The Threshold Theorem of Epidemiology

Consider the situation where a small group of people having an infectious disease is inserted into a large population which is capable of catching the disease.

- What happens as time evolves?

- Will the disease die out rapidly, or will an epidemic occur?

- How many people will ultimately catch the disease?

To answer these questions we will derive a system of differential equations which govern the spread of infectious disease within a population, and analyze the behavior of its solutions.

This approach will also lead to the famous Threshold Theorem of epidemiology which states that an epidemic will occur only if the number of people who are susceptible to the disease exceeds a certain threshold value.

We begin with the assumptions:

- the disease under consideration confers permanent immunity upon any individual who has completely recovered from it,

- the disease has a negligibly short incubation period.

This latter assumption implies that the individual who contracts the disease become infective immediately afterwards.

In this case we can divide the population into three classes of individuals:

- the susceptible class (S)
- the infective class (I)
- the removed class (R)

- The susceptible class consists of those individuals who are not infective, but who are capable of catching the disease and becoming infective.

- The infective class consists of those individuals who are capable of transmitting the disease to others.

- The removed class consists of those **individuals** who have had the disease and are dead, or have recovered and are permanently immune, or are isolated until recovery and permanent immunity occur.

The spread of the disease is presumed to be governed by the following rules.

Rule 1: The population remains at a fixed level N in the time interval under consideration. This means, of course, that we neglect births, deaths from causes unrelated to the disease under consideration, immigration and emigration.

Rule 2: The rate of change of the susceptible population is proportional to the product of the number of members of (S) and the number of members of (I).

Rule 3: Individuals are removed from the infectious class (I) at a rate proportional to the size of (I).

Let $S(t)$, $I(t)$, and $R(t)$ denote the number of individuals in classes (S), (I), and (R) respectively, at time t. It follows immediately from Rules 1-3 that $S(t)$, $I(t)$, $R(t)$ satisfies the system of differential equations

$$\begin{cases} \dfrac{dS}{dt} = -rSI \\[2mm] \dfrac{dI}{dt} = rSI - \gamma I \\[2mm] \dfrac{dR}{dt} = \gamma I \end{cases} \qquad (1)$$

for some positive constants r and γ.

The proportionality constant r is called the infection rate, and the proportionality constant γ is called the removal rate.

The first two equations of (1) do not depend on R. Thus, we need only consider the system of equations:

$$\begin{cases} \dfrac{dS}{dt} = -rSI \\[3mm] \dfrac{dI}{dt} = rSI - \gamma I \end{cases} \qquad (2)$$

for the two unknown functions $S(t)$ and $I(t)$. Once $S(t)$ and $I(t)$ are known, we can solve for $R(t)$ from the third equation of (1).

Alternately, observe that $\dfrac{d}{dt}(S + I + R) = 0$.

Thus, $\qquad S(t) + I(t) + R(t) = \text{constant} = N$

so that $\quad R(t) = N - S(t) - I(t)$.

The trajectories of (2) are the solution curves of the 1st-order equation

$$\frac{dI}{dS} = \frac{rSI - \gamma I}{-rSI} = -1 + \frac{\gamma}{rS}. \qquad (3)$$

Integrating this differential equation gives

$$\left[\begin{array}{l} I(S) = -S + \dfrac{\gamma}{r} \ln S + C, \quad I(S_0) = I_0 \; (I.C.) \Longrightarrow. \\[3mm] I_0 = -S_0 + \dfrac{\gamma}{r} \ln S_0 + C \Rightarrow \quad C = I_0 + S_0 - \dfrac{\gamma}{r} \ln S_0 \Rightarrow. \end{array} \right]$$

$$\boxed{I(S) = I_0 + S_0 - S + \rho \ln \frac{S}{S_0}} \qquad (4)$$

where S_0 and I_0 are the number of susceptibles and infectives at the initial time $t = t_0$, and $\rho = \dfrac{\gamma}{r}$.

To analyze the behavior of the curves (4), we compute

$$I'(S) = -1 + \frac{\rho}{S}$$

$$\left[\begin{array}{l} I'(S) = 0: \quad S = \rho \\ I''(S) = -\frac{\rho}{S^2} \quad I''(\rho) = -\frac{\rho}{\rho^2} = -\frac{1}{\rho} < 0 \text{ rel. max} \end{array}\right]$$

The quantity $-1 + \frac{\rho}{S}$ is negative for $S > \rho$, and positive for $S < \rho$. Hence, $I(S)$ is an increasing function of S for $S < \rho$, and a decreasing function of S for $S > \rho$.

Next, we observe that

$$I(0) = -\infty \quad \text{and} \quad I(S_0) = I_0 > 0.$$

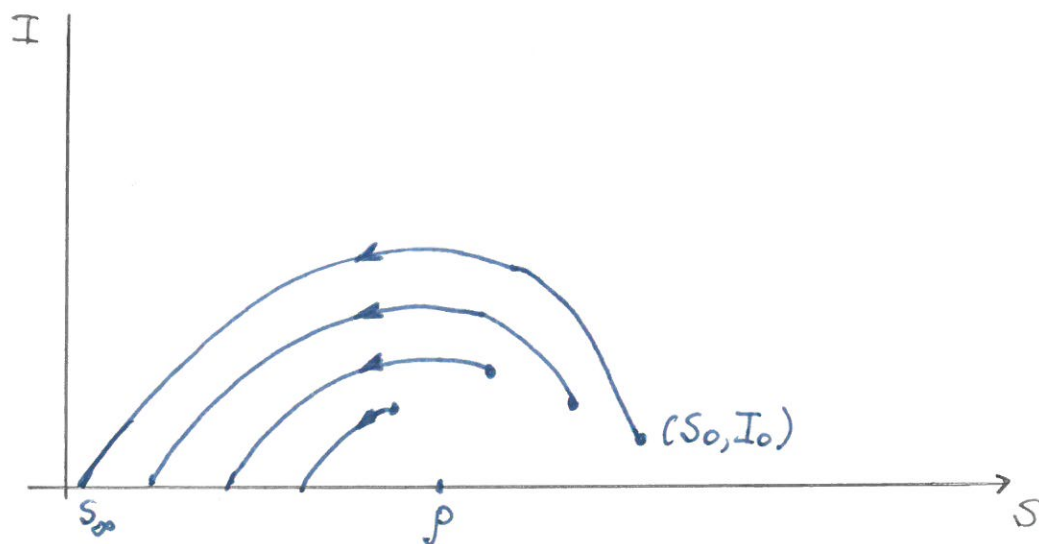Consequently, there exists a unique point S_∞, with $0 < S_\infty < S_0$, such that $I(S_\infty) = 0$, and $I(S) > 0$ for $S_\infty < S \leq S_0$. The point $(S_\infty, 0)$ is an equilibrium point of (2) since

$$\frac{dS}{dt}\Big|_{I=0} = 0$$

$$\frac{dI}{dt}\Big|_{I=0} = 0$$

Thus, the trajectories of (2), for $t_0 \leq t < \infty$, have the following form:

What all this implies about the spread of the disease within the population.

- As t runs from t_0 to ∞, the point $(S(t), I(t))$ travels along the curve (4), and it moves along the curve in the direction of decreasing S, since $S(t)$ decreases monotonically with time. Consequently, if S_0 is less than ρ, then $I(t)$ decreases monotonically to zero, and $S(t)$ decreases monotonically to S_∞. Thus, if a small group of infectives I_0 is inserted into a group of susceptibles S_0, with $S_0 < \rho$, then the disease will die out rapidly.

- On the other hand, if S_0 is greater than ρ, then $I(t)$ increases as $S(t)$ decreases to ρ, and it achieves a maximum value when $S = \rho$. It only starts decreasing when the number of susceptibles falls below the threshold value ρ.

From these results we may draw the following conclusions.

Conclusion 1: An epidemic will occur only if the number of susceptibles in a population exceeds the threshold value $\rho = \gamma / r$.

Conclusion 2: The spread of the disease does not stop for lack of a susceptible population; it stops only for lack of infectives. In particular, some individuals will escape the disease altogether.

Conclusion 1 corresponds to the general observation that epidemics tend to build up more rapidly when the density of susceptibles is high due to overcrowding, and the removal rate is low because of ignorance, inadequate isolation and inadequate medical care. On the other hand, outbreaks tend to be of only limited extent when good social conditions entails lower densities of susceptibles, and when removal rates are high because of good public health vigilance and control.

If the number of susceptibles S_0 is initially greater than, but close to, the threshold value ρ, then we can estimate the number of individuals who ultimately contract the disease. Specifically, if $S_0 - \rho$ is small compared to ρ, then the number of individuals who ultimately contract the disease is approximately $2(S_0 - \rho)$. This is the famous Theshold Theorem of epidemiology, which was first proven in 1927 by the mathematical biologists Kermack and McKendrick.

Theorem 1 (Threshold Theorem of Epidemiology).

Let $S_0 = \rho + \gamma$ and assume that γ/ρ is very small compared to one. Assume moreover, that the number of initial infectives I_0 is very small. Then, the number of individuals who ultimately contract the disease is 2γ. In other words, the level of susceptibles is reduced to a point as far below the threshold as it originally was above it.

Proof Letting t approach infinity in (4) gives

$$0 = I_0 + S_0 - S_\infty + \rho \ln\left(\frac{S_\infty}{S_0}\right).$$

If I_0 is very small compared to S_0, then we can neglect it, and write

$$0 = S_0 - S_\infty + \rho \ln\left(\frac{S_\infty}{S_0}\right)$$

$$= S_0 - S_\infty + \rho \ln\left[\frac{S - (S_0 - S_\infty)}{S_0}\right]$$

$$= S_0 - S_\infty + \rho \ln\left[1 - \left(\frac{S_0 - S_\infty}{S_0}\right)\right].$$

Now, if $S_0 - \rho$ is small compared to ρ, then $S_0 - S_\infty$ will be small compared to S_0. Consequently, we can truncate the Taylor series

$$\ln\left[1 - \left(\frac{S_0 - S_\infty}{S_0}\right)\right] = -\left(\frac{S_0 - S_\infty}{S_0}\right) - \frac{1}{2}\left(\frac{S_0 - S_\infty}{S_0}\right)^2 + \dots$$

after two terms.

Then,

$$0 = S_0 - S_\infty - \rho\left(\frac{S_0 - S_\infty}{S_0}\right) - \frac{\rho}{2}\left(\frac{S_0 - S_\infty}{S_0}\right)^2$$

$$= (S_0 - S_\infty)\left[1 - \frac{\rho}{S_0} - \frac{\rho}{2S_0^2}(S_0 - S_\infty)\right].$$

Solving for $S_0 - S_\infty$, we get

$$S_0 - S_\infty = \frac{1 - \frac{\rho}{S_0}}{\rho/(2S_0^2)} = \frac{S_0 - \rho}{S_0} \cdot \frac{2S_0^2}{\rho} = \frac{2S_0(S_0 - \rho)}{\rho} = \frac{2S_0\rho\left(\frac{S_0}{\rho} - 1\right)}{\rho}$$

$$= 2S_0\left(\frac{S_0}{\rho} - 1\right) = 2(\rho + \nu)\left[\frac{\rho + \nu}{\rho} - 1\right] = 2(\rho + \nu)\frac{\nu}{\rho}$$

$$= 2\rho\left(1 + \frac{\nu}{\rho}\right)\frac{\nu}{\rho} = 2\left(1 + \frac{\nu}{\rho}\right)\nu \approx 2\nu.$$

During the course of an epidemic it is impossible to accurately ascertain the number of new infectives each day or week, since the only infectives who can be recognized and removed from circulation are those who seek medical aid. Public health statistics thus record only the number of new removals each day or week, not the number of new infectives. Therefore, in order to compare the results predicted by the model with data from actual epidemics, we must find the quantity dR/dt as a function of time. This is accomplished in the following manner. We observe first that

$$\frac{dR}{dt} = \gamma I = \gamma(N - R - S).$$

Second, we observe that

$$\frac{dS}{dR} = \frac{dS/dt}{dR/dt} = -\frac{rSI}{\gamma I} = -\frac{S}{\rho}.$$

Hence, $\quad S(R) = S_0 e^{-\frac{R}{\rho}}\quad$ and

$$\frac{dR}{dt} = \gamma\left(N - R - S_0 e^{-R/\rho}\right). \tag{5}$$

Equation (5) is separable, but cannot be solve explicitly. However, if the epidemic is not very large, then R/ρ is small and we can truncate the Taylor series

$$e^{-R/\rho} = 1 - \frac{R}{\rho} + \frac{1}{2}\left(\frac{R}{\rho}\right)^2 + \cdots$$

after three terms. With this approximation,

$$\frac{dR}{dt} = \gamma\left[N - R - S_0\left[1 - \frac{R}{\rho} + \frac{1}{2}\left(\frac{R}{\rho}\right)^2\right]\right]$$

$$= \gamma\left[N - S_0 + \left(\frac{S_0}{\rho} - 1\right)R - \frac{S_0}{2}\left(\frac{R}{\rho}\right)^2\right]$$

The solution of this equation is

$$R(t) = \frac{\rho^2}{S_0}\left[\frac{S_0}{\rho} - 1 + \alpha \tanh\left(\frac{1}{2}\alpha\gamma t - \varphi\right)\right] \qquad (6)$$
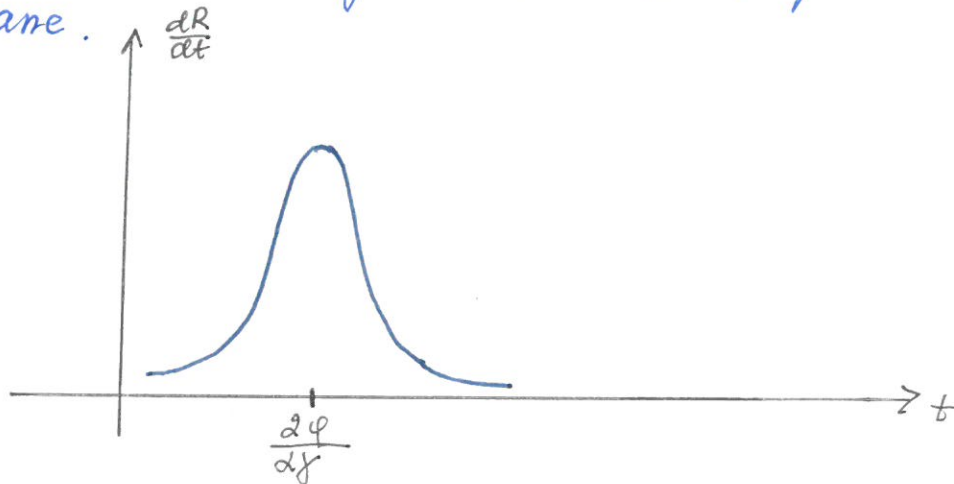
where

$$\alpha = \left[\left(\frac{S_0}{\rho} - 1\right)^2 + \frac{2S_0(N - S_0)}{\rho^2}\right]^{1/2}, \qquad \varphi = \tanh^{-1}\frac{1}{\alpha}\left(\frac{S_0}{\rho} - 1\right)$$

and the hyperbolic tangent function $\tanh z$ is defined by

$$\tanh z = \frac{e^z - e^{-z}}{e^z + e^{-z}}, \qquad \frac{d}{dz}\tanh z = \text{sech}^2 z = \frac{4}{(e^z + e^{-z})^2}$$

Hence,

$$\frac{dR}{dt} = \frac{\gamma\alpha^2\rho^2}{2S_0}\text{sech}^2\left(\frac{1}{2}\alpha\gamma t - \varphi\right) \qquad (7)$$

Equation (7) defines a symmetric bell shaped curve in the $(t, \frac{dR}{dt})$-plane.

Kermack and McKendrick compared the values predicted for dR/dt from (7) with data from the actual plague in Bombay which spanned the last half of 1905 and the first half of 1906. They set

$$\frac{dR}{dt} = 890 \operatorname{sech}^2(0.2t - 3.4)$$

with t measured in weeks, and compared these values with the number of deaths per week from the plague. This quantity is a very good approximation of dR/dt, since almost all cases terminated fatally.

As can be seen from the figure below, there is excellent agreement between the actual values of dR/dt, denoted by •, and the values predicted by (7). This indicates, of course, that the system of differential equations (1) is an accurate and reliable model of the spread of an infectious disease within a population of fixed size.
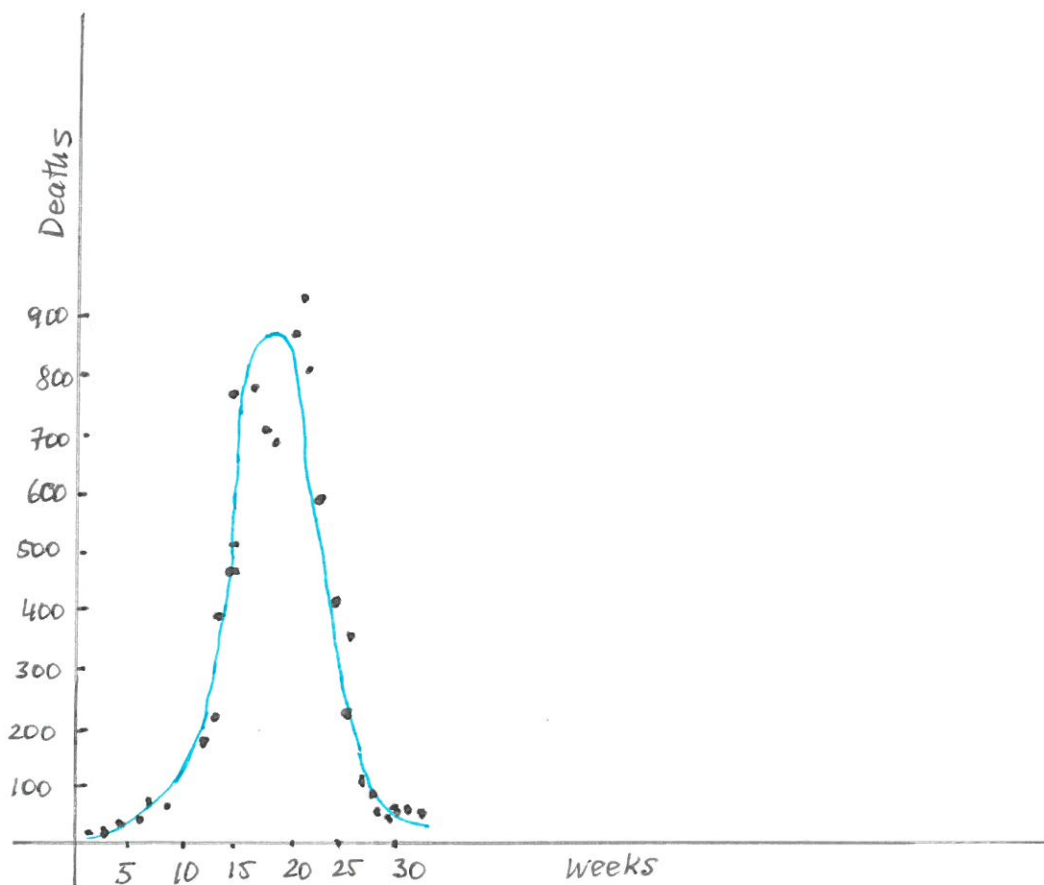
Example 1 It was observed in London during the early 1900's, that large outbreaks of measles epidemics recurred about once every two years. The mathematical biologist M. E. Soper tried to explain this phenomenon by assuming that the stock of susceptibles in constantly replenished by new recruits to the population. Thus, he assumed that

$$
\begin{cases}
\dfrac{dS}{dt} = -rSI + \mu \\[4mm]
\dfrac{dI}{dt} = rSI - \gamma I
\end{cases}
$$

(8)

for some positive constants r, γ, and μ.

(a) Show that $S = \dfrac{\gamma}{r}$, $I = \dfrac{\mu}{\gamma}$ is the only equilibrium solution of (8).

(b) Show that every solution $S(t)$, $I(t)$ of (8) which starts sufficiently close to this equilibrium point must ultimately approach it as t approaches infinity.

(c) It can be shown that every solution $S(t)$, $I(t)$ of (8) approaches the equilibrium solution $S = \gamma/r$, $I = \mu/\gamma$ as t approaches infinity. Conclude, therefore, that the system (8) does not predict recurrent outbreakes of measles epidemics. Rather, it predicts that the disease will ultimately approach a steady state.

Solution (a) $\dfrac{dS}{dt} = 0 = -rSI + \mu$ $\left.\begin{matrix}\\[6mm]\\ \end{matrix}\right\}\ \overset{+}{\Rightarrow}$ $\mu - \gamma I = 0 \Rightarrow I = \dfrac{\mu}{\gamma}$

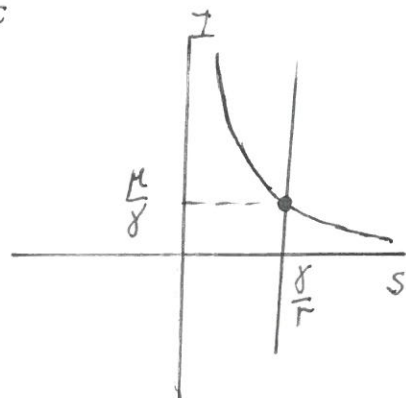$\dfrac{dI}{dt} = 0 = rSI - \gamma I$ $S = \dfrac{\gamma I}{rI} = \dfrac{\gamma}{r}$

The equilibrium point is $S = \dfrac{\gamma}{r}$, $I = \dfrac{\mu}{\gamma}$.

$rSI = \mu$ or $I = \dfrac{\mu}{rs}$

$rSI - \gamma I = 0$

$S = \dfrac{\gamma}{r}$

(6) To answer the question we investigate the stability of the equilibrium solution of the linearized system.

We set $\quad u = S - \frac{\gamma}{r}$

$$v = I - \frac{\mu}{\gamma}$$

$$\frac{du}{dt} = \frac{dS}{dt} = -r\left(u + \frac{\gamma}{r}\right)\left(v + \frac{\mu}{\gamma}\right) + \mu = -r\left(uv + \frac{\mu}{\gamma}u + \frac{\gamma}{r}v + \frac{\mu}{r}\right) + \mu$$

$$= -ruv - \frac{r\mu}{\gamma}u - \gamma v - \cancel{\mu} + \cancel{\mu} = -\frac{r\mu}{\gamma}u - \gamma v - ruv$$

$$\frac{dv}{dt} = \frac{dI}{dt} = r\left(u + \frac{\gamma}{r}\right)\left(v + \frac{\mu}{\gamma}\right) - \gamma\left(v + \frac{\mu}{\gamma}\right) = r\left(uv + \frac{\mu}{\gamma}u + \frac{\gamma v}{r} + \frac{\mu}{r}\right) - \gamma v - \mu$$

$$= ruv + \frac{\mu r}{\gamma}u + \cancel{\gamma v} + \cancel{\mu} - \cancel{\gamma v} \cancel{- \mu} = \frac{\mu r}{\gamma}u + ruv$$

$$\frac{d}{dt}\begin{pmatrix} u \\ v \end{pmatrix} = \underbrace{\begin{pmatrix} -\frac{r\mu}{\gamma} & -\gamma \\ \frac{\mu r}{\gamma} & 0 \end{pmatrix}}_{A}\begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} -ruv \\ ruv \end{pmatrix}$$

$$p(\lambda) = \det\begin{pmatrix} -\frac{\mu r}{\gamma} - \lambda & -\gamma \\ \frac{\mu r}{\gamma} & -\lambda \end{pmatrix} = \left(\frac{\mu r}{\gamma} + \lambda\right)\lambda + \mu r = \lambda^2 + \frac{\mu r}{\gamma}\lambda + \mu r$$

Eigenvalues: $\quad \lambda^2 + \frac{\mu r}{\gamma}\lambda + \mu r = 0.$

$$D = \left(\frac{\mu r}{\gamma}\right)^2 - 4\mu r$$

$$\lambda_1 = -\frac{\mu r}{2\gamma} - \frac{1}{2}\sqrt{\left(\frac{\mu r}{\gamma}\right)^2 - 4\mu r}$$

$$\lambda_2 = -\frac{\mu r}{2\gamma} + \frac{1}{2}\sqrt{\left(\frac{\mu r}{\gamma}\right)^2 - 4\mu r}$$

There are three possibilities: i) $D > 0$, then $\lambda_1 < 0$ & $\lambda_1 < 0$

ii) $D < 0$, then $Re(\lambda_1) < 0$, $Re(\lambda_2) < 0$

iii) $D = 0$, then $\lambda_1 = \lambda_2 = -\frac{\mu r}{2\gamma} < 0.$

Every case provides that the equilibrium solution is asymptotically stable. It implies that every solution $S(t)$, $I(t)$ of (8) which starts sufficiently close to this equilibrium solution will approach it as $t \to \infty$.

(c) By Part (b) we have two possibilites, either the equilibrium point an asymptotically stable node or an asymptotically stable focus. These two possibilities do not give periodic solutions, which would predict recurrent outbreaks of measles epidemics, but instead it support that solutions approach a steady state (the equilibrium solution), especially in the case of the asymptotically stable focus.