolution. *Actualités Scient. Indust.*, **96**, 5–45.

*LAPAGE, G. (1963). *Animal Parasites in Man* (2nd edn). New York: Dover.
LOTKA, A. J. (1923). Contributions to the analysis of malaria epidemiology. *Amer. J. Hyg.*, 3 (Suppl. 1), 1–121.
MACDONALD, G. (1950a). The analysis of infection rates in diseases in which superinfection occurs. *Trop. Dis. Bull.*, 47, 907–15.
MACDONALD, G. (1950b). The analysis of malaria parasite rates in infants. *Trop. Dis. Bull.*, 47, 915–38.
MACDONALD, G. (1952a). The analysis of the sporozoite rate. *Trop. Dis. Bull.*, 49, 569–86.
MACDONALD, G. (1952b). The analysis of equilibrium in malaria. *Trop. Dis. Bull.*, 49, 813–29.
MACDONALD, G. (1953). The analysis of malaria epidemics. *Trop. Dis. Bull.*, 50, 871–89.
MACDONALD, G. (1955). The measurement of malaria transmission. *Proc. Roy. Soc. Med.*, 48, 295–301.
MACDONALD, G. (1957). *The Epidemiology and Control of Malaria.* London: Oxford Univ. Press.
MACDONALD, G. (1965a). On the scientific basis of tropical hygiene. *Trans. Roy. Soc. Trop. Med. Hyg.*, 59, 611–20.
MACDONALD, G. (1965b). Eradication of malaria. *Pub. Hlth Rep.*, 80, 870–80.
MACDONALD, G. (1973). *Dynamics of Tropical Diseases* (Collected papers, eds L. J. Bruce-Chwatt and V. J. Glanville). London: Oxford Univ. Press.
MACDONALD, G., CUELLAR, C. B. & FOLL, C. V. (1968). The dynamics of malaria. *Bull. Wld Hlth Org.*, 38, 743–55.
MACDONALD, G. & GÖCKEL, C. W. (1964). The malaria parasite rate and interruption of transmission. *Bull. Wld Hlth Org.*, 31, 365–77.
*McNEILL, W. H. (1977). *Plagues and Peoples.* Oxford: Blackwell.
MAJONE, G. (1975). The feasibility of social policies. *Policy Science*, 6, 49–69.
MARCATI, P. & POZIO, M. A. (1980). Global asymptotic stability for a vector disease model with spatial spread. *J. Math. Biol.*, 9, 179–87.
MARTINI, E. (1921). *Berechnungen und Beobachtungen zur Epidemiologie und Bekämpfung der Malaria.* Hamburg: Gente.
MAY, R. M. (1977). Threshold and breakpoints in ecosystems with a multiplicity of stable states. *Nature*, 269, 471–7.
MAY, R. M. & ANDERSON, R. M. (1979). Population biology of infectious diseases: Part II. *Nature*, 280, 455–61.
*MEADOWS, D. H., MEADOWS, D. L., RANDERS, J. & BEHRENS, W. W. (1972). *The Limits to Growth.* New York: Universe Books.
MEHRA, R. K. & LAINIOTIS, D. G. (eds) (1976). *System Identification: Advances and Case Studies.* London: Academic Press.
*MESAROVIC, M. & PESTEL, E. (1975). *Mankind at the Turning Point.* London: Hutchinson.
MODE, C. J. (1971). *Multitype Branching Processes.* New York: American Elsevier.
MOLINEAUX, L., DIETZ, K. & THOMAS, A. (1978). Further epidemiological evaluation of a malaria model. *Bull. Wld Hlth Org.*, 56, 565–71.
MOLINEAUX, L. & GRAMICCIA, G. (1980). *The Garki Project: Research on the Epidemiology and Control of Malaria in the Sudan Savanna of West Africa.* Geneva: WHO.
MOLINEAUX, L., SHIDRAWI, G. R., CLARKE, J. L., BOULZAGUET, J. R. & ASHKAR, T. S. (1979). Assessment of insecticidal impact on the malaria mosquito's vectorial capacity, from data on the man-biting rate and age-composition. *Bull. Wld Hlth Org.*, 57, 265–74.
MOLINEAUX, L., STOREY, J., COHEN, J. E. & THOMAS, A. (1980). A longitudinal study of human malaria in the West African savanna in the absence of control measures: relationships between different *Plasmodium* species, in particular *P. falciparum* and *P. malariae*. *Amer. J. Trop. Med. Hyg.*, 29, 725–37.
MOSHKOVSKII, Sh. D. (1950). *Basic Laws of the Epidemiology of Malaria.* Moscow. (In Russian.)
MOSHKOVSKII, Sh. D. (1963). The dynamics of malaria eradication. *Proc. 7th Int. Congress Trop. Med. & Malaria*, 5, 219–20. (Rio de Janeiro.)
MOSHKOVSKII, Sh. D. (1967). A further contribution to the theory of malaria eradication. *Bull. Wld Hlth Org.*, 36, 992–6.
MOSHKOVSKII, Sh. D. (1969). The principles of quantitative epidemiology in arthropod-borne diseases. *WHO Inter-regional Travelling Seminar on Vector Ecology*. Moscow.
MUENCH, H. (1959). *Catalytic Models in Epidemiology.* Cambridge, Mass.: Harvard Univ. Press.
NÁJERA, J. A. (1974). A critical review of the field application of a mathematical model of malaria eradication. *Bull. Wld Hlth Org.*, 50, 449–57.
NÅSELL, I. (1977). On transmission and control of schistosomiasis, with comments on Macdonald's model. *Theor. Pop. Biol.*, 12, 335–65.
NÅSELL, I. (1980a). A hybrid version of the Ross malaria model. (Report No. TRITA-MAT-1980-2, Dept. Math. Royal Inst. Technology, Stockholm.)
NÅSELL, I. (1980b). A hybrid model for malaria with superinfection. (Report No. TRITA-MAT-1980-3, Dept. Math. Royal Inst. Technology, Stockholm.)
NÅSELL, I. (1980c). On the Macdonald–Irwin treatment of superinfection in malaria. (Report No. TRITA-MAT-1980-4, Dept. Math. Royal Inst. Technology, Stockholm.)
NÅSELL, I. & HIRSCH, W. H. (1973). The transmission of schistosomiasis. *Comm. Pure Appl. Math.*, 26, 395–453.
NNOCHIRI, E. (1975). *Medical Parasitology in the Tropics.* London: Oxford Univ. Press.
OLAOFE, G. O. & OLAOFE, K. (1975). A simple model for tropical malaria epidemics. *Math. Biosci.*, 25, 205–15.
ONORI, E. & GRAB, B. (1980a). Indicators for the forecasting of malaria epidemics. *Bull. Wld Hlth Org.*, 58, 91–8.
ONORI, E. & GRAB, B. (1980b). Quantitative estimates of the evolution of a malaria epidemic in Turkey if remedial measures had not been applied. *Bull. Wld Hlth Org.*, 58, 321–6.
PAMPANA, E. (1969). *A Textbook of Malaria Eradication* (2nd edn). London: Oxford Univ. Press.
PAYNE, D., GRAB, B., FONTAINE, R. E. & HEMPEL, J. H. G. (1976). Impact of control measures on malaria transmission and general mortality. *Bull. Wld Hlth Org.*, 54, 369–77.
*PICKLES, W. N. (1939). *Epidemiology in Country Practice.* Bristol: John Wright.
PRESCOTT, N. M. (1979a). Comment on paper by D. J. Bradley (1979). (See BRADLEY above.)
PRESCOTT, N. M. (1979b). The economics of malaria, filariasis and human trypanosomiasis. (Private circulation.)
PULL, J.-H. (1976). La notion de risque dans les maladies parasitaires avec référence particulière au paludisme. *Rev. Epidém. Santé Publ.*, 24, 221–9.
PULL, J.-H. (1977). Epidémiologie du paludisme, considérations pratiques. *Rev. Epidém. Santé Publ.*, 25, 405–6.
PULL, J.-H. & GRAB, B. (1974). A simple epidemiological model for evaluating the malaria inoculation rate and the risk of infection in infants. *Bull. Wld Hlth Org.*, 51, 507–16.
RADCLIFFE, J. (1973). The initial geographical spread of host–vector and carrier-borne epidemics. *J. Appl. Prob.*, 10, 703–17.

RADCLIFFE, J. (1974). The periodicity of malaria. *J. Appl. Prob.*, **11**, 562–7.
RADCLIFFE, J. (1976). The severity of a viral host-vector epidemic. *J. Appl. Prob.*, **13**, 791–4.
RAO, N. R., VIG, O. P. & AGARWALA, S. N. (1974a). Transmission dynamics of malaria. Quantitative studies – Part I: A stochastic model. *Bull. Haffkine Inst.*, **2**, 71–8.
RAO, N. R., VIG, O. P. & AGARWALA, S. N. (1974b). Transmission dynamics of malaria. Quantitative studies – Part II: Mathematical nature of expected happenings. *Bull. Haffkine Inst.*, **2**, 112–16.
RAO, N. R., VIG, O. P., AGARWALA, S. N., SABNIS, S. S. & RAO, V. N. (1975). Transmission dynamics of malaria: A computer simulation approach to measurement. (Computer Soc. of India: 10th Annual Convention, Ahmedabad, 16–18 January 1975.)
ROSS, R. (1909). *Report on the Prevention of Malaria in Mauritius.* London: Churchill.
ROSS, R. (1910). *The Prevention of Malaria.* London: Murray.
ROSS, R. (1911). *The Prevention of Malaria* (2nd edn). (With Addendum on the Theory of Happenings.) London: Murray.
ROSS, R. (1915). Some *a priori* pathometric equations. *Brit. Med. J.*, **1**, 546–7.
ROSS, R. (1916). An application of the theory of probabilities to the study of *a priori* pathometry, I. *Proc. Roy. Soc.*, A, **92**, 204–30.
ROSS, R. & HUDSON, H. P. (1917a). An application of the theory of probabilities to the study of *a priori* pathometry, II. *Proc. Roy. Soc.*, A, **93**, 212–25.
ROSS, R. & HUDSON, H. P. (1917b). An application of the theory of probabilities to the study of *a priori* pathometry, III. *Proc. Roy. Soc.*, A, **93**, 225–40.
SEVASTYANOV, B. A. (1951). Theory of branching stochastic processes. *Uspehi Mat. Nauk*, **6** (6), 47–99. (In Russian.)
SINGER, B. & COHEN, J. E. (1980). Estimating malaria incidence and recovery rates from panel surveys. *Math. Biosci.*, **49**, 273–305.
*SMITH, GEDDES (1943). *Plague on Us.* New York: Commonwealth Fund.
SOWUNMI, C. O. A. (1977). G. Adler's deterministic model of vector-borne epidemics – a threshold theorem. *Math. Biosci.*, **35**, 47–54.
SOWUNMI, C. O. A. On G. Adler's model of vector-borne epidemics. (Unpublished.)
STIRZAKER, D. (1976). Models for an epidemic. *J. Inst. Math. & Applic.*, **18**, 135–48.
TOMOVIĆ, R. & VUKOBRATOVIĆ. M. (1972). *General Sensitivity Theory.* New York: American Elsevier.
WALTON, G. A. (1947). On the control of malaria in Freetown, Sierra Leone, 1. *Plasmodium falciparum* and *Anopheles gambiae* in relation to malaria occurring in infants. *Ann. Trop. Med. Parasit.*, **41**, 380–407.
WAUGH, W. A. O'N. (1958). Conditioned Markov processes. *Biometrika*, **45**, 241–9.
WICHMANN, H. E. (1979). Asymptotic behaviour and stability in four models of venereal disease. *J. Math. Biol.*, **8**, 365–73.
WICKWIRE, K. H. (1977). Mathematical models for the control of pests and infectious diseases: a survey. *Theor. Pop. Biol.*, **11**, 182–238.
*WINSLOW, C. E. A. (1943). *The Conquest of Epidemic Disease.* Princeton, N.J.: Princeton Univ. Press.
*WORLD HEALTH ORGANIZATION (1978). *World Health*, Aug.-Sept. 1978, 38.
*WORLD HEALTH ORGANIZATION (1979). *WHO Expert Committee on Malaria (Seventeenth Report).* WHO Technical Report Series, No. 640.
*ZINSSER, H. (1935). *Rats, Lice and History.* Boston: Little, Brown & Co.

Index of subjects

Index of authors