Temporal Analysis of Epidemics I: Description and Comparison of Disease Progress Curves

It was soon realized that if assessments were expressed as a percentage of the foliage area destroyed, for whole plots or fields, and successive assessments were plotted on a time base, most valuable curves could be obtained depicting the progress of blight on typical crops in different years and different localities.

E. C. Large, 1952

8.1 INTRODUCTION

If the amount of disease present in a population of plants is assessed at several times, the results can be presented collectively as a disease progress curve. This disease progress curve, which is the graph of disease intensity versus some measure of time, is the “signature” of the epidemic. It represents an integration of all host, pathogen, and environmental effects occurring during the epidemic and provides an opportunity to analyze, compare, and understand plant disease epidemics.

Disease progress curves can be constructed for diseases caused by any pathogen in any population of host plants. The host plants can be annuals or perennials growing in a temperate, subtropical, or tropical zone, and the diseases can occur primarily on roots, stems, leaves, fruit, or any combination of these plant parts. The time scale can be relatively short (a few weeks) or long (several years) and can be measured in calendar days or on a physiological basis. The area in which the disease occurs can vary in scale from a small plot to a continent. The pathosystem can involve wild or cultivated plant species and can occur in a natural or managed habitat.

From disease progress curves various characteristics of the epidemic can be discerned. Curves for some viral diseases, a bacterial disease, and three foliar and three root diseases caused by fungi are illustrated in Figure 8.1. It is immediately apparent that there is a great deal of variation in the curves and that the relative length of epidemic duration (days, weeks, or years) varies among the pathosystems. For each epidemic, the time of disease
onset, the initial amount of disease ($y_0$), the rate of disease increase ($r_y$), the area under the disease progress curve (AUDPC), the shape of the curve, the maximum ($y_{max}$) and final ($y_f$) amounts of disease, and the overall duration of the epidemic can be determined by inspection, or interpolation, or extrapolation, or can be estimated statistically.

For epidemics that are not curtailed by the harvest of an annual crop or the loss of leaves in a perennial crop, the duration of the **progressive and degressive** phases of the epidemic, the rate of disease increase or decrease during each phase, and all of the characteristics identified above can be described, quantified, and compared. Epidemics of this type are most likely to occur in tropical or semitropical climatic zones on perennial crops. The disease progress curve for such a disease, black sigatoka on banana (caused by *Mycosphaerella fijiensis var. diffusa*) in Mexico is presented in Figure 8.2. In this example the progressive and degressive phases of the epidemic correspond primarily to seasonal changes in weather conditions.

The purpose of the temporal analysis of an epidemic, or collection of epidemics, determines the needed precision and complexity of the analysis. In general, three somewhat distinct levels of analysis are possible. One level seeks gross comparisons among experimental treatments such as cultivars or fungicide spray schedules in order to evaluate strategies for disease management. Little effort is spent on modeling disease progress or interpreting disease progress curves. This level is perhaps the most applied type of analysis and is used extensively (Fry, 1982; Berger, 1977, 1988). At a second, more complex level, changes in specific environmental factors, pathogen or vector biotypes, or host resistance within a pathosystem induce changes in the epidemic that are reflected by alterations in the disease progress curve. Models of disease progress are used and variations in curves are interpreted. The third level of analysis corresponds to comparative epidemiology (Kranz 1974b, 1978, 1988b; Palti and Kranz, 1980). The purpose is to identify similarities and differences among epidemics based on the nature or behavior of disease progress curves and to test theories concerning the elements that serve as primary determinants of factors such as the rate of increase during epidemics of plant diseases.

Once the purpose of the temporal analysis of the epidemic has been es-

---

Figure 8.1. Disease progress curves for nine pathosystems. A. Percent infection of cucumber plants by cucumber mosaic virus at five times after inoculation (data from Leebenstein et al., 1966). B. Percent infection of peach trees by prune dwarf and prunus necrotic ring spot viruses over several years (data from Smith et al., 1977). C. Percentage of trees diseased at various times after inoculation of a central fugal path with *Xanthomonas campestris pv. citri*, causal agent of Asian citrus canker, in a Duncan grapefruit nursery in Argentina (data from Gottwald et al., 1990). D. Disease severity of *Cercospora* blight caused by *Cercospora apii* on celery in Florida (data from Berger, 1973). E. Percent of maximum number of lesions per 0.3 m of row for early leaf spot of peanut (caused by *Cercospora arachidicola*) in North Carolina (data from Alderman et al., 1987a). F. Percent severity of stripe rust of wheat (caused by *Puccinia striiformis*) versus days since inoculation in region (data from Shrum, 1975). G. Disease incidence (percent dead plants) versus weeks after disease initiation for *Phymatotrichum* root rot (caused by *Phymatotrichum omnivorum*) on cotton in Texas (data from Rush and Lyda, 1984). H. Disease incidence (percent symptomatic plants) of black shank of tobacco (caused by *Phytophthora parasitica var. nicotianae*) in North Carolina (data from Campbell et al., 1984). I. Disease incidence (percent plants with foliar symptoms) of *Verticillium* wilt on cotton (caused by *Verticillium dahliae*) in California (data from Ashworth et al., 1979).

---

Figure 8.2. Increase and decrease in severity of black sigatoka disease of banana (caused by *Mycosphaerella fijiensis var. diffusa*) in Tabasco, Mexico during 1985–1987 (courtesy of G. Ramirez). Most epidemic models do not account for an increase and decrease in disease.
established, there are several general approaches or methodologies that can be used. The approach most appropriate for a given temporal analysis depends on the specific purpose of the analysis (see preceding paragraph) and the quantity and quality of the disease progress data available. The approaches to the analysis of disease progress data are not mutually exclusive, so that more than one approach may be appropriate and useful for a given study. In one approach, disease epidemics are described using one or several descriptive parameters, such as the rate of disease progress. For this approach the most common type of models used are growth curve models, which have been borrowed from other scientific disciplines such as ecology or animal science (Madden, 1980 (see Section 8.3)). Vanderplank (1963) used some of these models but gave little credit to the work that had been done in other disciplines. The growth curve models can be utilized as biological (mechanistic) models or as empirical (or statistical) models. A second approach is the “synoptic description” of disease development using a quantifier such as area under the disease progress curve ( Shaner and Finney, 1977; Fry, 1978). In this approach only disease level and duration are considered without making any of the assumptions associated with growth curve models that may be appropriate for describing the epidemic (see Section 8.4). A third approach is one in which no one parameter or variable value is used to characterize the entire epidemic, but disease level at each time of observation is compared through a repeated-measures type of statistical analysis (see Section 8.5.2).

Our presentation on the temporal analysis of epidemics is divided into two parts. In this chapter we discuss the proposals made by Vanderplank (1963) for epidemic analysis and the basics of temporal analysis using each of the three approaches outlined. Certain simplifying assumptions are made implicitly and the emphasis is on analysis and comparison of epidemics that conform (or nearly conform) to those assumptions. In Chapter 9, more advanced and problematical topics are considered. Additionally, the implicit assumptions inherent in the analytical and comparative methods described in Chapter 8 are made explicit and reexamined, and methods for examining disease progress data are expanded.

8.2 THE PROPOSALS OF VANDERPLANK

The way in which phytopathologists considered disease epidemics and the analysis of these epidemics changed significantly in 1963 with the publication of Vanderplank’s Plant Diseases: Epidemics and Control. This treatise was a follow-up to a chapter by Vanderplank (1960) in which he originally introduced several important concepts regarding epidemics. Earlier, workers such as Fracker (1936), Barratt (1945), Chester (1946), Large (1952), and Schmitt et al. (1959) had commented on the “seasonal advance” or “tempo” of disease development. They had also presented illustrations of disease increase over time, suggested the importance of knowing how fast epidemics progressed, and provided initial techniques for the analysis of disease progress data. These workers, along with Gärömm through his excellent treatise (1946), had provided the background for the Vanderplankian revolution in plant disease epidemiology.

Vanderplankian analysis is based “largely on infection rates and on the relation between the amount of inoculum and the amount of disease it produces” (Vanderplank, 1963). Two models, the logistic and monomolecular, formed the basis of the system, although some references were made to the exponential (= logarithmic) model (see Section 8.3.1). The models were applied as biological models; that is, specific biological assumptions were made about the pathosystems to which the models were applied. What emerged was a simple yet elegant system for analyzing and comparing epidemics.

For purposes of illustration, disease increase was considered analogous to the accumulation of interest on money. Two types of interest were important, simple and compound (Section 8.7). With simple interest, interest accrues based only on the amount of principal present and no interest is gained on the interest. By analogy, certain diseases increase only from a reservoir of inoculum with only one cycle of effective inoculum production per growing season; that is, new infections do not lead to further infections. These monocyclic diseases could thus be called “simple interest” diseases. With compound interest, interest accrues based on the amount of the principal plus the accumulated interest. In other words, one gets interest on the interest as well as on the principal. By analogy, certain diseases have many cycles of infection and inoculum production during the course of a single season’s epidemic and inoculum (and thus disease) increase is proportional to the initial amount present plus the amount produced subsequently. These polycyclic diseases could thus be called “compound interest” diseases.

The analogy between interest accumulated on money and disease increase is presented merely for illustration purposes and as an introduction to a way of examining the pathogen component of plant disease epidemics. The analogy between money and plant disease breaks down in at least four ways (Vanderplank, 1963):

1. Infections occur intermittently, not continuously.
2. There is a limit to the amount of host tissue available and thus to the amount of disease.
3. Newly infected host tissue does not immediately become infective.
4. Diseases often occur in foci.

Thus the money analogy serves as an intuitive, although not entirely realistic, starting point and helps establish a basis for the use of specific models to
describe disease progress. Items 1, 3, and 4 are considered in Chapter 9, and item 2 is considered in this chapter.

The distinction between polycyclic and monocyclic diseases is of paramount importance in Vanderplankian analysis. With polycyclic (compound interest) diseases, the pathogen "may multiply through successive generations in the course of the epidemic." An example presented was stem rust of wheat (caused by *Puccinia graminis* var. *tritici*) where the epidemic begins with only a few pustules that eventually erupt and release ureldiospores, which in turn form more pustules. For this type of disease increase, where the disease spreads from lesion to lesion or from plant to plant if the disease is systemic, Vanderplank proposed the linearized form of a logistic model such that one would plot \( \ln[y/(1-y)] \) versus time, where \( y \) is the proportion of disease (incidence or severity). This represents a modified logarithmic increase of disease: "logarithmic because the increase follows a compound interest pattern, but it is modified to allow for a decreasing proportion, \( 1-y \), of tissue left for infection" (Vanderplank, 1963) (see Section 8.3.1.3).

For monocyclic (simple interest) diseases, there is an increase in disease without multiplication of the pathogen in the sense that the pathogen does not move from plant to plant and inoculum produced during a growing season (or other unit of time) does not move to uninfected plant tissue and cause infections. One of the examples presented by Vanderplank (1963) was cotton wilt (caused by *Fusarium oxysporum* f.sp. *vasinfectum*). For this disease, the inoculum present in the soil at the beginning of the season remains the main source of inoculum. The increase in the number of wilted plants over time is not caused "primarily by the fungus spreading from one plant to another." For this type of disease, Vanderplank (1963) suggested that one plot \( \ln[1/(1-y)] \) versus time and proposed the use of the monomolecular model for disease increase, which "implies that the progress of disease with time is not logarithmic, and allows for a diminishing proportion, \( 1-y \), of tissue available for infection" (see Section 8.3.1.2). Other examples presented by Vanderplank (1963) for simple-interest type diseases were spotted wilt of tomato (caused by a virus), Pierce's disease of grapevine (caused by *Xylella fastidiosa*), and covered smut or bunt of wheat (caused by *Tilletia* spp.). Of course a monocyclic disease for a single growing season may be polycyclic over many growing seasons.

The description of disease progress was thus divided into two biological categories: diseases with effective multiplication of inoculum during a season and diseases without effective inoculum multiplication. Although the proposals of Vanderplank (1963) served to stimulate thinking and research in plant disease epidemiology, they also served to place several unfortunate limitations on the ways in which researchers viewed the analysis of disease progress. These limitations (or misdirections) arose because of the biological rationale proposed for using the logistic and monomolecular models, from the choice and emphasis of examples presented, and from the subsequent incomplete or inaccurate interpretation of some of the proposals of Vanderplank.

In the years following the publication of Vanderplank's *Plant Diseases: Epidemics and Control*, the general practice of epidemic analysis followed the reasoning that if a disease was one in which multiplication of inoculum occurred during the season, then the logistic model must be appropriate for describing disease progress. If there was no inoculum increase, the monomolecular model should be utilized. So by a relatively common, although erroneous, extension the reasoning was that the logistic model should be used for foliar diseases and the monomolecular model should be used for root diseases. Also, because of the appeal and apparent biological "sense" of this inductive reasoning, the temptation was to make conclusions about the biological nature of the pathosystem based on whether the monomolecular or logistic model fit the disease progress data (Pfender, 1982).

The problem was that although the universal application of two models to disease progress data based on the characteristics of pathogen reproduction may have made biological sense, there was no caveat that the statistical fit of the models to the data is important (see Chapter 7; Madden, 1980, 1986). Proper comparison of epidemics obviously requires that chosen models fit the data. Furthermore, when the statistical fit of a model to data was considered as a criterion for deciding which model to use, the biological nature of the pathosystem really could not be ascertained just because one model was statistically more appropriate than the other (Campbell, 1986; Pfender, 1982). The main difficulty was that the models were being used interchangeably as biological (mechanistic) and statistical (or empirical) models, often without regard to the appropriate biological or statistical assumptions. As we show below, and more so in the next chapter, for example, certain biological assumptions can lead to compound-interest diseases being described by the logistic model. However, the logistic model providing a good fit to data does not necessarily indicate how disease is increasing. Before we can further discuss the interpretation of modeling results, the important growth curve or population dynamic models must be presented.

### 8.3 Models for Analyzing Disease Progress

Dynamic processes, including the change in amount of disease intensity in a population of plants over time, are defined by their rate of change with time. If \( y \) is the amount of disease measured as either severity or incidence, then an epidemic can be described in terms of \( dy/dt \), the change in \( y \) with an infinitesimals change in time \( t \) (see Chapter 7 for a discussion of absolute rates and their estimates). The term \( dy/dt \) represents the absolute rate of disease increase or absolute growth rate. For an epidemic with a degressive phase, \( dy/dt \) can also represent the absolute rate of disease decrease. The
quantitative description of epidemics is accomplished by expressing $\frac{dy}{dt}$ as a function of $y$, $t$, or other variables.

The description and analysis of disease progress curves have been reviewed and discussed extensively (Campbell, 1986; Gilligan, 1985c, 1990; Hau and Kranz, 1977; Jowett et al., 1974; Madden, 1980, 1986; Madden and Campbell, 1986, 1990; Rouse, 1985; Thresh, 1983; Waggoner, 1986). Currently, analysis of disease progress data consists largely of using methods and models adapted from the broader field of growth curve analysis and population dynamics, and in recent years, models and methods adapted specifically for plant disease epidemics. In this section, we attempt to synthesize the most common of these models and to provide examples of their application. In general these models are appropriate for epidemics without a degressive phase.

### 8.3.1 Models with Three or Fewer Parameters

Five models are considered in this section: exponential, monomolecular, logistic, Gompertz, and log-logistic. The first three models have formed the foundation of analysis of disease progress curves since the proposals of Vanderplank (1963). None of the models was developed specifically for applications in plant pathology and thus care should be exercised in attaching strict biological interpretations to the variables and parameters of the models. The differential equations, integrated forms, and linearized forms are summarized in Table 8.1.

#### 8.3.1.1 Exponential

One of the simplest models is generally called the exponential model; Vanderplank (1963) also referred to it as the logarithmic model. Its use dates back implicitly to Malthus (1798) for use in predicting increase in the human population. The model can be written as

$$\frac{dy}{dt} = r_E y$$  \hspace{1cm} (8.1)

in which $r_E$ is a rate parameter specific for the exponential model with units of time$^{-1}$. In eq. 8.1, $r_E$ is the absolute rate of change in $y$ relative to the level of $y$ [$r_E = (dy/dt)(1/y)$]. If $dy/dt$, for example, represents new lesions per day, then $r_E$ equals new lesions per lesion per day. We use a capital letter subscript to identify the rate parameter associated with each specific model. If a rate parameter in general is referred to without regard to a specific model, we use $r_E$.

In the biological interpretation, eq. 8.1 indicates that the absolute rate of disease increase is directly proportional to the amount of disease (and indirectly, inoculum). A greater level of disease leads to greater disease in-
Figure 8.3. Absolute rate (\(dy/dt\) vs. \(t\)) and disease intensity (\(y\) vs. \(t\)) curves at two rates \((r_E)\) for six models used to describe disease progress: A, exponential; B, monomolecular; C, Gompertz; D, log-logistic; and F, Richards with \(m = 2.7\).

creease. The graph of \(dy/dt\) versus \(t\) (or \(y\)) (Fig. 8.3, A) indicates that the absolute rate increases throughout the epidemic when eq. 8.1 is appropriate.

In practice, the level of disease at several times is observed in the population of plants, not the rate of absolute increase. Therefore the exponential model, as well as the other models we consider, is integrated to express \(y\) as a function of \(t\). The integrated form of eq. 8.1 is written as

\[
y = y_0 \exp(r_E t)
\]  

(8.2)

where \(\exp\) represents \(e\) raised to a specific power and \(y_0\) is a constant of integration that also represents the initial disease level if one assumes that the epidemic starts at \(t = 0\). The dependent variable \(y\) could be listed with a \(t\) subscript, \(y_t\), but this generally is not necessary. Equation 8.2 is derived without differential equations in Section 8.7. The graph of \(y\) versus \(t\) (Fig. 8.3, A) shows unlimited increase in \(y\) over time.

The exponential model may be appropriate when there is no limitation on disease increase. Although it is perhaps overly simplistic for most complete epidemics, Vanderplank (1963) suggested that the model may be appropriate for describing the very early stages of a polycyclic epidemic when \(y\) is low, for example, \(y < 0.05\), or perhaps \(y < 0.15\), and host tissue is not limiting.

The linear form of eq. 8.2 is obtained by taking the logarithms of both sides of the equation to produce

\[
\ln(y) = \ln(y_0) + r_E t
\]  

(8.3)

If \(r_E\) is constant, a plot of \(\ln(y) = y^a\) versus \(t\) will produce a straight line with a slope of \(r_E\). The intercept is given by \(\ln(y_0)\). (In Chapter 7, we used the general terms \(b_0\) and \(b_1\) for the intercept and slope, respectively.) Both \(\ln(y_0)\) and \(r_E\) can be estimated by linear regression.

The exponential model can be used in various ways to understand disease progress better. For instance, disease doubling, that is, increase from \(y\) to \(2y\), can be described as

\[
2y = y \exp(r_E t_d)
\]

in which \(2y\) and \(y\) replace \(y\) and \(y_0\), respectively, of eq. 8.2. The doubling time is given by \(t_d\). Dividing both sides of the equation by \(y\), taking the logarithm of both sides of the equation, and rearranging produces

\[
t_d = \frac{\ln(2)}{r_E} = \frac{0.693}{r_E}
\]

which indicates that the doubling time is inversely related to \(r_E\). If \(r_E = 0.693\), \(y\) is doubling every unit of \(t\) (e.g., day). When \(r_E = 0.1/\text{day}\), \(t_d = 6.9\) days. The advantage of using \(t_d\) is that it permits a physical interpretation for specific values of the rate parameter \(r_E\).

### 8.3.1.2 Monomolecular

The monomolecular model is based on monomolecular chemical reactions of the first order. It has also been called the negative exponential model and
has been used for describing many phenomena including cell expansion, response of crops to nutrients (fertilizer), and general growth of plants and animals (Hunt, 1982; Mitscherlich, 1909; Richards, 1969). The rate equation can be written as

\[ \frac{dy}{dt} = r_M(1 - y) \quad (8.4) \]

when it is assumed that the maximum amount of disease \((y_{\text{max}}) = 1 (= 100\%)\). The term \((1 - y)\) represents the proportion of plant tissue or plants that is apparently disease free. According to eq. 8.4, the absolute rate of change in disease is proportional to the level of apparently healthy tissue or proportion of apparently healthy plants. The term apparently healthy is used because not all plant tissue that is infected becomes symptomatic immediately after infection. The quantity \(dy/dt\) declines over time from a maximum at the beginning of the epidemic (Fig. 8.3, B) and has the form of a negative exponential probability density function. Equation 8.4 can also be written as \(dy/dt = r_M - r_M y\). When \(y\) is low (close to 0), \(dy/dt\) is approximately equal to \(r_M\). Thus it is easy to see how \(dy/dt\) does not depend on disease level at low \(y\).

The integrated form of eq. 8.4 can be written as

\[ y = 1 - B \exp(-r_M t) \quad (8.5) \]

in which \(B\) is a constant of integration equal to \((1 - y_0)\) and \(r_M\) once again has units of time \(^{-1}\). A graph of \(y\) versus \(t\) is concave to the time axis and approaches the maximum level of disease, 1, asymptotically (i.e., may come very close to but never quite reach 1). At low \(y\), integration of eq. 8.4 leads to \(y = y_0 + r_M t\), which is the simple-interest model of eq. 8.4.1 (Section 8.7) with \(r_M = kl_0\). Equation 8.5 can be linearized to

\[ \ln \left( \frac{1}{1-y} \right) = \ln \left( \frac{1}{1-y_0} \right) + r_M t \quad (8.6) \]

where \(\ln[1/(1 - y_0)]\) is the \(y\)-axis intercept and \(r_M\) is the slope of the line. If \(r_M\) is constant, the graph of \(\ln[1/(1 - y)]\) versus \(t\) will be a straight line. The quantity \(\ln[1/(1 - y)]\) [or \(-\ln(1 - y)\)] is also known as the multiple infection transformation (Gregory, 1948), but has a different derivation when used for multiple infection (see Section 9.12). At low \(y\), \(\ln[1/(1 - y)]\) and \(y\) are about equal, but as \(y\) increases, \(\ln[1/(1 - y)]\) becomes much larger than \(y\).

In the proposals of Vanderplank (1963), the rate parameter \(r_M\) was viewed as the product of two terms, the amount of inoculum \((q)\) and the rate at which this inoculum causes infections \((w)\) (e.g., \(w\) equals the proportion of spores that result in infection). As such, \(w\) represents the number of new lesions or new diseased plants per unit of inoculum per time. In general, it is not possible to measure both terms independently, and other approaches must be taken to incorporate biological factors into models, particularly for root diseases (Gilligan, 1990). When one is dealing with just a disease progress curve, only the product \(qw\) (\(= r_M\)) can be estimated.

The monomolecular model has been used successfully, for example, to analyze epidemics of tobacco black shank (caused by Phytophthora parasitica var. nicotiana) (Campbell and Powell, 1980; Campbell et al., 1984; Kannwischer and Mitchell, 1978), lettuce drop (caused by Sclerotinia minor) (Jarvis and Hawthorne, 1972), common root rot of wheat (caused by Cochliobolus sativus) (Verma et al., 1974), powdery mildew on barley (caused by Erysiphe graminis) (Hau and Kranz, 1977), and some epidemics caused by maize dwarf mosaic virus (Madden et al., 1987a).

### 8.3.1.3 Logistic

The logistic model has probably been the most important for temporal analysis of disease progress because of its widespread application and appropriateness for describing many epidemics. It was originally proposed by Verhulst (1838) to describe the growth of the human population and was subsequently derived independently by M'Kendrick and Pai (1911) and Pearl and Reed (1920) for use in population growth studies. The model is sometimes referred to as the autocatalytic model because of the type of chemical reactions it can be used to describe.

The differential equation for the logistic model can be written as

\[ \frac{dy}{dt} = r_L y(1 - y) \quad (8.7) \]

where \(r_L\) is a rate parameter with units of time \(^{-1}\) and maximum \(y\) equals 1. The parameter is known as the intrinsic rate in ecology. Vanderplank (1963) called it the "apparent infection rate" because what is actually observed is the apparently diseased or symptomatic tissue. In general, it is not possible to observe the total amount of diseased tissue visually because not all infected tissue will have passed through one incubation period (i.e., the time from infection to symptom expression). If the incubation period is constant, one generally can ignore it in calculating \(r_L\). For convenience, the apparent infection rate is often written without a subscript. If disease level, and thus inoculum (from initial and secondary infections), is considered the principal determining factor in the epidemic, \(dy/dt\) increases as \(y\) increases. This is the same as the exponential model. However, as more and more plants become infected, or greater amounts of tissue become diseased, there is less tissue available to infect and the absolute rate decreases. This is the same as the monomolecular model. In conclusion, when \(r_L\) is constant, the plot...
of $\frac{dy}{dt}$ versus $t$ (Fig. 8.3, C) increases initially, reaches a maximum, and declines to zero. The maximum absolute rate or inflection point occurs at $y = 0.5$ or 50% disease. The plot of $\frac{dy}{dt}$ versus $t$ is symmetrical about this inflection point.

Integration of eq. 8.7 results in

$$ y = \frac{1}{1 + \exp[-(B + r_{LT})]} \quad (8.8) $$

in which $B$ is the constant of integration. In this case, $B$ can be written as $\ln[y_0/(1 - y_0)]$. An alternative expression can be written as

$$ y = \frac{1}{1 + B \exp(-r_{LT})} \quad (8.8a) $$

in which $B$ is then $(1 - y_0)/y_0$. The plot of $y$ versus $t$ is an S-shaped or sigmoid curve (Fig. 8.3, C) that is symmetrical about $y = 0.5$. The time of occurrence of the inflection is $-\ln[y_0/(1 - y_0)]/r_{LT}$. To demonstrate this, $\frac{dy}{dt}$ and $y$ versus $t$ are plotted in the same figure (Fig. 8.4). Maximum $\frac{dy}{dt}$ is about 0.062 and occurs at about $t = 20$. At this time, $y$ equals 0.5.

Equation 8.8 can be linearized to

$$ \ln\left(\frac{y}{1 - y}\right) = \ln\left(\frac{y_0}{1 - y_0}\right) + r_{LT} \quad (8.9) $$

Values of $\ln[y/(1 - y)]$ approach negative and positive infinity as the value of $y$ approaches 0 and 1, respectively. At $y = 0.5$, $\ln[y/(1 - y)]$ is 0.

As a note on terminology, the quantity $\ln[y/(1 - y)]$ was called a “logit” by Berkson (1944) with a subsequent redefinition of logit by Finney (1952). Vanderplank (1963, p. 311) stated that: “We have avoided logit in this book because it is unfamiliar, but regard its use in future as inevitable and to be welcomed.” Indeed, the term logit has become fairly widely used in the literature of plant disease epidemiology. Although we would not discourage such usage, the jargon of epidemiology has increased as more of the models discussed below are utilized in the literature. Berger (1981) used the term “gompit” to refer to the linearizing transformation of the Gompertz model (see below) and, at least in oral presentations, the terms “monit” and “weibit” have been used to refer to transformations associated with the monomolecular and Weibull models. We believe it would be best not to use such jargon, but simply to refer to each of these quantities as the linearizing transformation for the logistic, Gompertz, monomolecular, or Weibull model.

### 8.3.1.4 Gompertz

Gompertz (1825) proposed a population growth model for animals more than a decade before Verhulst (1838) proposed the logistic model. In phytopathology, Analytis (1973b) compared the Gompertz model with a range of growth models for describing disease progress curves and Berger (1981) reiterated the appropriateness of the Gompertz model over the logistic model for describing 113 plant disease epidemics in nine foliar pathosystems.

The differential equation can be written as

$$ \frac{dy}{dt} = r_0 y [\ln(1) - \ln(y)] \quad (8.10) $$

As with the logistic model, a plot of $\frac{dy}{dt}$ versus $t$ (Fig. 8.3, D) increases to a maximum (the inflection point) and then declines to 0. The curve for $\frac{dy}{dt}$ versus $t$ for the Gompertz, unlike the logistic, model is not symmetrical around the inflection point. The absolute rate approaches the inflection point more rapidly and declines more slowly than in the logistic model. The inflection point occurs at $y = 0.37$ ($1/e$) and the curve is skewed positively; in other words, a larger portion of the area under the curve is to the right of the inflection point. As a biological interpretation of this phenomenon, Waggoner (1986) stated that the Gompertz model indicates “that in equal small intervals of time the organism [pathogen] loses equal proportions of its power to increase.”

Integration of eq. 8.10 yields

$$ y = \exp[-B \exp(-r_{gt} t)] \quad (8.11) $$

in which $B$, a constant of integration, equals $-\ln(y_0)$. The plot of $y$ versus
8.3 Models for Analyzing Disease Progress

One of the important aspects of temporal analysis of epidemics is the selection of an appropriate model for describing disease progress data. Model selection is important because the parameters estimated for the model form the basis for statistical analysis and comparison of disease progress curves. Although it is tempting to decide upon the suitability of one of the growth curve models presented based only on the intuitive biological meaning of the differential equation of the model, most often a more appropriate means for evaluating the models is by the criteria identified in Chapter 7, such as the value of the coefficient of determination $R^2$, the mean square error, the standard deviation (error) of the parameter estimates, and the plot of the standardized residuals versus predicted values.

As an example of the application of the models presented in this section, we consider epidemics of potato late blight (caused by Phytophthora infestans) occurring on four cultivars in 1980 in New York. The data for these epidemics were kindly supplied by Dr. W. E. Fry, Department of Plant Pathology, Cornell University. Data were obtained from four replicate plots of each of four potato cultivars. Disease was initiated by focal inoculation with a suspension of P. infestans on 29 July and disease was assessed at 3–4 day intervals after inoculation. Disease assessments for these epidemics are presented in Table 8.2 and the disease progress curves are illustrated in Figure 8.5. For convenience, disease is listed as a percentage, not proportion. The analyses are based on proportions.

Inspection of the disease progress curves shows a clear leveling off of $y$ at high $t$. Therefore the exponential model would not be appropriate. The curves are S-shaped, so the monomolecular model would not be appropriate. Calculation of $dy/dt$ using eq. 7.37 (or 7.40) confirms that the maximum rate is in the middle portion of the epidemics. For instance, average $y$ at each time was calculated for Katahdin and estimated $dy/dt$ plotted versus $t$. Clearly maximum estimated $dy/dt$ (0.064/day) was in the middle part of the epidemic (Fig. 8.6). However, there are not enough data points to use this curve alone to choose the most appropriate model. A larger rate could have occurred a little earlier or later than the time of the calculated maximum.

The linearized forms of the monomolecular, logistic, and Gompertz models were evaluated for goodness-of-fit to the entire set of disease progress data using the General Linear Models procedure (PROC GLM) of the Statistical Analysis System (SAS) (1985). (The monomolecular model is tested only for demonstration purposes.) The SAS statements for the analysis are given in Table 8.3 and a summary of the output and criteria for
TABLE 8.2 Disease Progress Data (% Severity) for Late Blight of Potato (Caused by Phytophthora infestans) for Four Cultivars in New York in 1980 (courtesy W. E. Fry)

<table>
<thead>
<tr>
<th>Days after Inoculation</th>
<th>Replication</th>
<th>Late Blight Severity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Katahdin</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>11.8</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3.9</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>23.2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>26.2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>23.4</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>21.5</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>37.8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>42.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>32.5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>37.3</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>56.3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>58.8</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>56.3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>55.3</td>
</tr>
<tr>
<td>29</td>
<td>1</td>
<td>83.5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>85.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>72.5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>78.0</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>89.3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>92.5</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>91.0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>87.0</td>
</tr>
<tr>
<td>37</td>
<td>1</td>
<td>98.3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>98.3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>98.0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>98.3</td>
</tr>
</tbody>
</table>

evaluating the models are presented in Table 8.4. Plots of the standardized residuals versus predicted values for several of the individual epidemic-model combinations are given in Figure 8.7.

Based on the criteria mentioned above, we concluded that the logistic model was the most appropriate for describing the disease progress data for the cultivar Monona and that the Gompertz model provided the best description of disease progress for the cultivars Katahdin, Kennebec, and Sebago. The estimated rate parameter values were \( r_L = 0.435 \), \( r_G = 0.214 \), \( r_G = 0.164 \), and \( r_G = 0.131 \) per day for disease progress on the cultivars Monona, Katahdin, Kennebec, and Sebago, respectively. How to compare \( r_L \) values from different models is shown in Section 8.3.2.1.

The decision as to which one model is the best may be easier in some cases than others. For example, for Sebago the \( R^2 \) for the Gompertz and logistic models was 97.9 and 94.2%, respectively. (Note that the cumulative nature of a disease progress curve leads to high \( R^2 \) values, even if an inappropriate model is used.) Both these values are desirable and could be acceptable; however, the mean square error for the Gompertz was 0.023 and for the logistic was tenfold larger at 0.230 (Table 8.4). It should be pointed out that one cannot directly compare \( R^2 \) or MSE when different transformations of \( y \) are used (see Chapter 7). \( R^2 \) can be compared and the Gompertz model had a slightly higher value. Also the residual plot for the estimates provided from the logistic model has a distinct and undesirable pattern (Fig. 8.7, G, H) that is not present in the case of the Gompertz model. In the plot of predicted and actual mean disease severity values versus time (Fig. 8.8), the predicted values from the Gompertz models are quite similar to the observed values. Predicted disease severity generated from the logistic model overestimated disease level early and late in the epidemic and underpredicted disease level 18–21 days after inoculation. The conclusion here
is relatively clear—the Gompertz model is superior to the logistic model for describing disease progress on Sebago in 1980. In the case of disease progress on Katahdin, however, the decision is more difficult. The various statistics are quite similar and the residual plots (Fig. 8.7, A, B) both have a slight but probably not unacceptable curving pattern. The plots of predicted versus observed disease severity are also not definitive for differentiating between the two models. The Gompertz model described the early portion (e.g., up to day 21) of the epidemic better than the logistic model did. On days 24 and 29, however, the Gompertz model overpredicted disease level and predicted values deviated more from the observed values than values from the logistic model. The standard deviation of estimated \( r_L \) was slightly less than for \( r_L \), and we subjectively decided that the residuals were slightly closer to random; therefore we chose the Gompertz over the logistic model.

The various linearizing transformations are undefined at some levels of \( y \). For instance, the logistic and Gompertz transformations are undefined at \( y = 0 \) and 1. The monomolecular transformation is undefined at \( y = 1 \). These values were therefore not used in the linear regression analyses. An alternative sometimes used is to substitute a value of \( y \) close to 0 (or 1) when \( y \) equals 0 (or 1). For instance, if the largest \( y \) other than 1 is 0.99, then one could substitute 0.995 for 1.0. When \( y \) represents disease incidence, an alternative is possible. One can define \( y \) as

\[
y = \frac{I + 0.5}{N + 1}
\]

**Figure 8.6.** Estimated \( dy/dt \) (eq. 7.37) versus \( t \) for potato cultivar Katahdin. The mean disease progress curve for the four replications was determined prior to estimating \( dy/dt \).
In the above exercise, the best possible fit to the data was searched for without concern for biological processes. Although the logistic model was slightly inferior to the Gompertz, one could use the logistic model for these epidemics with no serious errors in understanding or making comparisons. It would, however, be a serious error to use the monomolecular or exponential models.
8.3 Models for Analyzing Disease Progress

TABLE 8.5 Summary of Nonlinear Regression Statistics* Used in Evaluation of Three Growth Models for Appropriateness for Describing Disease Progress of Late Blight

<table>
<thead>
<tr>
<th>Model</th>
<th>R² (%)</th>
<th>MSE (x 10^4)</th>
<th>B</th>
<th>s(B)</th>
<th>r_a</th>
<th>s(r_a)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cultivar Katahdin</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monomolecular</td>
<td>91.4</td>
<td>113.1</td>
<td>0.98</td>
<td>0.01</td>
<td>0.079</td>
<td>0.006</td>
</tr>
<tr>
<td>Logistic</td>
<td>98.9</td>
<td>13.8</td>
<td>5.93</td>
<td>0.24</td>
<td>0.257</td>
<td>0.010</td>
</tr>
<tr>
<td>Gompertz</td>
<td>99.0</td>
<td>12.9</td>
<td>4.44</td>
<td>0.85</td>
<td>0.188</td>
<td>0.007</td>
</tr>
<tr>
<td><strong>Cultivar Kennebec</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monomolecular</td>
<td>85.0</td>
<td>17.8</td>
<td>0.98</td>
<td>0.01</td>
<td>0.056</td>
<td>0.005</td>
</tr>
<tr>
<td>Logistic</td>
<td>98.5</td>
<td>17.6</td>
<td>7.99</td>
<td>0.28</td>
<td>0.398</td>
<td>0.014</td>
</tr>
<tr>
<td>Gompertz</td>
<td>99.2</td>
<td>10.2</td>
<td>5.70</td>
<td>0.82</td>
<td>0.162</td>
<td>0.006</td>
</tr>
<tr>
<td><strong>Cultivar Monona</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monomolecular</td>
<td>91.4</td>
<td>13.4</td>
<td>0.96</td>
<td>0.01</td>
<td>0.112</td>
<td>0.009</td>
</tr>
<tr>
<td>Logistic</td>
<td>99.6</td>
<td>6.7</td>
<td>7.99</td>
<td>0.28</td>
<td>0.398</td>
<td>0.014</td>
</tr>
<tr>
<td>Gompertz</td>
<td>99.3</td>
<td>11.2</td>
<td>1.19</td>
<td>3.01</td>
<td>0.275</td>
<td>0.012</td>
</tr>
<tr>
<td><strong>Cultivar Sebago</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monomolecular</td>
<td>86.2</td>
<td>123</td>
<td>0.98</td>
<td>0.01</td>
<td>0.046</td>
<td>0.004</td>
</tr>
<tr>
<td>Logistic</td>
<td>97.4</td>
<td>23.1</td>
<td>5.97</td>
<td>0.31</td>
<td>0.205</td>
<td>0.011</td>
</tr>
<tr>
<td>Gompertz</td>
<td>98.2</td>
<td>15.9</td>
<td>29.33</td>
<td>4.54</td>
<td>0.131</td>
<td>0.006</td>
</tr>
</tbody>
</table>

* R² = coefficient of determination; MSE = mean square error; B = constant of integration; s(B) = estimated standard deviation of estimated B; r_a = rate parameter for listed model; s(r_a) = estimated standard deviation of estimated r_a.

* Caused by Phytophthora infestans on four potato cultivars in New York in 1980 (data courtesy of W. E. Fry).

8.3.2 Models with a Shape Parameter

Each of the models discussed in Section 8.3.1 had a fixed or static shape with the exception of the log-logistic model. It is not entirely reasonable to assume that all plant diseases will have disease progress curves for which dy/dt versus t plot is similar, nor is it even necessarily true that all disease progress curves for a given disease will have the same shaped plot of dy/dt versus t. In order to be able to account for the possible variations in both rate of change in y and shape as separate characteristics of the disease progress curves for a temporal analysis, models are needed that incorporate a shape parameter. With the incorporation of a shape parameter into disease progress models, a more complete description of the epidemic is possible. The use of a shape parameter results in a formal way of comparing r_a for epidemics with different shapes. Shape cannot be separated from rate with the log-logistic model. With a shape parameter in the model, it is no longer possible simply to compare rate (r_a) parameters among epidemics without regard for the shape of the curve.

8.3.2.1 Von Bertalanffy—Richards

Von Bertalanffy (1938, 1957) proposed that change in animal weight is proportional to the difference in rates of anabolism and catabolism. With certain assumptions about the rates of these processes, he developed a growth model with an inflection point at y < 1/e. Richards (1959, 1969) generalized the von Bertalanffy model to account for the many shapes of growth or absolute rate curves, that is, curves of dy/dt versus t.

The Richards model can be written as

\[
\frac{dy}{dt} = r_By(1 - y^{m-1})/(m-1)
\]

(8.16)
in which \( r_R \) is the rate parameter and \( m \) is a shape parameter that can range from 0 to infinity. When \( m = 0 \), eq. 8.15 reduces to the monomolecular model; at \( m = 2 \), the logistic model; and as \( m \) approaches 1 in the limit, the Gompertz (Fig. 8.9). The variation in the plot of \( \frac{dy}{dt} \) versus \( t \) is due not only to the change in \( y, 1 - y, \) or \( -\ln(y) \), as in the fixed models, but depends on a power function of \( y \) (i.e., \( y \) raised to the power \( m - 1 \)). A specific case, when \( m = 2.7 \), is illustrated in Figure 8.3, F.

The differential equation (eq. 8.16) can be integrated to

\[
y = [1 - B \exp(-r_R t)]^{1/(1-m)} \tag{8.17a}
\]

when \( m < 1 \), and

\[
y = [1 + B \exp(-r_R t)]^{1/(1-m)} \tag{8.17b}
\]

when \( m > 1 \), if all parameters are treated as nonnegative numbers. The graph of \( y \) versus \( t \) indicates that a wide range of curve shapes can be obtained by varying \( m \) (Fig. 8.9).

The Richards model can be linearized and the linear form of the model also depends on \( B \):

\[
\ln \left( \frac{1}{1 - y^{1-m}} \right) = -\ln(B) + r_R t \tag{8.18a}
\]

if \( m < 1 \), and

\[
\ln \left( \frac{1}{y^{1-m} - 1} \right) = -\ln(B) + r_R t \tag{8.18b}
\]

if \( m > 1 \). The constant of integration, \( B \), varies with the level of \( m \). If \( m < 1 \), \( B = 1 - (y_0^{1-m}) \) and if \( m > 1 \), then \( B = (y_0^{1-m}) - 1 \).

The Richards model with a fixed value of \( m \) (i.e., the von Bertalanffy model) has had some success as a disease progress model (Analytis, 1973b) but as with applications to plant growth analysis (Hunt, 1982; Venus and Causton, 1979), it will probably be more applicable when the shape parameter is allowed to vary (Jeger, 1982b; Park and Lim, 1985).

The values of \( r_R \) cannot be compared directly if the value of \( m \) is not the same. The reason for this is that the parameter \( r_R \) represents the change in transformed \( y \) (\( y^* \)) per unit change in time. However, \( y^* \) depends on \( m \), so that in actuality, changing \( m \) results in a change in the dependent variable and in the meaning of the slope, \( r_R \). In order to compare rate parameters a new parameter is needed that incorporates both \( r_R \) and \( m \). This new rate parameter is called the weighted mean absolute rate of disease increase (Richards, 1959), and is denoted as \( \rho \) such that

\[
\rho = \frac{r_R}{2m + 2} \tag{8.19}
\]

One can also use this expression when one is comparing \( r_a \) from different single-shape or nonflexible models. For instance, in the potato late blight example, one divides the Gompertz rates (\( r_G \)) by \( 4 \) (\( = (2)(1) + 2 \)) and the logistic rate (\( r_L \)) by \( 6 \) (\( = (2)(2) + 2 \)) in order to make direct comparisons. When \( m \) is statistically estimated as well as \( r_R \), the variance of \( \rho \) is approximated by

\[
\sigma^2 = \frac{r_R^2}{(2m + 2)^2} \left[ \frac{s^2(r_R)}{r_R^2} + \frac{s^2(m)}{(2m + 2)^2} - \frac{2s(r_R, m)}{r_R(2m + 2)} \right] \tag{8.20}
\]
where \( s^2(r_R) \) is the variance of estimated \( r_R \), \( s^2(m) \) is the variance of estimated \( m \), and \( s(r_R, m) \) is the covariance of estimated \( r_R \) and \( m \). This is a large sample result. Its appropriateness for small samples is not known.

### 8.3.2.2 Weibull

Fisher and Tippett developed a statistical distribution in 1928 (quoted in Bailey, 1980) which was later derived independently by Weibull (1939). A later paper (Weibull, 1951) is often given as the primary citation for this model and as a result it has been named after Weibull. The model has been used extensively in life-testing and survival studies, but has also been applied as a growth model (Bailey, 1980; Yang and Smith, 1978), a dose–response model (Rawlings and Cure, 1985) and as a disease progress model (Pennypacker et al., 1980; Thal et al., 1984).

The Weibull model can be written as

\[
\frac{dy}{dt} = \frac{c}{b} \left( \frac{t - a}{b} \right)^{c-1} \exp \left[ -\left( \frac{t - a}{b} \right)^c \right] \tag{8.21}
\]

in which \( a, b, \) and \( c \) are parameters. The location parameter \( a \) represents the earliest occurrence of disease (with units of time) or the time of disease onset. The scale parameter \( b \) is inversely related to the rate of disease increase (units of time). The shape parameter \( c \) is unitless and controls the skewness or shape of the rate curve \((dy/dt) \) vs. \( t \) and the inflection point. A wide range of shapes similar to those for the Richards model (Fig. 8.9) can be obtained with the Weibull model.

Integrating eq. 8.21 results in

\[
y = 1 - \exp \left[ -\left( \frac{t - a}{b} \right)^c \right] \tag{8.22}
\]

which can be linearized to give

\[
\left[ \ln \left( \frac{1}{1 - y} \right) \right]^{1/c} = \frac{-a}{b} + \frac{t}{b} \tag{8.23a}
\]

with \( y \)-axis intercept of \(-a/b\) and a slope of \(1/b\), or to

\[
\ln \left[ \ln \left( \frac{1}{1 - y} \right) \right] = -c \ln(b) + c \ln(t - a) \tag{8.23b}
\]

with an intercept of \(-c \ln(b)\) and a slope of \(c\). Because eq. 8.23a has one of the parameters on the left-hand side of the equation, it cannot be used to estimate all three parameters. Conversely, to use eq. 8.23b the value of \( a \), the time of disease onset, must be known.

The integrated equation of the Weibull can take on many forms and appear quite similar to the Richards model. When \( c = 1 \), the Weibull model reduces to the monomolecular model with \( r_M = 1/b \) and \( B = \exp(a/b) \). There is an inflection point when \( c \) is greater than 1 which shifts to higher values of \( y \) as \( c \) increases. The inflection point occurs at \( y = 1 - \exp((1/c) - 1) \) and 

\[
t = b((c - 1)/c)^{1/c} + a.
\]

From this inflection point equation it can be determined that an inflection point of \( y = 0.5 \) (analogous to the logistic model) occurs when \( y = 3.26 \). Dubey (1967), however, showed that the curve is symmetrical when \( c = 3.6 \). Thus the logistic and Weibull models are not identical but, as Pennypacker et al. (1980) demonstrated, can be similar under a range of conditions. Thal et al. (1984) further extended this work by demonstrating the numerical similarity between the Weibull and the Gompertz, as well as the monomolecular and Richards models when \( m = 0.5 \).

The study of Thal et al. (1984) considered a wide range of \( r_s \) and \( y_0 \) parameters for several nonflexible models and other factors on the estimated parameters of the Weibull model. The value of the rate parameter \( r_s \) did not influence the estimated value of \( c \), but as expected, was related inversely to the estimated \( b \) values. Initial level of disease \((y_0)\) strongly influenced the shape parameter, especially when data were generated from the logistic model. Higher estimates of \( c \) were obtained with lower levels of \( y_0 \). At all levels of \( y_0 \), however, there was no overlap of the shape parameters for the different growth models.

When the Weibull model was fitted to cultivar Katahdin (Table 8.2) data, the following parameters (and their standard errors) were estimated: \( a = 9.2 \) days (1.6), \( b = 16.1 \) days (1.6), and \( c = 2.3 \) (0.3). \( R^2 \) equaled 99.3%. This confirmed the earlier conclusion that the Gompertz model was the most appropriate nonflexible model.

The Weibull model has been criticized as not having any clear biological interpretation and of merely providing a means of empirically describing a wide range of disease progress curves (Rouse, 1985; Waggoner, 1986). Although it is true that the Richards model has a more apparent intuitive biological foundation, there is no reason to expect that the Richards model will be superior to the Weibull in terms of curve-fitting ability. Because many investigators use disease progress models empirically rather than as biological models, and the biological nature of the pathosystem cannot reliably be ascertained from the fit of a particular model to a data set (Pfender, 1982; Campbell, 1986; Madden and Campbell, 1990), there is no a priori reason to choose the Richards over the Weibull or vice versa. Madden and Campbell (1990) have contended that the Weibull model may be more realistic in accounting for the effect of putative infections (rather than disease intensity) on disease increase than the Richards model; however, disease progress
The parameter \( r_* \) also represents the absolute rate relative to a function of \( y \). For instance, for exponential growth, \( r_E = (dy/dt)/y \). For Gompertz growth, \( r_G = (dy/dt)y[-\ln(y)] \). It can be considered an overall “speedometer” for the change in disease over time.

The parameter \( r_* \) is determined by the susceptibility of the host, aggressiveness of the pathogen, and favorability of the environment. Because almost all abiotic and biotic factors can influence \( r_* \), researchers must be cautious in attributing causes for differences in \( r_* \) among epidemics. In general, one should use a valid experimental design to attribute properly one or more causes (e.g., fungicide dosage) for the magnitude of \( r_* \).

It should be realized that the parameters \( y_0 \) and \( r_* \) are not of equal importance in polycyclic epidemics. Of course, if \( y_0 \) or \( r_* \) is 0, no epidemic occurs. However, at \( y_0 > 0 \), \( r_* \) has a relatively greater influence on disease progress than does \( y_0 \). To see this, we consider here the concept of sanitation and the exponential model, which is similar to several other models at low \( y \), and evaluate disease progress at different \( y_0 \) and \( r_E \) values. Suppose one could reduce \( y_0 \) to \( y_0s \), using some control. This could be done by reducing initial inoculum. Any control that reduces \( y_0 \) is known as sanitation; \( y_0s \) is defined as the initial disease after sanitation. With the exponential growth, one can write:

\[
y_0 = y_{0s} \exp(r_E t_s)
\]

in which \( y_{0s} \), \( y_0s \), and \( t_s \) replace \( y \), \( y_0 \), and \( t \), respectively, in eq. 8.2. This equation merely describes the increase from \( y_{0s} \) to \( y_0 \); \( t_s \) is then the time to increase from \( y_{0s} \) to \( y_0 \) at rate \( r_E \). It is the time savings due to sanitation. Rearrangement and logarithms produces

\[
t_s = \frac{1}{r_E} [\ln(y_0) - \ln(y_{0s})]
\]

\[
t_s = \frac{1}{r_E} \ln \left( \frac{y_0}{y_{0s}} \right)
\]

showing that \( t_s \) is inversely proportional to \( r_E \). As \( r_E \) increases, \( t_s \) declines so that the time savings are diminished.

The second form of eq. 8.27 shows that only the ratio of \( y_{0s}/y_0 \) influences the results, not the actual values. The same \( t_s \) would be obtained if \( y_0 = 0.01 \) and \( y_{0s} = 0.0005 \), or \( y_0 = 0.1 \) and \( y_{0s} = 0.005 \). The ratio \( y_{0s}/y_0 (= 20 \) in these examples) is known as the sanitation ratio (SR).

As an example, let \( y_0 = 0.02 \) and \( y_{0s} = 0.002 \), so that SR = 10. If \( r_E = 0.2/\text{day}, \ t_s = 11.5 \) days. It would take 11.5 days for disease to reach the initial starting level without sanitation. If \( r_E \) is halved to 0.1/day, \( t_s \) is doubled to 23 days. However, if SR is doubled to 20, \( t_s \) is increased only from 11 to
15 days. The sanitation ratio has to be increased an order of magnitude (from 10 to 100) to obtain a value of \( t_s \) equal to 23 days.

The equation for \( t_s \) is more complicated when other models hold. If \( y_0 > 0.05 \), for polycyclic diseases, one would need to use other equations depending on the disease progress model. The general equation can be written as

\[
t_s = \frac{1}{r_a} (y_0^* - y_{0s}^*)
\]  
(8.28)

in which \( y^* \) is the appropriate linearizing transformation of \( y \) and \( r_a \) is the corresponding rate parameter. For the logistic model, for instance, \( y_0^* = \ln[y_0/(1 - y_0)] \). One can no longer make general conclusions based on the ratio \( y_0/y_{0s} \).

The above calculations hold when \( r_a \) is not changed by changes in \( y_0 \). This may be true for some diseases, but there are now several examples that show that reductions in \( y_0 \) are associated with increases in the rate parameter (see review by Berger, 1988). When this occurs, the effect of reducing \( y_0 \) can be completely negated by the increase in \( r_a \).

The situation is somewhat different for monocyclic diseases. As mentioned before, \( r_M \) is composed of inoculum \( q \) and infections per unit of inoculum \( (w) \). Reducing \( q \) automatically reduces \( r_M \). Vanderplank (1963) felt that time should be ignored for this type of disease. He stated that, for example, a 10-fold reduction in \( q \) would result in a 10-fold reduction in \( \ln[1/(1 - y)] \), the presumed number of infections. If \( y = 0.75 \) without sanitation, \( \ln[1/(1 - y)] = 1.39 \). A 10-fold reduction equals 1.39/10 = 0.14. Back-transformation of 0.14 shows that \( y \) after sanitation is 0.13. Another approach is to use the monomolecular form of eq. 8.28 but also divide \( r_M \) by the reduction in inoculum (10 in the example). If \( y_0 = 0.02, y_{0s} = 0.002 \), and \( r_M \) before sanitation was 0.03/day, then \( t_s = (1/0.0030)(0.0202 - .0020) = 6.07 \) days, because the new \( r_M = 0.03/10 = 0.003/day \).

8.4 AREA UNDER THE DISEASE PROGRESS CURVE

For some plant disease epidemics, none of the considered growth models may be appropriate for describing an epidemic, possibly due to fluctuations in \( dy/dt \) or the irregular shape of the plot of \( y \) versus \( t \). In such cases, the purpose is to summarize a disease progress curve for comparative or analytical purposes, the area under the disease progress curve (AUDPC) (Fig. 8.10, A) can be used as a descriptor for the epidemic (Shaner and Finney, 1977). Variation in time of disease onset, \( r_a \), and final \( y \) (\( y_f \)) are all incorporated into AUDPC.

AUDPC is simply \( y \) (disease intensity) integrated between two times

(see Section 7.5 for a discussion). It can be approximated using the midpoint rule or so-called trapezoidal integration method. As illustrated in Figure 8.10, B, the disease progress curve is divided into a series of rectangles and the area of each rectangle is summed to approximate the total area under the curve (see eq. 7.42). The AUDPC can be estimated as

\[
\text{AUDPC} = \sum_{i}^{n} \left( \frac{y_i + y_{i+1}}{2} \right) (t_{i+1} - t_i)
\]  
(8.29)

in which \( n \) is the number of assessment times; AUDPC has units of percent-days or proportion-days. The first term in eq. 8.29 is the height of the rectangle (estimated as the midpoint between \( y_i \) and \( y_{i+1} \)) and the second term is the width of the rectangle. The accuracy of eq. 8.29 as an estimate of the true area depends on the size of the interval between disease assessments. The narrower the interassessment intervals, the more accurate the determination of the AUDPC. The AUDPC can be standardized by dividing the AUDPC value by the total time duration \( (t_n - t_1) \) of the epidemic (Fry, 1977) for making the comparisons among epidemics of different durations. This results in an AUDPC that can range from 0 to the maximum \( y \) (100 with percentages, 1 with proportions). This is advisable when the duration of assessments is not the same for each epidemic.

For the epidemics of late blight on the four potato cultivars in New York, 1980, we calculated the AUDPC from the disease progress curve for each replication. The 16 values are presented in Table 8.6. AUDPC ranged from
TABLE 8.6 Analysis of Variance and Mean Separation Test for Area Under the Disease Progress Curve (AUDPC) for Four Replications of Late Blight on Each of Four Potato Cultivars in 1980 in New York (data courtesy W. E. Fry)

<table>
<thead>
<tr>
<th>Replication</th>
<th>Katahdin</th>
<th>Kennebec</th>
<th>Monona</th>
<th>Sebago</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1365.50</td>
<td>1053.25</td>
<td>1664.45</td>
<td>791.25</td>
</tr>
<tr>
<td>2</td>
<td>1449.95</td>
<td>969.95</td>
<td>1703.55</td>
<td>849.10</td>
</tr>
<tr>
<td>3</td>
<td>1311.65</td>
<td>990.85</td>
<td>1669.00</td>
<td>818.60</td>
</tr>
<tr>
<td>4</td>
<td>1327.75</td>
<td>931.65</td>
<td>1691.90</td>
<td>920.10</td>
</tr>
<tr>
<td>Mean</td>
<td>1363.71</td>
<td>986.42</td>
<td>1682.22</td>
<td>844.76</td>
</tr>
</tbody>
</table>

Analysis of Variance for AUDPC

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replication</td>
<td>3</td>
<td>1.393</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>Cultivar</td>
<td>3</td>
<td>572.885</td>
<td>203.80</td>
<td>0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>9</td>
<td>2.811</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mean Separation Based on Fisher's Least Significant Difference (P = .05)

<table>
<thead>
<tr>
<th>Cultivar</th>
<th>Mean Separation</th>
<th>Grouping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monona</td>
<td>1682.22</td>
<td>A</td>
</tr>
<tr>
<td>Katahdin</td>
<td>1363.71</td>
<td>B</td>
</tr>
<tr>
<td>Kennebec</td>
<td>986.42</td>
<td>C</td>
</tr>
<tr>
<td>Sebago</td>
<td>844.76</td>
<td>D</td>
</tr>
</tbody>
</table>

* AUDPC was calculated in eq. 8.29. Proportion-days are obtained by dividing by 100.

791 \%-days for replication 1 for Sebago to 1703 \%-days for replication 2 for Monona. Dividing by time range 26 (= 37 - 11) produces standardized AUDPCs of 30.4 and 65.5\% for these two cases. Mean AUDPC was 845, 986, 1364, and 1682 \%-days for the cultivars Sebago, Kennebec, Katahdin, and Monona, respectively.

8.5 STATISTICAL COMPARISON OF DISEASE PROGRESS CURVES AND DESCRIPTIVE PARAMETERS

A description of a single disease progress curve can give important and useful information about an epidemic; however, researchers generally wish to compare several curves from several epidemics. The goal is to determine if two or more disease progress curves differ significantly from each other. The calculated parameters from disease progress models, AUDPC, and the actual levels of disease at various times are useful characteristics for comparison. Madden (1986) discussed many of the approaches and issues in the statistical analysis and comparison of disease progress curves; other useful techniques for comparison can be found in many standard statistics texts (Neter et al., 1985; Draper and Smith, 1981; Rawlings, 1988; Steel and Torrie, 1980). We present a summary of the most commonly used methods here.

8.5.1 Comparison of Model Parameters and AUDPC Values

8.5.1.1 Standard Error of the Parameter Estimates

Each parameter estimated for a disease progress model, for example, $r_*$, has an associated standard error (deviation) that can be used in statistical comparisons. (As discussed in Chapter 7, for simplicity we do not use "hats" over the parameter symbols to represent estimates.) If $\theta_1$ and $\theta_2$ represent the estimated parameters for two different epidemics, for example, $r_{*1}$ and $r_{*2}$, then a confidence interval for the difference between the two parameters can be estimated by

$$(\theta_1 - \theta_2) \pm t(P/2; n_1 + n_2 - (2p))s[d]$$

where $p$ is the number of parameters in each model; $n_1$ and $n_2$ are the number of observations for the two different disease progress curves (without replications, these would be the number of assessment times); $t(\cdot)$ is the critical value from a $t$ table with significance level $P$ and $n_1 + n_2 - (2p)$ degrees of freedom; and $s[d]$ is the standard error of the difference of parameter estimates. With "one-sided" tables, one "looks up" $P/2$ in the table to obtain $t$ for a significance level of $P$. The standard error of the difference, $s[d]$, is calculated as the square root of the sum of the estimated variances of the two estimated parameters:

$$s[d] = (s^2[\theta_1] + s^2[\theta_2])^{1/2}$$

If the confidence interval (eq. 8.30) does not include zero, then one can reject the null hypothesis that the two parameter values are equal in favor of the alternative hypothesis of inequality.

In order to use directly the confidence interval approach to compare the parameter estimates from two epidemics, the same model must be used for each epidemic. For example, a rate parameter value estimated from a logistic model ($r_L$) cannot be compared directly to a rate parameter value estimated from a Gompertz model ($r_G$). If epidemics described by different models are to be compared, use of the weighted mean absolute rate ($p$) (eq. 8.19) from the Richards model can, however, be calculated for each epidemic and used in the $t$ interval given in eq. 8.30 with the appropriate estimates of the variance of estimated $p$. If separate (nonflexible) models are used, then one should divide the standard deviation of the estimated $r_*$ by $2m + 2$. For instance, $s(r_M)$ and $s(r_L)$ are divided by 2 and 6, respectively, and then squared to obtain $s^2[p]$ for each epidemic.

For example, we wish to compare the rate of disease progress in the late
blight example (Table 8.2) for the potato cultivars Sebago and Katahdin. Both disease progress curves were adequately described by the Gompertz model (Table 8.4), so this is an appropriate comparison. For Sebago the estimate of \( r_G \) was 0.131 per day with the standard error (deviation) of the estimate \([s(\text{r}_G)] = 0.004. For Katahdin, \( r_G = 0.214 \) and \( s(\text{r}_G) = 0.008.\) From eq. 8.31, \( s(\text{d}) = 0.009.\) From eq. 8.30 the confidence interval would equal \((0.131 - 0.214) \pm (2.0)(0.009)\) because \( f(0.025; 28 + 28 - (2)(2)) = 2.0\) with 52 degrees of freedom (7 usable assessment times and four replications per time \(= 28).\) The calculated interval, \(0.083 \pm 0.02,\) does not include zero; therefore we conclude that the rate parameter values are different at \( P = .05.\) There are also more elaborate ways of comparing regression models (i.e., parameters) and these can be found in standard texts (Neter et al., 1985; Rawlings, 1988).

The validity of the confidence interval approach depends on a valid estimate of the standard deviations of the estimated parameters. The cumulative nature of disease progress curves (i.e., \( y \) at \( t \) is made up of \( y \) at \( t - 1 \) plus the increase) may result in incorrect estimates of the standard deviations (Madden, 1986). This is because the \( y \)'s, and possibly the residuals, are not independent. When the residuals exhibit so-called positive autocorrelation, the estimated standard deviations will be lower than the true values, giving a false impression of high precision.

After one performs a linear or nonlinear regression on disease progress curves, the residuals should be tested for autocorrelation. A correlation coefficient or the Durbin-Watson statistic can be calculated with most statistical computer programs (Neter et al., 1985). If the correlation is high, then remedial measures can be taken. Only the simplest approach is given here. If linear regression is being used, one can calculate first differences, \( y_{i}^{*} - y_{i-1}^{*} \) and \( t_{i} - t_{i-1},\) where \( i \) is the \( i \)th observation. For example, consider the three data points and assume that the logistic model is appropriate:

\[
\begin{array}{cccc}
 t_{i} & y_{i} & \ln[y/(1 - y)] & y_{i}^{*} - y_{i-1}^{*} & t_{i} - t_{i-1} \\
 7 & 0.10 & -2.20 & & \\
 14 & 0.30 & -0.85 & 1.35 & 7 \\
 21 & 0.65 & 0.62 & 1.47 & 7 \\
\end{array}
\]

The last two columns form the dependent and independent variables for linear regression. Then a linear regression is performed without a \( y \)-axis intercept. This can be written as a statistical model:

\[
y_{i}^{*} - y_{i-1}^{*} = r_{*}(t_{i} - t_{i-1}) + (\varepsilon_{i} - \varepsilon_{i-1}) \tag{8.32}
\]

This can be effective because the new error term \((\varepsilon_{i} - \varepsilon_{i-1})\) will be inde-

pendent when the original \( \varepsilon \)'s are highly correlated. Details on using this and other methods are provided by Madden (1986). With the potato late blight example, the residuals did not have a significant autocorrelation and therefore no corrections were necessary.

The confidence interval approach can be expanded to more than two disease progress curves and confidence intervals for all pairwise comparisons can be calculated. In the late blight example, because there are four epidemics, there would be \( 4(4 - 1)/2 = 6 \) pairwise comparisons for each estimate of parameter. If each interval is calculated at \( P = .05, \) the overall significance level could be as high as \( 1 - (1 - P)^{6} = .26.\) Some statisticians recommend that the individual \( P \) should be reduced to \( P/f, \) with \( f \) comparisons, to achieve an overall (family) significance level equal to or less than \( P.\) This approach may be cumbersome when there are many comparisons and may result in the detection of few significant differences because \( P/f \) becomes very small. The argument can be made that what is important is the individual comparisons and that controlling the overall significance level serves no purpose (Carmer and Walker, 1982). Epidemiologists should be aware that this is a controversial area.

### 8.5.1.2 Analysis of Variance

When experiments to determine disease progress are divided into blocks or replications, there are several approaches for comparing the disease progress curves. In many cases, the replicate values of disease at each time are used in the regression analysis and the rate parameter, as in the potato late blight example (Table 8.4), is calculated from all the replicate observations. Parameters can then be compared by confidence interval calculation as in Section 8.5.1.1 if the disease progress curves for all replications are described by the same model. If the Richards family of models is appropriate, then \( r \) can be compared. An alternative, however, would be to calculate the model parameters or AUDPC for the disease progress curve in each replication of the study separately and compare the treatments, potato cultivars in the late blight example, by analysis of variance (ANOVA). In this case the parameter or AUDPC estimates, which are random variables, are handled in the same way as variables measured directly in the field.

For AUDPC (Table 8.6), the \( F \)-test for cultivar effect was significant at \( P = .0001, \) indicating that differences exist among the cultivars. Using a mean separation test with preassigned \( P = .05, \) we determine that the AUDPC for each cultivar is different from that of every other. Similarly, if we calculate the weighted mean absolute rate, \( \rho \) (eq. 8.19), for disease progress in each replication for each cultivar, these values (Table 8.7) can also be used in an ANOVA. In this case the \( F \)-test for the cultivar effect was significant at \( P = .0001, \) indicating that the cultivars differed with regard to rate of disease progress. Also, from the mean separation test (Table 8.7), rate of disease progress was not different between plots of the cultivars Kennebec and Sebago, but both of these were different from cultivars Mo-
TABLE 8.7  Weighted Mean Absolute Rate (μ) of Disease Increase, Analysis of Variance, and Mean Separation Test for μ Values for Four Replications of Late Blight on Each of Four Potato Cultivars in 1980 in New York (data courtesy W. E. Fry)

<table>
<thead>
<tr>
<th>Replication</th>
<th>Weighted Mean Absolute Rate (μ) of Disease Increase*</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Katahdin</td>
<td>Kennebec</td>
</tr>
<tr>
<td>1</td>
<td>0.055</td>
<td>0.045</td>
</tr>
<tr>
<td>2</td>
<td>0.044</td>
<td>0.041</td>
</tr>
<tr>
<td>3</td>
<td>0.054</td>
<td>0.039</td>
</tr>
<tr>
<td>4</td>
<td>0.048</td>
<td>0.039</td>
</tr>
</tbody>
</table>

Analysis of Variance for AUDPC

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replication</td>
<td>3</td>
<td>0.000023</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>Cultivar</td>
<td>3</td>
<td>0.001165</td>
<td>38.83</td>
<td>0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>9</td>
<td>0.000030</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mean Separation Based on Fisher’s Least Significant Difference (P = .05)

<table>
<thead>
<tr>
<th>Cultivar</th>
<th>Mean</th>
<th>Grouping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monona</td>
<td>0.073</td>
<td>A</td>
</tr>
<tr>
<td>Katahdin</td>
<td>0.050</td>
<td>B</td>
</tr>
<tr>
<td>Kennebec</td>
<td>0.041</td>
<td>C</td>
</tr>
<tr>
<td>Sebago</td>
<td>0.033</td>
<td>C</td>
</tr>
</tbody>
</table>

* μ values were calculated as $\mu = r/(2m + 2)$ where $r$ is the rate parameter calculated from an appropriate model and $m$ is the assumed shape parameter that corresponds to that model, for example, $m = 0$ for monomolecular, $1 = Gompertz$, and $2 = logistic$. Units are per day.

The Gompertz model best described the disease progress curves for Katahdin (replications 1, 3), Kennebec (replications 1–4), and Sebago (replications 1–4); thus $m = 1$ was used in calculation of μ values. The logistic model best described the disease progress curves for Katahdin and Sebago (replications 2, 4); thus $m = 2$ was used in calculation of the μ values.

variables are highly correlated and the MANOVA allows the relationship among the variables to be analyzed.

For epidemics such as those in the potato late blight example, often only the estimated initial amount of disease ($y_0$) and the rate parameter ($r$) are considered in comparison of disease progress curves. Other variables such as final disease level ($y_f$) and AUDPC may also be appropriate for the overall comparison of the epidemics and all these variables may be highly correlated. With a typical ANOVA performed separately for each variable, misleading results may be obtained from the collection of separate ANOVAs because of the high degree of intercorrelation among the variables.

MANOVA can be used to analyze the vectors of variables to determine if treatment differences exist. Contrasts of the mean vectors can then be utilized to separate treatments. Because the interpretation of multivariate contrasts can be very difficult, Madden (1986) suggests first performing MANOVA, where appropriate, and if certain factors or interactions are statistically significant, then, as a follow-up, ANOVA and contrasts of the means can be used for each variable separately. Although the computational and interpretational difficulties of MANOVA will probably discourage many epidemiologists, MANOVA should nevertheless be a useful and general technique for analyzing disease progress data.

8.5.2 Repeated-Measures Comparison

In some cases it is not possible to describe disease progress data with a growth curve model and AUDPC is not considered to be a desirable alternative. Plots of $y$ versus $t$ may show cyclic trends (Fig. 8.2), discontinuities, or other unusual features such as a decline in disease level near the end of the epidemic (Fig. 8.1, E; Thal and Campbell, 1988; Campbell et al., 1988a). These make the typical epidemic models inappropriate. Interaction effects between time and treatments (e.g., cultivars, fertilizer rates), for example, may have as much or more biological importance in interpreting treatment effects as the possible difference in rate parameters. In such cases an analysis involving the actual disease values at each time, such as a repeated-measures ANOVA, may be useful.

Repeated-measures designs are similar to split-plot, split-unit, or split-block designs and repeated-measures experimental designs are thoroughly described by Gill (1978). In plant disease epidemiology, measurements of $y$ at $t$ times in all experimental units provide a restriction on randomization over time that is analogous to restriction on randomization in space of the “subplot” in agronomic studies. For such repeated-measures designs there are at least two different error variances (mean square errors) for the $F$-test of significance; special attention must be given to determining the appropriate error term to use for testing the significance of treatment effects. The sums of squares and expected mean squares for each source of variation in the ANOVA are given by Gill (1978), Monlezu et al. (1984), and elsewhere.
### 8.6 SUGGESTED READING


### 8.7 APPENDIX: SIMPLE AND COMPOUND INTEREST

It is useful to consider the concepts of simple interest and compound interest accumulation of money in some detail to help plant pathologists understand the models proposed by Vanderplank for disease increase. An example with one interest rate and two time periods is considered here. Let \( t \) represent time (e.g., 1 or 10 yr), \( k \) the interest rate (0.1/yr, i.e., 10% per year), and \( I_0 \) the initial amount of money (100). For simple interest accumulation, total amount of money at any time \( I \) is based on \( I_0 \) and is given by

\[
I = I_0(1 + kt)
\]

At \( t = 1 \) yr, \( I = 100(1 + 0.1) = 100(1.1); \) at \( t = 10 \) yr, \( I = 100(1 + 1) = 100(2). \) This model could be used for monocyclic diseases when there is no
limitation to increase. Equation 8.1 also represents a straight line with slope 
$I_0 k$ and intercept $I_0$.
For compound interest, interest is based on $I_0$ and accumulated interest. 
For discontinuous compound interest, $I$ is given by

$$I = I_0 \left(1 + \frac{k}{n}\right)^{nt} \quad (8.2)$$

in which $n$ is the number of times per unit time period (year in the example) 
that interest is added to the total amount of money. When interest is ac-
cumulated once per year ($n = 1$), eq. 8.2 reduces to $I = I_0(1 + k)^t$. At $t$
= 1 yr, $I = 100(1 + 0.1)^1 = 100(1.1)$, the same as simple interest. At $t$
= 10 yr, however, $I = 100(1 + 0.1)^{10} = 100(2.594)$. When $n = 4$ (quarterly 
accumulation), and $t = 1$ yr, $I = 100(1 + 0.025)^4 = 100(1.104)$; at $t = 10$
yr, $I = 100(1 + 0.025)^{40} = 100(2.685)$. Finally, when $n = 365$ (daily ac-
accumulation) and $t = 1$ yr, $I = 100(1.105)$; at $t = 10$ yr, $I = 100(2.718)$.
Note that the term in parentheses for 10 yr is approaching $e$, the base of 
the natural logarithm system ($e = 2.71828, \ldots$). This is because as $n$ gets 
large (i.e., goes to infinity),

$$\lim_{n \to \infty} (1 + k/n)^{nt} = [(1 + 1/n)^n]^{kt}$$

The term within brackets $[(1 + 1/n)^n]$ equals $e$ at large $n$. When $kt$ equals 
1, at yr = 10 in the example ($0.1 \times 10$), $I_0$ is simply multiplied by $e$.
Daily accumulation of money ($n = 365$) is close to continuous for most practical 
purposes. Therefore, at large $n$ (when interest is being added almost con-
tinuously) eq. 8.2 can be written as

$$I = I_0 e^{kt} \quad (8.3)$$

This equation could be used as a biological model for polycyclic diseases 
when there is no limitation to disease increase.