Three Basic Epidemiological Models

Herbert W. Hethcote

Contents

1. Introduction

There are three basic types of deterministic models for infectious diseases which are spread by direct person-to-person contact in a population. Here these simplest models are formulated as initial value problems for systems of ordinary differential equations and are analysed mathematically. Theorems are stated regarding the asymptotic stability regions for the equilibrium points and phase plane portraits of solution paths are presented. Parameters are estimated for various diseases and are used to compare the vaccination levels necessary for herd immunity for these diseases. Although the three models presented are simple and their mathematical analyses are elementary, these models provide notation, concepts, intuition and foundation for considering more refined models. Some possible refinements are disease-related factors such as the infectious agent, mode of transmission, latent period, infectious period, susceptibility and resistance, but also social, cultural, Ecology by providing a sound intuitive understanding and complete proofs for the three most basic epidemiological models for microparasitic infections.

The study of disease occurrence is called epidemiology. An epidemic is an unusually large, short term outbreak of a disease. A disease is called endemic if it persists in a population. The spread of an infectious disease involves not only disease-related factors such as the infectious agent, mode of transmission, latent period, infectious period, susceptibility and resistance, but also social, cultural, demographic, economic and geographic factors. The three models considered here are the simplest prototypes of three different types of epidemiological models. It

S. A. Levin et al. (eds.), *Applied Mathematical Ecology*

Table 1. Classification of infectious diseases by agent and mode of transmission

	Mode of transmission			
Agent	Person → person	Person → environment environment → person	Reservoir → vector vector → person	Reservoir → person
Virus	Measles Chickenpox Mumps Rubella Smallpox Influenza Poliomyelitis Herpes HIV (AIDS virus)		Arboviruses: yellow fever dengue fever encephalitis tick fever sandfly fever	Rabies
Bacteria	Gonorrhea Tuberculosis Pneumonia Meningitis Strep throat	Typhoid fever Cholera	Plague	Brucellosis Tuleramia Anthrax
Protozoa	Syphillis	Amebiasis	Malaria Trypanosomiasis	
Helminths			Schistosomiasis Filariasis Onchocerciasis	Trichinosis

is important to understand their behaviour before considering general models incorporating more of the factors above.

Table 1 classifies diseases by agent and method of transmission. This useful classification scheme is similar to one presented by K. Dietz in 1974. The models considered here are suitable for diseases which are transmitted directly from person to person. More complicated models must be used when there is transmission by insects called vectors or a reservoir of nonhuman infectives. Epidemiological models are now widely used as more epidemiologists realize the role that modeling can play in basic understanding and policy development.

Justifications of mathematical modeling of the transmission of infectious diseases are given in the next section. The essential assumptions and terminology are given in Section 3. The SIS model analysed in Section 4 is for diseases for which infection does not confer immunity. SIR models for diseases where infection does confer immunity are considered for epidemics in Section 5 and for endemic situations in Section 6. Section 7 is devoted to herd immunity and its implication for vaccination for specific diseases. The discussion in Section 8 summarizes and refers to more complicated models.

2. Why Do Epidemiologic Modeling?

Even though vaccines are available for many infectious diseases, these diseases still cause suffering and mortality in the world, especially in developing countries. In developed countries chronic diseases such as cancer and heart disease have received more attention than infectious diseases, but infectious diseases are still a more common cause of death in the world. Recently, the human immunodeficiency virus (HIV) which can lead to acquired immunodeficiency syndrome (AIDS) has become an important infectious disease in both developing and developed countries.

The transmission mechanism from an infective to susceptibles is understood for nearly all infectious diseases and the spread of diseases through a chain of infections is known. However, the transmission interactions in a population are very complex so that it is difficult to comprehend the large scale dynamics of disease spread without the formal structure of a mathematical model. An epidemiological model uses a microscopic description (the role of an infectious individual) to predict the macroscopic behavior of disease spread through a population.

In many sciences it is possible to conduct experiments to obtain information and test hypotheses. Experiments with infectious disease spread in human populations are often impossible, unethical or expensive. Data is sometimes available from naturally occurring epidemics or from the natural incidence of endemic diseases; however, the data is often incomplete due to underreporting. This lack of reliable data makes accurate parameter estimation difficult so that it may only be possible to estimate a range of values for some parameters. Since repeatable experiments and accurate data are usually not available in epidemiology, mathematical models and computer simulations can be used to perform needed theoretical experiments. Calculations can easily be done for variety of parameter values and data sets.

Mathematical models have both limitations and capabilities that must recognized. Sometimes questions cannot be answered by using epidemiological models, but sometimes the modeler is able to find the right combination of available data, an interesting question and a mathematical model which can lead to the answer.

Comparisons can lead to a better understanding of the processes of disease spread. Modeling can often be used to compare different diseases in the same population, the same disease in different populations, or the same disease at different times. Comparisons of diseases such as measles, rubella, mumps, chickenpox, whooping cough, poliomyelitis and others are made in London and Yorke (1973), Yorke and London (1973), Yorke et al. (1979), Hethcote (1983), Anderson and May (1982) and in the article on rubella in this volume by Hethcote (1989).

Epidemiological models are useful in comparing the effects of prevention or control procedures. Hethcote and Yorke (1984) use models to compare gonorrhea control procedures such as screening, rescreening, tracing infectors, tracing infectees, post-treatment vaccination and general vaccination. Communicable disease models are often the only practical approach to answering questions about which prevention or control procedure is most effective. Quantitative predictions

of epidemiological models are always subject to some uncertainty since the models are idealized and the parameter values can only be estimated. However, predictions of the relative merits of several control methods are often robust in the sense that the same conclusions hold over a broad range of prameter values and a variety of models. Strategies for rubella vaccination are compared using a cost benefit analyses in the article on rubella by Hethcote (1989) in this volume.

Optimal strategies for vaccination can be found theoretically by using modeling. Longini, Ackerman and Elveback (1978) use a epidemic model to decide which age groups should be vaccinated first to minimize cost or deaths in an influenza epidemic. Hethcote (1988) uses a modeling approach to estimate the optimal age of vaccination for measles. A primary conclusion of this paper is that better data is needed on vaccine efficacy as a function of age in order to better estimate the optimal age of vaccination. Thus epidemiological modeling can be used to identify crucial data that needs to be collected.

An underrecognized value of epidemiological modeling is that it leads to a clear statement of the assumptions about the biological and sociological mechanisms which influence disease spread. The parameters used in an epidemiological model must have a clear interpretation such as a contact rate or a duration of infection. Models can be used to assess many quantitative conjectures. For example, one could check a conjecture that AIDS incidence would decrease if 90% of the sexually active heterosexual population started using condoms consistently. Epidemiological models can sometimes be used to predict the spread or incidence of a disease. For example, Hethcote (1983) predicted that rubella and Congenital Rubella Syndrome will eventually disappear in the United States because the current vaccination levels using the combined measles-mumps-rubella vaccine are significantly above the threshold required for herd immunity for rubella. An epidemiological model can also be used to determine the sensitivity of predictions to changes in parameter values. After the parameters are identified which have the greatest influence on the predictions, it may be possible to design studies to obtain better estimates of these parameters.

3. Assumptions and Notation

The population under consideration is divided into disjoint classes which change with time t. The susceptible class consists of those individuals who can incur the disease but are not yet infective. The infective class consists of those who are transmitting the disease to others. The removed class consists of those who are removed from the susceptible-infective interaction by recovery with immunity, isolation, or death. The fractions of the total population in these classes are denoted by $S(t)$, $I(t)$ and $R(t)$, respectively.

In the epidemiological models here, the following assumptions are made:

1. The population considered has constant size N which is sufficiently large so that the sizes of each class can be considered as continuous variables. If the model is to include vital dynamics, then it is assumed that births and natural deaths

occur at equal rates and that all newborns are susceptible. Individuals are removed by death from each class at a rate proportional to the class size with proportionality constant μ which is called the daily death removal rate. This corresponds to a negative exponential age structure with an average lifetime of $1/\mu$.

2. The population is homogenously mixing. The daily contact rate λ is the average number of adequate contacts per infective per day. An adequate contact of an infective is an interaction which results in infection of the other individual if he is susceptible. Thus the average number of susceptibles infected by an infective per day is λS, and the average number of susceptibles infected by the infective class with size NI per day in λSNI. The daily contact rate λ is fixed and does not vary seasonally. The type of direct or indirect contact adequate for transmission depends on the specific disease. The number of cases per day λSNI, which is called the incidence, is a mass action law since it involves the product of S and I.

3. Individuals recover and are removed from the infective class at a rate proportional to the number of infectives with proportionality constant γ, called the daily recovery removal rate. The latent period is zero (it is defined as the period between the time of exposure and the time when infectiousness begins). Thus the proportion of individuals exposed (and immediately infective) at time t_0 who are still infective at time $t_0 + t$ is $\exp(-\gamma t)$, and the average period of infectivity is $1/\gamma$ (Hethcote, Stech and van den Driessche, 1981c).

The removal rate from the infective class by both recovery and death is $\gamma + \mu$ so that the death-adjusted average period of infectivity is $1/(\gamma + \mu)$. Thus the average number of adequate contacts (with both susceptibles and others) of an infective during the infectious period is $\sigma = \lambda/(\gamma + \mu)$, which is called the contact number. This quantity is also called the basic reproductive rate (Anderson and May, 1981, 1982; May, 1986) even though it is a number and not a rate. Since the average number of susceptibles infected by an infective during the infectious period is σS, the quantity σS is called the replacement number.

If recovery does not give immunity, then the model is called an SIS model, since individuals move from the susceptible class to the infective class and then back to the susceptible class upon recovery. If individuals recover with permanent immunity, then the model is an SIR model. If individuals recover with temporary immunity so that they eventually become susceptible again, then the model is an SIRS model as considered in Hethcote (1976) and Hethcote, Stech and van den Driessche (1981a). If individuals do not recover, then the model is an SI model. In general, SIR models are appropriate for viral agent diseases such as measles, mumps, and smallpox, while SIS models are appropriate for some bacterial agent diseases such as meningitis, plague, and venereal diseases, and for protozoan agent diseases such as malaria and sleeping sickness (see Table 1).

A basic concept in epidemiology is the existence of thresholds; these are critical values for quantities such as the contact number, population size or vector density that must be exceeded in order for an epidemic to occur or for a disease to remain endemic. The formulations used here are somewhat different from the more classical formulations of Hamer (1906), Ross (1911), Kermack and McKendrick (1927) and others, as given in Bailey (1975). Here they involve the fractions of the populations in the classes instead of the numbers in the classes because these formulations have

much more intutive threshold conditions involving the contact number instead of the population sizes. See Hethcote (1976, p. 339) or Hethcote and Van Ark (1987) for further comparisons of formulations in terms of proportions and numbers in the classes.

4. The SIS Model

The first model is for diseases for which infection does not confer immunity. It is called an SIS model since individuals return to the susceptible class when they recover from the infection. Using the notation in Section 3, the compartmental diagram for an SIS model is given in Fig. 1. Naturally occurring births and deaths (vital dynamics) are included, but the behavior of solutions is similar when vital dynamics are not included.

The initial value problem (IVP) for this SIS model formulated in terms of class sizes is

$$\begin{aligned} (NS(t))' &= -\lambda SNI + \gamma NI + \mu N - \mu NS \\ (NI(t))' &= \lambda SNI - \gamma NI - \mu NI \\ NS(0) &= NS_0 > 0, \quad NI(0) = NI_0 > 0, \quad NS(t) + NI(t) = N \end{aligned} \tag{4.1}$$

where λ is a positive constant and primes denote derivatives with respect to time t. If each equation above is divided by the constant population size N, then the IVP in terms of the fractions in the classes is

$$\begin{aligned} S'(t) &= -\lambda IS + \gamma I + \mu - \mu S \\ I'(t) &= \lambda IS - \gamma I - \mu I \\ S(0) &= S_0 > 0, \quad I(0) = I_0 > 0, \quad S(t) + I(t) = 1. \end{aligned} \tag{4.2}$$

Note that the IVP (4.2) involves the daily contact and removal rates, but not the population size N. This model is appropriate for some bacterial agent diseases such as gonorrhea, meningitis and streptococcal sore throat. Here all parameters in (4.2) are nonnegative and only nonnegative solutions are considered since negative solutions have no epidemiological significance.

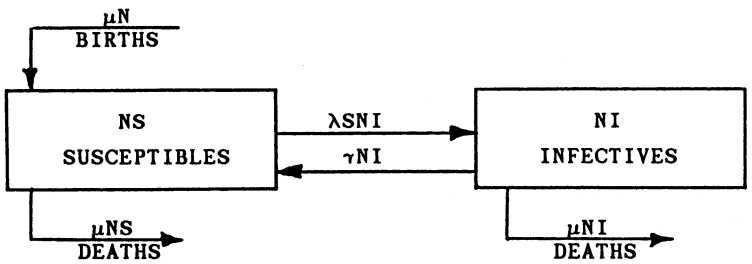

Fig. 1. The compartmental diagram for the SIS model.

Since $S(t)$ can be found from $I(t)$ by using $S(t) = 1 - I(t)$, it is sufficient to consider

$$I'(t) = [\lambda - (\gamma + \mu)]I - \lambda I^2$$
$$I(0) = I_0 > 0. \tag{4.3}$$

Since this is a Bernoulli differential equation, the substitution $y = I^{-1}$ converts (4.3) into a linear differential equation from which the unique solution of (4.3) is found to be

$$I(t) = \begin{cases} \dfrac{e^{(\gamma+\mu)(\sigma-1)t}}{\sigma[e^{(\gamma+\mu)(\sigma-1)t} - 1]/(\sigma - 1) + 1/I_0} & \text{for } \sigma \neq 1 \\ \dfrac{1}{\lambda t + 1/I_0} & \text{for } \sigma = 1 \end{cases} \tag{4.4}$$

where σ is the contact number $\lambda/(\gamma + \mu)$ defined in Section 3. The theorem below follows from the explicit solution (4.4).

Theorem 4.1. *The solution* $I(t)$ *of* (4.3) *approaches* $1 - 1/\sigma$ *as* $t \to \infty$ *if* $\sigma > 1$ *and approaches* 0 *as* $t \to \infty$ *if* $\sigma \leqq 1$.

This theorem means that for a disease without immunity with any positive initial infective fraction, the infective fraction approaches a constant endemic value if the contact number exceeds 1; otherwise, the disease dies out. Although the model (4.2) reduces to a one dimensional IVP (4.3), we show SI phase diagrams for this model in Fig. 2 so that they can be compared with the phase diagrams for the other models.

Here the threshold quantity is the contact number σ and the critical threshold value is 1. Note that the replacement number σS is 1 at the endemic equilibrium point. A threshold result for an SI model is obtained from Theorem 4.1 by taking the removal rate γ to be zero in the model. If both the removal rate γ and the birth and death rate μ are zero, then $\sigma = \infty$ so that there is no threshold and eventually everyone is infected. This model with $\gamma = \mu = 0$ is the "simple epidemic

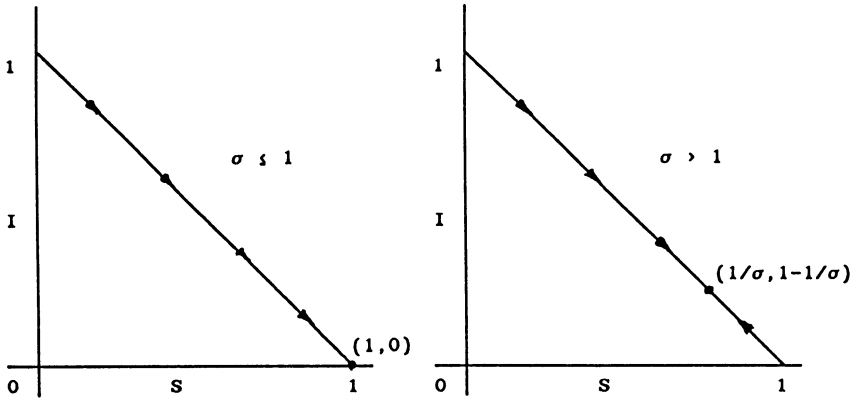

Fig. 2. Phase diagrams for the SIS model. Note that the paths are on the line $S + I = 1$.

model" considered in Bailey (1975, p. 20). Some authors such as May (1986) use other terminology; he uses the basic reproductive rate R_0 is place of the contact number σ and effective reproductive rate $R_0 S$ instead of the replacement number σS.

The prevalence is defined as the number of cases of a disease at a given time so that it corresponds to NI. Since the incidence is defined to be the number of new cases per unit time, it corresponds to the λSNI term in model (4.1). At an endemic equilibrium the prevalence is equal to the incidence times the average duration of infection $1/(\gamma + \mu)$ since the right side of the second equation in (4.1) is zero at an equilibrium.

The incidence and the prevalence of some diseases oscillate seasonally in a population. This oscillation seems to be caused by seasonal oscillation in the contact rate λ. For example, the incidence of childhood diseases such as measles and rubella increase each year in the winter when children aggregate in schools (London and Yorke, 1973; Yorke and London, 1973; Dietz, 1976; Schenzle, 1984).

If the contact rate λ changes with time t, then the λ in models (4.1)–(4.3) are replaced by $\lambda(t)$. If $\lambda(t)$ is periodic with period p, then Hethcote (1973) has found the asymptotic behavior of solutions $I(t)$ of (4.3). If the average contact number $\bar{\sigma} = \bar{\lambda}/(\gamma + \mu)$ satisfies $\bar{\sigma} \leqq 1$, then $I(t)$ damps in an oscillatory manner to 0 for large t. However, if $\bar{\sigma} > 1$, then $I(t)$ approaches an explicit periodic solution for large t. These behaviors are shown in Figs. 3 and 4.

Gonorrhea is an example of a disease for which infection does not confer immunity. Fig. 5 shows the actual seasonal oscillation of reported cases of gonorrhea from 1946 to 1984. Numerous models for gonorrhea transmission dynamics and control including a seasonal oscillation model are presented in Hethcote and Yorke (1984).

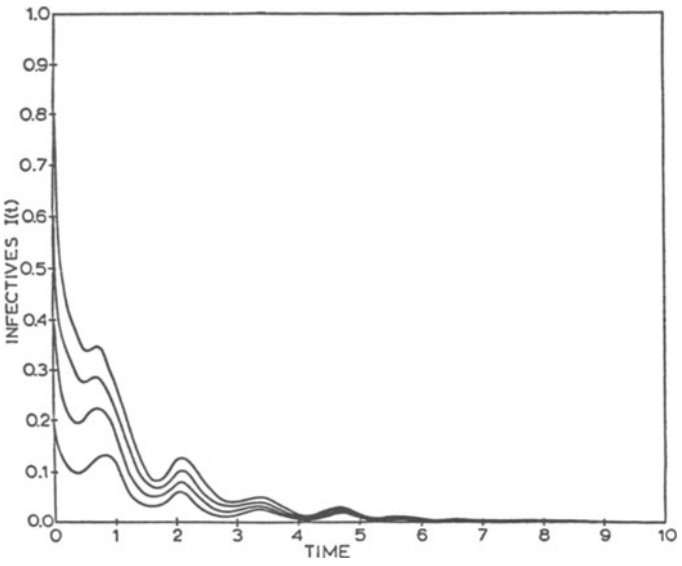

Fig. 3. Solutions of initial value problem (4.3) with periodic $\lambda(t)$ and various values of I_0. Here $\lambda(t) = 2 - 1.8 \cos 5t$ and $\gamma = 4$ so that the average contact number is $\bar{\sigma} = 0.5 < 1$. From Hethcote (1973).

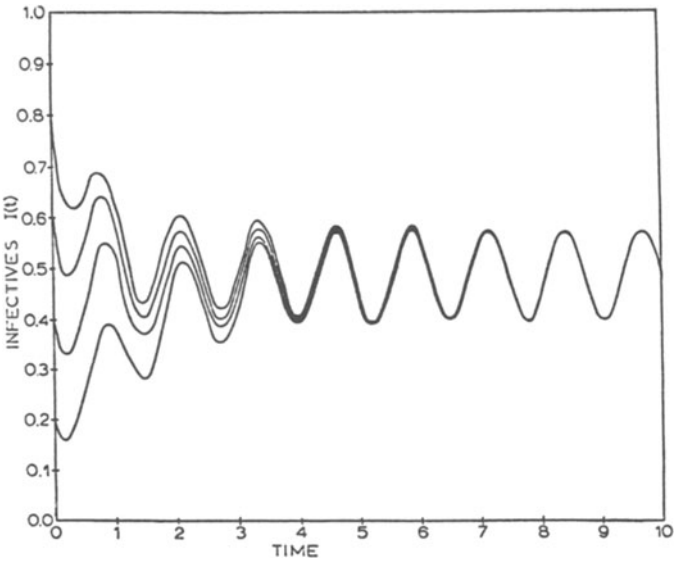

Fig. 4. Solutions as in Fig. 3 except that here $\gamma = 1$ so that $\bar{\sigma} = 2 > 1$. From Hethcote (1973).

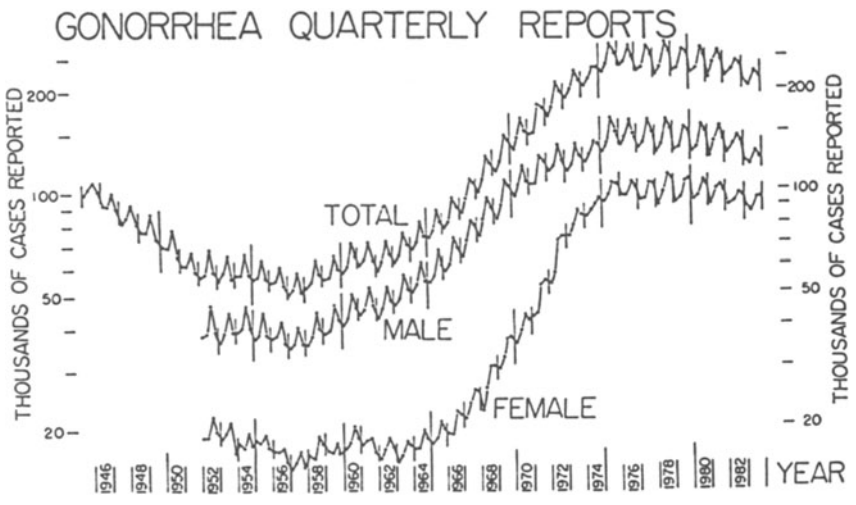

Fig. 5. Reported cases of gonorrhea in women and men in the United States. From Hethcote and Yorke (1984).

5. The SIR Model Without Vital Dynamics

Here and in Sect. 6 we consider diseases for which infection confers permanent immunity. When such an SIR disease goes through a population in a relatively short time (less than one year), then this disease outbreak is called an epidemic.

Fig. 6. The compartmental diagram for the SIR model without vital dynamics.

Since an epidemic occurs relatively quickly, the model does not include births and deaths (vital dynamics). Epidemics are common for diseases such as influenza, measles, rubella and chickenpox. Using the notation in Sect. 3, the compartmental diagram for this model is given in Fig. 6.

The initial value problem (IVP) for the SIR model without vital dynamics given in Fig. 6 is

$$\begin{aligned} &(NS(t))' = -\lambda SNI \\ &(NI(t))' = \lambda SNI - \gamma NI \\ &(NR(t))' = \gamma NI \\ &\quad NS(0) = NS_0 > 0, \quad NI(0) = NI_0 > 0, \quad NR(0) = NR_0 \geqq 0 \\ &NS(t) + NI(t) + NR(t) = N \end{aligned} \tag{5.1}$$

where λ and γ are positive constants.

If each equation in (5.1) in divided by the constant population size N, then the IVP for the fractions $S(t)$ and $I(t)$ is

$$\begin{aligned} &S'(t) = -\lambda SI \\ &I'(t) = \lambda SI - \gamma I \\ &S(0) = S_0 > 0, \quad I(0) = I_0 > 0. \end{aligned} \tag{5.2}$$

Since $R(t)$ can always be found from $S(t)$ and $I(t)$ by using $R(t) = 1 - S(t) - I(t)$, it is sufficient to consider the IVP (5.2) in the SI phase plane. The epidemiologically reasonable region in the SI plane is the triangle given by

$$T = \{(S, I) | S \geqq 0, I \geqq 0, S + I \leqq 1\}. \tag{5.3}$$

Theorem 5.1. *Let* $(S(t), I(t))$ *be the solutions of* (5.2). *If* $\sigma S_0 \leqq 1$, *then* $I(t)$ *decreases to zero as* $t \to \infty$. *If* $\sigma S_0 > 1$, *then* $I(t)$ *first increases up to a maximum value* Im *equal to* $1 - R_0 - 1/\sigma - [\ln(\sigma S_0)]/\sigma$ *and then decreases to zero as* $t \to \infty$. *The susceptible fraction* $S(t)$ *is a decreasing function and the limiting value* $S(\infty)$ *is the unique root in* $(0, 1/\sigma)$ *of the equation*

$$1 - R_0 - S(\infty) + [\ln(S(\infty)/S_0)]/\sigma = 0. \tag{5.4}$$

The threshold quantity in Theorem 5.1 is the initial replacement number σS_0 where $\sigma = \lambda/\gamma$ is the contact number. The natural logarithm is denoted by ln. This theorem states that if the initial replacement number is greater than one, then an epidemic occurs since the prevalence (the infective fraction) increases to a peak and then decreases to zero. Otherwise, there is no epidemic since the prevalence decreases to zero. The infection spread stops during an epidemic because the replacement number $\sigma S(t)$ becomes less than one when $S(t)$ becomes small; however,

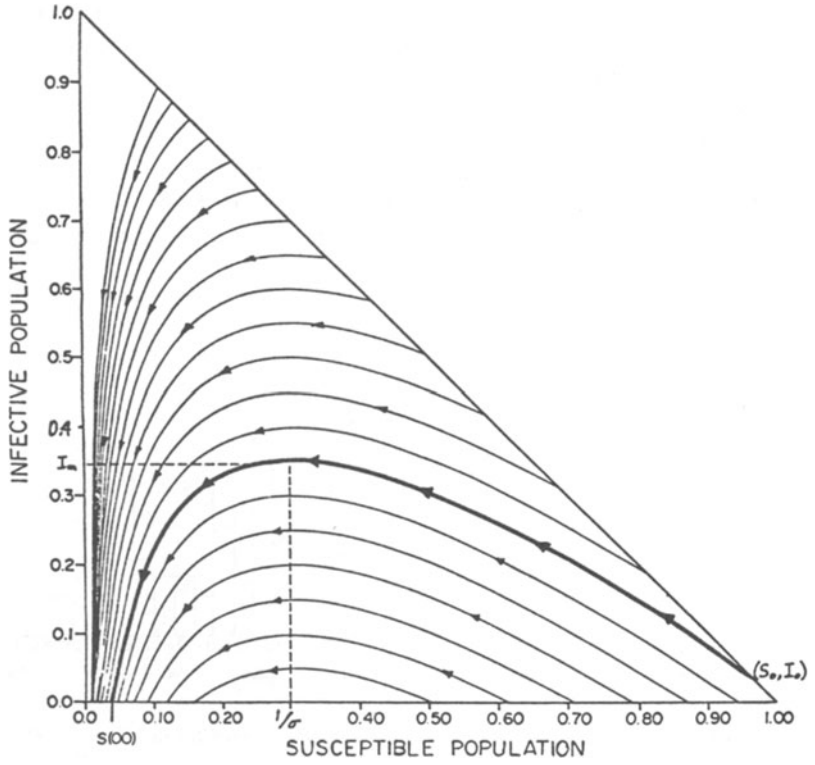

Fig. 7. Phase diagram for the SIR model without vital dynamics with $1/\sigma = 0.30$.

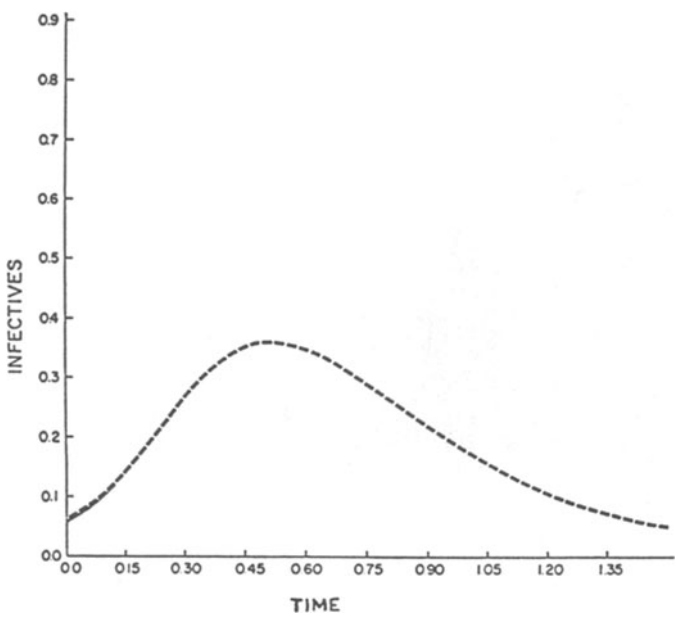

Fig. 8. An epidemic curve starting at $I_0 = 0.06$ with $\lambda = 10$ and $\gamma = 3$ so that $\sigma = 10/3$.

the final susceptible population $S(\infty)$ is not zero. A phase portrait corresponding to system (5.2) is given in Fig. 7. The proof of Theorem 5.1 is given in the Appendix.

Figure 8 shows an epidemic curve which is the prevalence $I(t)$ as a function of time; the incidence or number of new cases per day would also increase to a peak and then decrease. Incidences for examples of epidemics are given in Figs. 9 to 13.

If an epidemic occurs in a homogeneous population and there is no vaccination during the epidemic, then it is possible to estimate the contact number for the disease in that population from epidemic data (Hethcote and Van Ark, 1987). Since

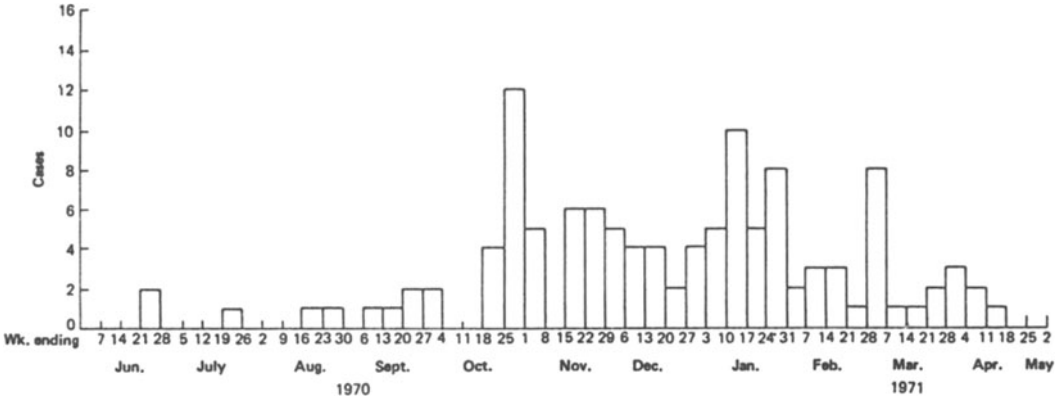

Fig. 9. An epidemic curve for infectious hepatitis in Barren County, Kentucky, USA in 1970 and 1971. The data are irregular, but the general shape is consistent with Fig. 8. Figure from CDS (1971a).

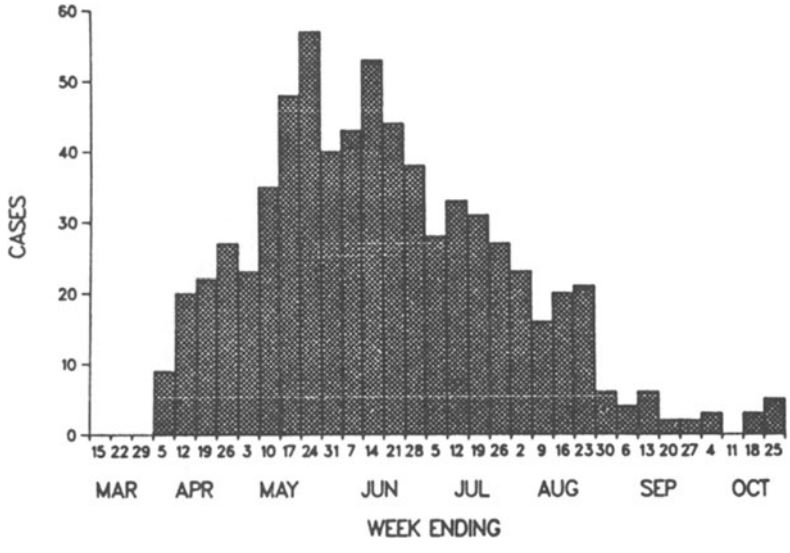

Fig. 10. Cases of non-A, non-B hepatitis in a refugee camp in Tug Wajale, Somalia from March 15 to October 25, 1986. Figure from CDS (1987b).

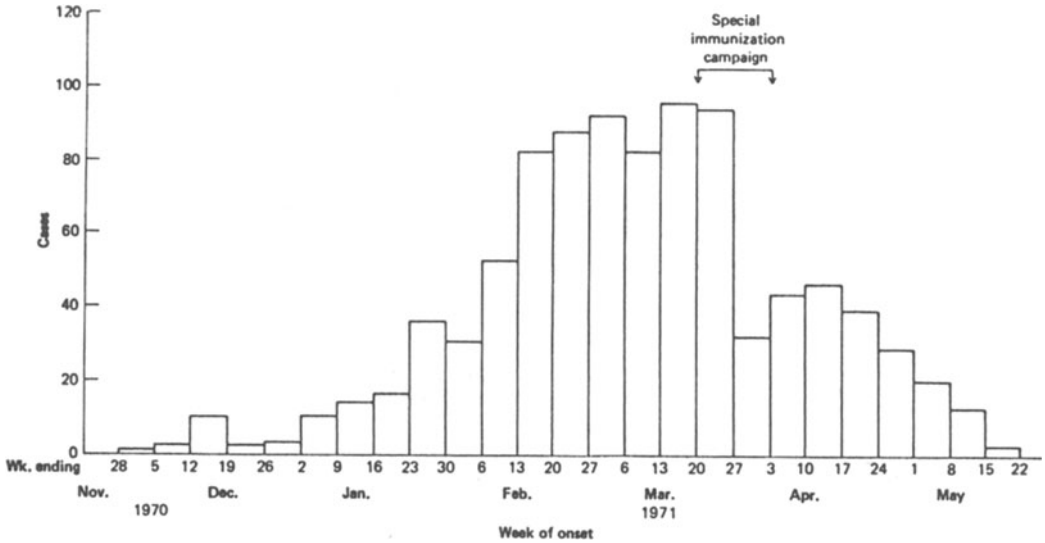

Fig. 11. An epidemic curve for measles cases in Dallas, Texas, USA in 1970 and 1971. Note the interruption due to the special immunization campaign. Figure from CDS (1971b).

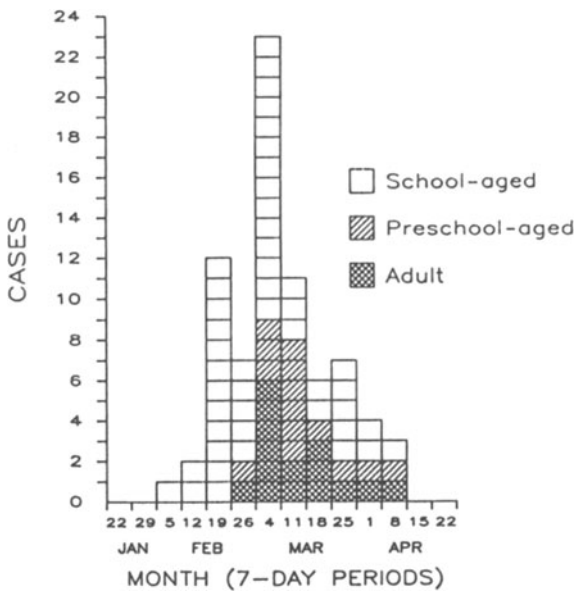

Fig. 12. Measles cases by data of onset in Hobbs, New Mexico from January 22 to April 22, 1984. Figure from CDC (1984).

an epidemic enters a population as one or very few cases, I_0 is negligibly small so that $S_0 = 1 - R_0$. Then (5.4) can be solved for σ to obtain

$$\sigma = \frac{\ln(S_0/S(\infty))}{S_0 - S(\infty)}. \tag{5.5}$$

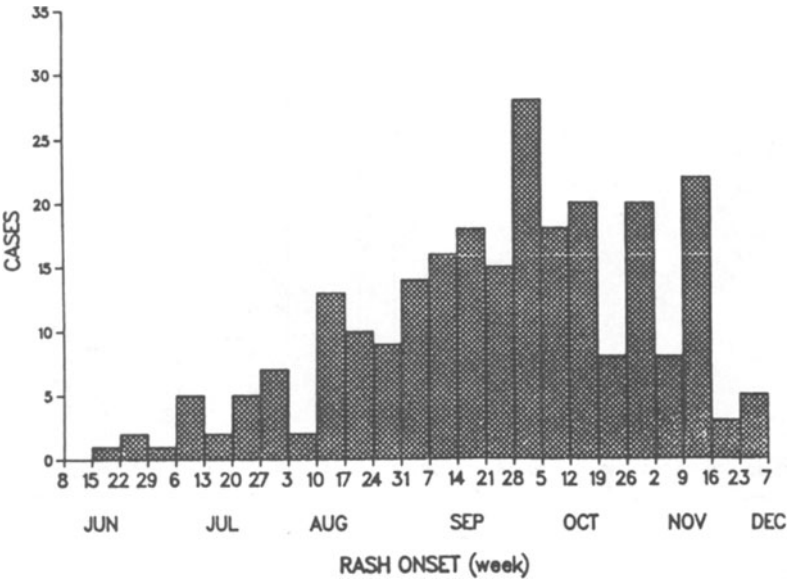

Fig. 13. Reported measles cases by data of rash onset in Dade County, Florida in 1986. The data are somewhat irregular since the latent period for measles is about 2 weeks. Figure from CDC (1987a).

If the susceptible fractions before the epidemic (S_0) and after the epidemic ($S(\infty)$) are measured by serologic studies (i.e., testing immune responses in blood samples), then the contact number σ can be estimated using (5.5).

Evans (1982) reports on serosurveys conducted on freshman at Yale University. The fractions susceptible to rubella at the beginning and end of their freshman year were 0.25 and 0.0965 so that (5.5) leads to the estimate $\sigma = 6.2$. For influenza, the fraction susceptible at the start and end of their freshman year were 0.911 and 0.5138, which leads to a contact number estimate of $\sigma = 1.44$.

6. The SIR Model with Vital Dynamics

An SIR epidemiological model is considered as in Sect. 5, but here we model the disease behavior in the population over a long time period. A disease is called endemic if it is present in a population for more than 10 or 20 years. Because of the long time period involved, a model for an endemic disease must include births as a source of new susceptibles and natural deaths in each class. Using the notation and assumptions in Sect. 3, the compartmental diagram for the SIR model with vital dynamics is given in Fig. 14.

The initial value problem (IVP) for the SIR model with vital dynamics is

$$\begin{aligned}(NS(t))' &= -\lambda SNI + \mu N - \mu NS\\ (NI(t))' &= \lambda SNI - \gamma NI - \mu NI\\ (NR(t))' &= \gamma NI - \mu NR\end{aligned} \tag{6.1}$$

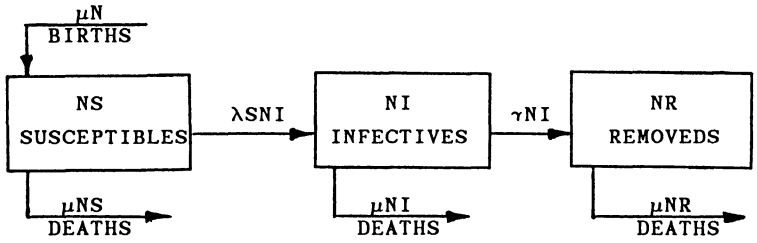

Fig. 14. The compartmental diagram for the SIR model with vital dynamics.

$$NS(0) = NS_0 > 0, \quad NI(0) = NI_0 \geqq 0, \quad NR(0) = NR_0 \geqq 0$$
$$NS(t) + NI(t) + NR(t) = N$$

where the contact rate λ, the removal rate constant γ and the death rate constant μ are positive constants.

If each equation in (6.1) is divided by N, then the IVP in terms of $S(t)$ and $I(t)$ is

$$\begin{aligned} S'(t) &= -\lambda SI + \mu - \mu S \\ I'(t) &= \lambda SI - \gamma I - \mu I \\ S(0) &= S_0 > 0, \quad I(0) = I_0 \geqq 0. \end{aligned} \tag{6.2}$$

As in Section 5, it is sufficient to consider the IVP (6.2) since $R(t)$ is given by $R(t) = 1 - S(t) - I(t)$. The asymptotic behaviors of solution paths in the SI phase plane are described in the following theorem.

Theorem 6.1. *If $\sigma \leqq 1$, then the triangle T defined by* (5.3) *is an asymptotic stability region for the equilibrium point* $(1,0)$. *If $\sigma > 1$, then $T - \{(S,0) | 0 \leqq S \leqq 1\}$ is an asymptotic stability region for the equilibrium point*

$$(1/\sigma, \mu(\sigma - 1)/\lambda). \tag{6.3}$$

Figures 15 and 16 are phase plane portraits for the two possibilities described in the theorem. The theorem above can be explained intuitively in terms of the contact number $\sigma = \lambda/(\gamma + \mu)$, which is the threshold quantity. If the contact number is less than one so that an infective replaces itself with less than one new infective, then the disease dies out. Moreover, the susceptible fraction eventually approaches one since everyone is susceptible when the disease has disappeared and all of the removed people who are immune have died.

If the contact number is greater than one, the initial infective fraction I_0 is small, and the initial susceptible fraction S_0 is large so that $\sigma S_0 > 1$, then S decreases and I first increases to a peak and then decreases just as it would for an epidemic (compare Figs. 16 and 8). However, after the infective fraction has decreased to a low level, the susceptible fraction slowly starts to increase due to the births of new susceptibles. When the susceptible fraction gets large enough, there is a second smaller epidemic and so on as the path spirals into the equilibrium point (6.3). At this endemic equilibrium point, the replacement number σS is 1 since if the replacement number were greater or less than 1, then the infective fraction would

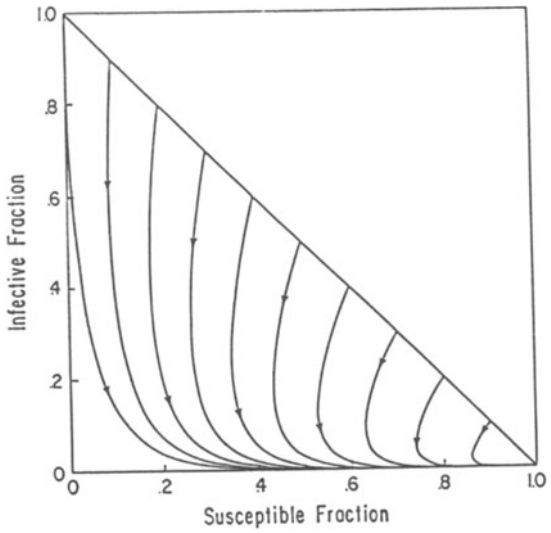

Fig. 15. Phase plane portrait for SIR model with vital dynamics when the contact number is $\sigma = 0.5 < 1$. From Hethcote (1976).

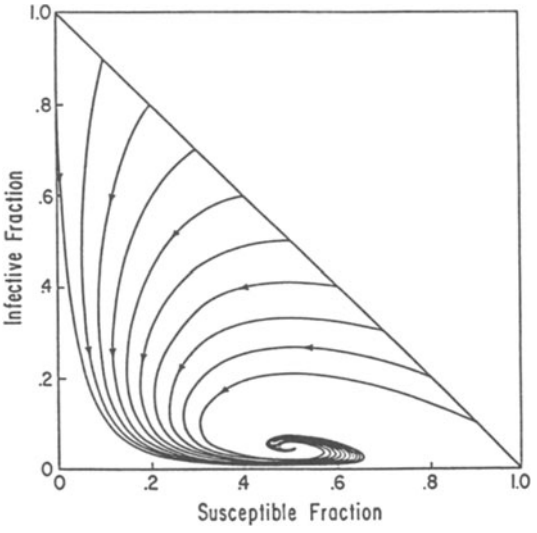

Fig. 16. Phase plane portrait for SIR model with vital dynamics when the contact number is $\sigma = 2 > 1$. From Hethcote (1976).

be increasing or decreasing, respectively. One advantage of precise threshold results such as Theorem 6.1 is that the effects of changes in the parameter values on the asymptotic behavior can be determined directly. The proof of Theorem 6.1 is given in the Appendix.

A method was presented at the end of Section 5 for estimating the contact number from epidemic data. There are two ways to estimate contact numbers from data for SIR diseases which are endemic. The first method involves estimating the susceptible fraction S_e from a serological survey (i.e., testing immune responses in blood samples). It is assumed that the sample is randomly chosen from a

homogeneously mixing population and that the disease has reached an endemic equilibrium in the population. Since $S_e = 1/\sigma$ in (6.3), the contact number σ can be estimated using

$$\sigma = 1/S_e. \tag{6.4}$$

This estimation is also valid for the SIS model considered in Section 4 and for endemic SIR disease when part of the population has been vaccinated (Hethcote and Van Ark, 1987).

The second method is useful for childhood diseases in which the susceptible fraction is a decreasing function of age. Dietz (1975) has used an age-structured model in an unvaccinated homogeneously mixing population to derive the formula

$$\sigma = 1 + L/A \tag{6.5}$$

where L is the average lifetime and A is the average age of attack for an endemic disease. Thus if L and A are estimated for an endemic disease in a population, then (6.9) can be used to estimate the contact number σ.

The formula (6.5) can be derived heuristically from the model (6.2). The incidence rate at the endemic equilibrium is $\lambda I_e S_e$ so that λI_e is the incidence proportionality constant. The waiting time to get the infection is distributed as a negative exponential so that the average age at first infection is

$$A = 1/\lambda I_e = 1/[\mu(\sigma - 1)]. \tag{6.6}$$

Solving this equation for σ yields (6.5) since the average lifetime L is $1/\mu$.

7. Herd Immunity and Vaccination

A population is said to have herd immunity for a disease if enough people are immune so that the disease would not spread if it were suddenly introduced somewhere in the population. If the population is homogeneously mixing and the immune people are distributed uniformly in the population, then herd immunity will be obtained if a large enough uniformly distributed fraction is immune. The contact number σ gives the average number of adequate contacts (i.e., those which are sufficient for transmission if all contacted people were susceptible) of an infective during the infectious period. In order to prevent the spread of infection from an infective, enough people must be immune so that the replacement number satisfies $\sigma S < 1$. That is, the susceptible fraction must be small enough so that the average infective infects less than one person during the infectious period.

Herd immunity in a population is achieved by vaccination of susceptibles in the population. If R is the fraction of the population which is immune due to vaccination, then since $S = 1 - R$ when $I = 0$, herd immunity is achieved if $\sigma(1 - R) < 1$ or

$$R > 1 - 1/\sigma. \tag{7.1}$$

For example, if the contact number is 5, at least 80% must be immune to have

Table 2. Estimates of contact numbers and herd immunity fractions from data (Anderson, 1982) on average ages of attack and average lifetimes

Disease	Location	A	L	$\sigma = 1 + A/L$	Minimum R for herd immunity
Measles	England and Wales, 1956–1959	4.8	70	15.6	0.94
	USA, 1912–1928	5.3	60	12.3	0.92
	Nigeria 1960–1968	2.5	40	17.0	0.94
Whooping cough	Maryland, USA, 1943	4.3	70	17.3	0.94
	England and Wales, 1944–1978	4.5	70	16.5	0.94
Chickenpox	Maryland, USA, 1943	6.8	70	11.3	0.91
Diphtheria	Virginia and New York, USA 1934–1947	11.0	70	7.4	0.86
Scarlet fever	Maryland, USA, 1908–1917	8.0	60	8.5	0.88
Mumps	Maryland, USA, 1943	9.9	70	8.1	0.88
Rubella	England and Wales, 1979	11.6	70	7.0	0.86
	West Germany, 1972	10.5	70	7.7	0.87
Poliomyelitis	USA, 1955	17.9	70	4.9	0.80
	Netherlands, 1960	11.2	70	4.3·	0.86
Smallpox	India	12	50	5.2	0.81

herd immunity. If σ is 10, then 90% must be immune for herd immunity. If σ is 20, then 95% immunity corresponds to herd immunity. These results are intuitively reasonable since a higher contact number corresponds to a more easily spread disease so that a larger percentage must be immune to achieve herd immunity.

Table 2 contains data (Anderson, 1982) on the average age A of attack and the average lifetime L for various diseases. The estimates of contact numbers σ in Table 2 are calculated using (6.5) and the minimum immune fraction R is estimated using (7.1). Although the estimates of contact numbers σ in Table 2 are based on many simplifying assumptions, they do lead to crude comparisions of the approximate immunity levels necessary for herd immunity for these diseases.

Attainment of herd immunity for a disease can be quite difficult. Although smallpox was eliminated by vaccination from most developed countries by 1958, it remained endemic in some developing countries. The World Health Organization started a program in 1958 to eradicate smallpox throughout the world (WHO, 1980). Even though high vaccination percentages were achieved in some countries, the disease persisted, primarily because the vaccinations were not uniformly distributed in the population. Eventually the disease was eliminated from more and more countries until the last case occurred in Somalia in 1977. The eradication of smallpox was partly due to herd immunity and partly due to containment efforts such as surveillance, patient isolation and vaccination of all possible contacts when a case occurred (Fenner, 1983). The contact number for smallpox is estimated (see Table 2) to be 5 from data in India. Since eradication in the world of smallpox, which has a low contact number, was difficult, it seems that eradication in the

world of diseases with higher contact numbers would be even more difficult.

For various reasons a small fraction of those who are vaccinated do not become immune. This fraction of primary vaccine failures is usually about 0.05 or 0.10, but it can be 0.2 or 0.4 for some influenza vaccines. Vaccine efficacy (VE) is defined as the fraction of those vaccinated who become immune. For example, for measles or rubella vaccination at age 15 months, the vaccine efficacy is approximately 0.95 (Hethcote, 1983). Since the immune fraction R satisfies $R = (V)$ (VE) where V is the vaccinated fraction in the population, inequality (7.1) implies that herd immunity is achieved if the vaccinated fraction V satisfies

$$V > (1 - 1/\sigma)/\text{VE}. \tag{7.2}$$

Measles and rubella have some similarities so it is interesting to compare them. The contact number for rubella is approximately 7 so that herd immunity is obtained if the immune fraction R satisfies $R > 0.86$. If a vaccine efficacy of 0.95 is used, then herd immunity occurs in a homogeneously mixing population if the vaccinated fraction V satisfies $V > 0.91$. The contact number for measles is approximately 15 in a modern developed country so that herd immunity occurs if the immune fraction R satisfies $R > 0.94$. If the vaccine efficiency is 0.95, then herd immunity is achieved if the vaccinated fraction V satisfies $V > 0.99$.

It initially appears that measles may be about twice as difficult to eradicate by herd immunity as rubella since for herd immunity, the *unimmune percentage* must theoretically be less than 14% for rubella and less than 6% for measles. However, it is actually much harder to achieve herd immunity for measles since the *unvaccinated percentage* must be less than 9% for rubella and less than 1% for measles. Indeed, in the USA measles has persisted despite major elimination efforts, while rubella incidence seems to be decreasing (CDC, 1981; CDC, 1986b). Although measles is very difficult or impossible to eradicate with a one dose program, it is easier to achieve herd immunity with a two dose program (Hethcote, 1983).

Hence, although the endemic SIR model is very simple, it has been possible to estimate parameters from it and to use these estimates to get a rough comparision between the immune fractions necessary for herd immunity for various diseases. For further discussion of models with vaccination and applications, see Hethcote (1978), Anderson (1982), Anderson and May (1982, 1983, 1985), Hethcote (1983), May (1986) and Hethcote and Van Ark (1987).

8. Discussion

The SIS model in Sect. 4 and the SIR model with vital dynamics in Sect. 6 have two intuitively appealing features. The first is that the disease dies out if the contact number σ satisfies $\sigma \leqq 1$ and the disease remains endemic if $\sigma > 1$. The second is that at an endemic equilibrium, the replacement number is 1; i.e., the average infective replaces itself with one new infective during the infectious period. Although the contact number threshold criterion is the same for diseases without and with immunity, the infective fraction approached asymptotically for large time

is higher for diseases without immunity than for diseases with immunity (compare Figs. 2 and 16). In Hethcote, Stech and van den Driessche (1981c) the Soper (1929) model for an SIR disease with vital dynamics is shown to be ill-posed since some solution paths leave the triangle T and R becomes negative.

By comparing Theorems 5.1 and 6.1 it is clear that the asymptotic behaviors for SIR models without and with vital dynamics are very different. The SIR model without vital dynamics might be appropriate for describing an epidemic outbreak during a short time period, whereas the SIR model with vital dynamics would be appropriate over a longer time period. Viral agent diseases such as measles, chickenpox, mumps, and influenza may have occasional large outbreaks in certain communities and yet be endemic at a low level in larger population groups. The threshold quantity for the SIR model without vital dynamics is the initial replacement number σS_0. In this model, no epidemic occurs if $\sigma S_0 < 1$ and an epidemic occurs if $\sigma S_0 > 1$.

The latent period is the time in which an individual is infected but is not yet infectious. The latent period is approximately 15 days for chickenpox, 10 days for measles, and 2 days for influenza. The latent period has been ignored in the three basic models considered here because the thresholds and asymptotic behaviors are essentially the same for the models which include latent periods. The fraction of the population that is in the latent period is often called $E(t)$ or the exposed fraction. Various SEIS models are analysed in Hethcote, Stech and van den Driessche (1981b). Some SEIR models with vital dynamics are considered in Hethcote and Tudor (1980). Longini (1986) shows that the formula (5.5) also holds for an SEIR model without vital dynamics.

Instead of assumption 2 in Section 3, it is sometimes assumed that susceptibles become infectious at a rate proportional to the product of the number of susceptibles NS and the number of infectives NI with proportionality constant β. By comparing the resulting initial value problem with (4.1), (5.1) or (6.1), we see that $\beta = \lambda/N$ and thus the assumption that β is constant implies that the daily contact rate λ is proportional to the population size N. The daily contact rate would probably increase if the population within a fixed region increased (i.e., the population density increased). However, it seems more likely that the daily contact rate λ is independent of population size since λ might be the same for a large population in a large region and a small population in a small region. Hethcote and Van Ark (1987) consider model formulation for heterogeneous populations and discuss a "city and villages" model where confusion between β and λ has led to misleading results. Consequently, it seems best to carefully separate the daily contact rate λ and the population size N as we have done in assumption 2. Moreover, threshold statements involving contact numbers are more appealing intuitively than the population size threshold statements as given in Bailey (1975).

Although the models discussed here do provide some insights and useful comparisons, most models now being applied to specific diseases are more complicated. Hethcote, Stech and van den Driessche (1981c) have surveyed the mathematical epidemiology literature using the classifications introduced in this article. More recent references are given below for some more refined models. Many more complicated models are considered in other articles in this volume.

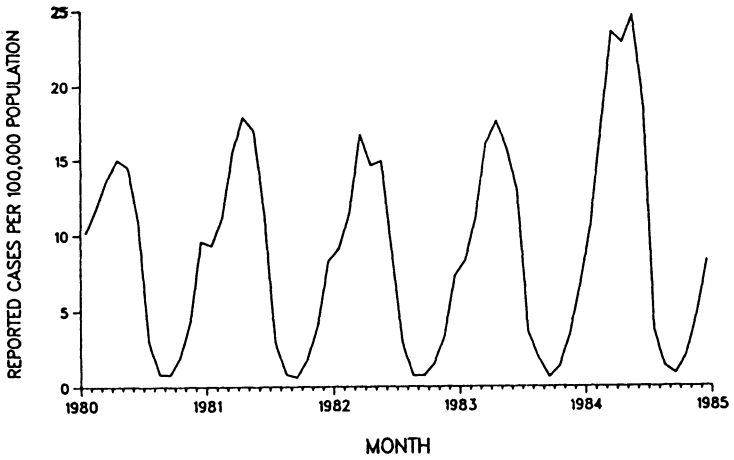

Fig. 17. Seasonal oscillation in the incidence of chickenpox (varicella) in the United States Between 1980 and 1984. Figure from CDC (1986a).

The prevalence for many diseases varies periodically because of seasonal changes in the daily contact rates. For example, the seasonal oscillation in the incidence of chickenpox is shown in Fig. 17. An SIS model with a periodic contact rate has been considered briefly in Sect. 4. Other epidemiological models with periodic contact rates are described in an article in this volume by Hethcote and Levin (1989). Other models without periodic contact rates can also have periodic solutions and are also described in the article mentioned above. These other models leading to periodic solutions have features such as a delay corresponding to temporary immunity, nonlinear incidence, variable population size or cross immunity with age structure.

The three basic epidemiological models in this article have assumed that the population being considered is uniform and homogeneously mixing; however, most infectious diseases actually spread in a diverse or dispersed population. Hence it is desirable to consider a population divided into different subpopulations. Mathematical aspects of models for heterogeneous populations are described in the survey of Hethcote, Stech and van den Driessche (1981c) and, more recently, in Hethcote and Thieme (1985) and in Hethcote and Van Ark (1987). Since gonorrhea transmission occurs in a very heterogeneous population, the models in Hethcote and Yorke (1984) for gonorrhea involve from 2 to 8 subpopulations. A spatially heterogeneous "city and villages" example is considered in May and Anderson (1984a, 1984b) and again in Hethcote and Van Ark (1987). Parameter estimation methods similar to those presented in Sections 5 and 6 are developed for heterogeneous population models in Hethcote and Van Ark (1987).

Models for populations where the disease causes enough deaths to influence the population size are considered by Anderson and May (1979) and May and Anderson (1979). Since contact rates between age groups vary greatly, it is often important to consider models with age structure. These models are considered in papers such as Kermack and McKendrick (1927), Dietz (1975), Hoppensteadt

(1975), Longini et al. (1978), Hethcote (1983), and Anderson and May (1983, 1985). Age-structured models have been used to compare the UK and USA strategies for rubella vaccination as described in the article by Hethcote (1989) in this volume. Models for measles are considered in Fine and Clarkson (1982), Hethcote (1983) and Anderson and May (1983). Epidemiological models for influenza with age structure and cross immunity are presented in the article by Castillo-Chavez et al. (1989) in this volume.

Epidemiological models with spatial spread are surveyed by Mollison (1977) and more recently by Mollison and Kuulasmaa (1985). The spatial spread of fox rabies has been considered by Anderson et al. (1981). See Radcliffe and Rass (1986) and the references cited therein for thresholds, final sizes, pandemic theorems and asymptotic speeds of propagation of travelling epidemic waves. The spread of influenza throughout the world has recently been modeled and is described in Rvachev and Longini (1985).

As indicated in Sect. 2, infectious disease models are useful in comparing control procedures. See Wickwire (1977) for a survey of models for the control of infectious diseases. The optimal uses of vaccination for influenza are considered in Longini, Ackerman and Elveback (1978). Control strategies for rubella and comparisons using cost benefit analyses are described in the article by Hethcote (1989) on rubella in this volume. Gonorrhea control procedures are compared in Hethcote and Yorke (1984).

The purpose of this article has been to introduce the most basic ideas, assumptions, notation and formulations for epidemiological models in order to prepare the reader for the study of more refined models and their applications to specific diseases. There is a great need for individuals to understand and analyse specific diseases through modeling and to use modeling to investigate and compare methods for decreasing their incidence.

References

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

Anderson, R.M., Jackson, H.C., May, R.M., Smith, A.D.M. (1981) Populations dynamics of fox rabies in Europe. Nature *289*, 765–777.

Anderson, R.M., May, R.M. (1979) Population biology of infectious diseases I. Nature *280*, 361–367.

Anderson, R.M., May, R.M. (1981) The population dynamics of microparasites and their invertebrate hosts. Phil. Trans. Roy. Soc. London *B291*, 451–524.

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

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

Anderson, R.M., May, R.M. (1985) Vaccination and herd immunity to infectious diseases. Nature *318*, 323–329.

Bailey, N.T.J. (1975) The Mathematical Theory of Infectious Diseases, 2nd edn. Hafner, New York.

Castillo-Chavez, C., Hethcote, H.W., Andreasen, V., Levin, S.A., Liu, W.M. (1988) Cross-immunity in the dynamics of homogeneous and heterogeneous populations, In: T.G. Hallam, L. Gross, and S.A. Levin (eds.) Mathematical Ecology, World Scientific Publishing, Singapore, 303–316.

Centers for Disease Control (1971a) Infectious hepatitis—Kentucky. Morbidity and Mortality Weekly Report *20*, 136–137.

Centers for Disease Control (1971b) Measles—Dallas, Texas, Morbidity and Mortality Weekly Report *20*, 191–192.

Centers for Disease Control (1981) Rubella—United States, 1978–1981. Morbidity and Mortality Weekly Report *30*, 513–515.

Centers for Disease Control (1984) Measles in an immunized school-aged population—New Mexico. Morbidity and Mortality Weekly Report *34*, 52–59.

Centers for Disease Control (1986a) Annual summary 1984: reported morbidity and mortality in the United States. Morbidity and Mortality Weekly Report *33*(54).

Centers for Disease Control (1986b) Rubella and congenital rubella syndrome—United States 1984–1985. Morbidity and Mortality Weekly Report *35*, 129–135.

Centers for Disease Control (1987a) Measles—Dade County, Florida. Morbidity and Mortality Weekly Report *36*, 45–48.

Centers for Disease Control (1987b) Enterically transmitted non-A, non-B hepatitis—East Africa, Morbidity and Mortality Weekly Report *36*, 241–244.

Coddington, E.A., Levinson, N. (1955) Theory of Ordinary Differential Equations. McGraw-Hill, New York.

Coleman, C.S. (1978) Biological cycles and the fivefold way. In: Braun, M., Coleman, C.S., Drew, D.A. (eds.) Differential Equation Models. Springer, New York, pp. 251–278.

Dietz, K. (1975) Transmission and control of arbovirus diseases. In: Ludwig D. and Cooke, K.L (eds.) Epidemiology. SIMS 1974 Utah Conference Proceedings, SIAM, Philadelphia, pp. 104–121.

Dietz, K. (1976) The incidence of infectious diseases under the influence of season fluctuations. In: Berger, J., Buhler, R., Repges, R., Tantu, P. (eds.) Mathematical Models in Medicine. Lecture Notes in Biomathematics, vol. 11. Springer, New York, pp. 1–15.

Evans, A.S. (1982) Viral Infections of Humans 2nd edn. Plenum Medical Book Company, New York.

Fenner, F. (1983) Biological control, as exemplified by smallpox eradication and myxomatosis. Proc. Roy. Soc. London *B218*, 259–285.

Fine, P.E.M., Clarkson, J.A. (1982) Measles in England and Wales I: An analysis of factors underlying seasonal patterns, and II: The impact of the measles vaccination programme on the distribution of immunity in the population. Int. J. Epid. *11*, 5–14 and 15–24.

Guckenheimer, J., Holmes, P. (1983) Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields. Springer, New York.

Hamer, W.H. (1906) Epidemic disease in England. Lancet *1*, 733–739.

Hethcote, H.W. (1973) Asymptotic behavior in a deterministic epidemic model. Bull. Math. Biology *35*, 607–614.

Hethcote, H.W. (1974) Asymptotic behavior and stability in epidemic models. In: van den Driessche P. (ed.) Mathematical Problems in Biology. Lecture Notes in Biomathematics, vol. 2. Springer, Berlin Heidelberg New York, pp. 83–92.

Hethcote, H.W. (1976) Qualitative analysis for communicable disease models. Math. Biosci. *28*, 335–356.

Hethcote, H.W. (1978) An immunization model for a heterogeneous population. Theor. Prop. Biol. *14*, 338–349.

Hethcote, H.W. (1983) Measles and rubella in the United States. Am. J. Epidemiol. *117*, 2–13.

Hethcote, H.W. (1988) Optimal ages of vaccination for measles. Math. Biosci. *89*, 29–52.

Hethcote, H.W. (1989) Rubella. In: Levin, S.A., Hallam, T.G., Gross, L. (eds.) Applied Mathematical Ecology. Biomathematics, vol. 18. Springer, Berlin, Heidelberg, New York.

Hethcote, H.W., Levin, S.A. (1988) Periodicity in epidemiological models. In: Levin, S.A., Hallam, T.G. Gross, L. (eds.) Applied Mathematical Ecology. Biomathematics, vol. 18. Springer, Berlin, Heidelberg, New York, 193–211.

Hethcote, H.W., Stech, H.W., van den Driessche, P. (1981a) Nonlinear oscillations in epidemic models. SIAM J. Appl. Math. *40*, 1–9.

Hethcote, H.W. Stech, H.W., van den Driessche, P. (1981b) Stability analysis for models of diseases without immunity. J. Math. Biology *13*, 185–198.

Hethcote, H.W., Stech, H.W., van den Driessche, P. (1981c) Periodicity and stability in epidemic models: a survey. In: Busenberg, S. and Cooke, K.L. (eds.) Differential Equations and Applications in Ecology, Epidemics and Populations Problems. Academic Press, New York, pp. 65–82.

Hethcote, H.W., Tudor, D.W. (1980) Integral equation models for endemic infectious diseases. J. Math. Biol. *9*, 37–47.

Hethcote, H.W., Van Ark, J.W. (1987) Epidemiological models for heterogeneous populations: proportionate mixing, parameter estimation and immunization programs. Math. Biosci. *84*, 85–118.

Hethcote, H.W., Yorke, J.A. (1984) Gonorrhea Transmission Dynamics and Control. Lecture Notes in Biomathematics, vol. 56, Springer, Berlin Heidelberg New York.

Hoppensteadt, F. (1975) Mathematical Theories of Populations. Demographics, Genetics and Epidemics. SIAM, Philadelphia.

Jordan, D.W., Smith, P. (1977) Nonlinear Ordinary Differential Equations. Oxford University Press, Oxford.

Kermack, W.O., McKendrick, A.G. (1927) A contribution to the mathematical theory of epidemics. Proc. Roy. Soc. *A115*, 700–721.

London, W.A., Yorke, J.A. (1973) Recurrent outbreaks of measles, chickenpox and mumps. I. Am. J. Epid. *98*, 453–468.

Longini, I.M., Jr. (1986) The generalized discrete-time epidemic model with immunity: a synthesis. Math. Biosci. *82*, 19–41.

Longini, I.M., Jr., Ackerman, E., Elveback, L.R. (1978) An optimization model for influenza A epidemics. Math. Biosci. *38*, 141–157.

May, R.N. (1986) Population biology of microparasitic infections, In: Hallam T.G. and Levin, S.A. (eds.) Mathematical Ecology. Biomathematics, vol. 17. Springer, Berlin, Heidelberg, New York, pp. 405–442.

May, R.M., Anderson, R.M. (1979) Population biology of infectious diseases II. Nature *280*, 455–461.

May, R.M., Anderson, R.M. (1984a) Spatial heterogeneity and the design of immunization programs. Math. Biosci. *72*, 83–111.

May, R.M., Anderson, R.M. (1984b) Spatial, temporal, and genetic heterogeneity in host populations and the design of immunization programmes. IMA J. of Math. App. Med. Biol. *1*, 233–266.

Miller, R.K., Michel, A.N. (1982) Ordinary Differential Equations. Academic Press, New York.

Mollison, D. (1977) Spatial contact models for ecological and epidemic spread. J.R. Statist. Soc. Ser. *B39*, 283–326.

Mollison, D., Kuulasmaa, K. (1985) Spatial epidemic models: theory and simulations. In: Bacon, P.J. (ed.) Population Dynamics of Rabies in Wildlife. Academic Press, London, pp. 291–309.

Radcliffe, J., Rass, L. (1986) The asymptotic speed of propagation of the deterministic nonreducible n-type epidemic. J. Math. Biol. *23*, 341–359.

Ross, R. (1911). The Prevention of Malaria, 2nd edn. Murray, London.

Rvachev, L.A., Longini, I.M. Jr. (1985) A mathematical model for the global spread of influenza. Math. Biosci. *75*, 3–22.

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

Soper, H.E. (1929) Interpretation of periodicity in disease prevalence. J.R. Statist. Soc. *92*, 34—73.

World Health Organization (1980) The Global Eradication of Smallpox. Final report, WHO, Geneva.

Yorke, J.A., London, W.P. (1973) Recurrent outbreaks of measles, chickenpox and mumps II. Am. J. Epid. *98*, 469–482.

Yorke, J.A., Nathanson, N. Pianigiani, G. Martin, J. (1979) Seasonality and the requirements for prepetuation and eradication of viruses in populations. Am. J. Epidemiol. *109*, 103–123.

Appendix

Proof of Theorem 5.1. The triangle T given by (5.3) is positively invariant since no direction vectors at the boundary of T are outward. More precisely, $S = 0$ implies $S' = 0$, $I = 0$ implies $I' = 0$, and $S + I = 1$ implies $(S + 1)' = -\gamma I \leqq 0$. Moreover, every point on the S axis where $I = 0$ is an equilibrium point (EP). The EP for $S < 1/\sigma$ are neutrally stable and the EP for $S > 1/\sigma$ are neutrally unstable.

From (5.2) and the positive invariance of T, it follows that $S(t)$ is nondecreasing and $S(t) \geqq 0$ so that a unique limit $S(\infty)$ exists. Since $R'(t) = \gamma I \geqq 0$ and $R(t)$ is bounded above by 1, the limit $R(\infty)$ exists. Since $I(t) = 1 - S(t) - R(t)$, the limit $I(\infty)$ exists. Moreover, $I(\infty) = 0$ since otherwise $R'(t) > \gamma I(\infty)/2$ for t sufficiently large so that $R(\infty) = \infty$ which contradicts $R(\infty) \leqq 1$. The solution paths

$$I = 1 - R_0 - S + [\ln(S(t)/S_0)]/\sigma \tag{A.1}$$

are found from $dI/dS = -1 + 1/\sigma S$. The conclusions stated follow directly from (A.1) and the observation that a solution path has a maximum infective fraction Im when $\sigma S = 1$. □

Proof of Theorem 6.1. This proof uses standard phase plane methods found in differential equation books such as Coddington and Levinson (1955), Jordan and Smith (1977) and Miller and Michel (1982). This proof was first presented in Hethcote (1976). The equilibrium points (EP) in the SI phase plane are $(1, 0)$ and the EP given by (6.3). The characteristic roots of the linearization around the EP $(1, 0)$ are $-\mu$ and $(\gamma + \mu)(\sigma - 1)$ so that this EP is a stable node if $\sigma < 1$ and a saddle if $\sigma > 1$. The triangle defined by (5.4) is positively invariant since no path leaves through a boundary. More precisely, $S = 0$ implies $S'(t) = \mu > 0$, $I = 0$ implies $I'(t) = 0$, and $S + I = 1$ implies $(S + I)' = -\gamma I \leqq 0$. Moreover, there is a path along the S axis approaching the EP $(1, 0)$.

The Poincarè-Bendixson theorem (Coddington and Levinson, 1955; Miller and Michel, 1982) implies that bounded paths in the phase plane approach either an EP, a limit cycle or a cycle graph (Coleman, 1978). If $\sigma \leqq 1$, then $(1, 0)$ is the only EP in the triangle T. There is no limit cycle contained in T since limit cycles must contain at least one EP in their interior. There is no cycle graph in T since a homoclinic loop is not possible from a stable EP (Guckenheimer and Holmes, 1983). Thus all paths in T approach the EP $(1, 0)$ if $\sigma \leqq 1$.

If $\sigma > 1$, then the EP $(1, 0)$ is a saddle with an attractive path along the S axis and a repulsive path into T with slope $-1 + \gamma/(\gamma + \mu)$. If $\sigma > 1$, then the EP (6.3) is in the interior of T and it is locally asymptotically stable since the characteristic roots of the linearization around it have negative real parts. The Bendixson–Dulac test (Jordan and Smith, 1977, p. 91; Hethcote, 1976) with multiplying factor $1/I$ leads to

$$\frac{\partial}{\partial S}\left(-\lambda S + \frac{\mu}{I} - \frac{\mu S}{I}\right) + \frac{\partial}{\partial I}(\lambda S - \gamma - \mu) = -\lambda - \frac{\mu}{I} < 0$$

so that there are no limit cycles or cycle graphs in T. The only path in T approaching the EP $(1, 0)$ is the S axis. Thus all paths in T except the S axis approach the EP given by (6.3).

This stability result for $\sigma > 1$ can also be proved by using a Liapunov function. The Liapunov function used here also works for the SIRS model in Hethcote (1976), but the Liapunov function given in Hethcote (1974) does not. If $S = S_e(1 + U)$ and $I = I_e(1 + V)$ where (S_e, I_e) is the EP (6.3), then

$$U'(t) = -\lambda I_e U(1 + V) - \lambda I_e V - \mu U$$
$$V'(t) = (\gamma + \mu) U(1 + V)$$

and the positively invariant triangle T given by (5.3) becomes

$$T^* = \{(U, V): U \geqq -1, V \geqq -1, S_e U + I_e V \leqq 1 - S_e - I_e\}.$$

The Liapunov function

$$L = U^2/2 + \sigma I_e[V - \ln(1 + V)]$$

is positive definite for $V > -1$ and the Liapunov derivative is

$$L' = -\lambda I_e U^2(1 + V) - \mu U^2 \leqq 0.$$

The set $E = \{(U, V): L' = 0\}$ is the V axis where $U = 0$. Since $U = 0$ implies $U' = -\lambda I_e V$, the only positively invariant subset of E is the origin. By the Liapunov–Lasalle theorem (Miller and Michel, 1982, p. 226), the EP$(0, 0)$ in UV coordinates is locally asymptotically stable. Since the Liapunov curves $L(U, V) = C$ fill the upper half plane above $V = -1$ as C approaches infinity, this half plane is an asymptotic stability region for the EP$(0, 0)$. This implies the result stated for $\sigma > 1$ in Theorem 6.1. □