# A Flexible Model for Studying Plant Disease Progression

S. P. Pennypacker, H. D. Knoble, C. E. Antle, and L. V. Madden

First and fourth authors: associate professor and graduate assistant, respectively, Department of Plant Pathology; second author: research associate, The Computation Center; third author: professor, Department of Statistics, The Pennsylvania State University, University Park 16802.

Contribution 1073, Department of Plant Pathology, The Pennsylvania Agricultural Experiment Station. Authorized for publication 13 March 1979 as Journal Series Paper 5701.

Accepted for publication 22 August 1979.

## ABSTRACT

PENNYPACKER, S. P., H. D. KNOBLE, C. E. ANTLE, and L. V. MADDEN. 1980. A flexible model for studying plant disease progression. Phytopathology 70:232-235.

It is desirable for several reasons for epidemiologists to be able to depict with a single mathematical function the full range of shapes found in disease progress curves. Because of its simplicity and flexibility of application, the Weibull probability density function and cumulative distribution can be applied as such a disease progress model. Disease progress curves corresponding to "simple interest" and "compound interest" disease were accurately fit with the Weibull function. The Weibull model also was used to depict the full range of shapes possible for disease progress curves. Its flexibility and accuracy make the Weibull function a useful technique for modeling plant disease epidemics.

*Additional key words:* quantitative epidemiology, nonlinear model.

The proportion of diseased plants is a simple measure of disease progress during the course of an epidemic. Analysis of this type of data often begins with evaluation of a simple disease progress curve to identify the pattern of disease increase. The epidemic is then often characterized by an associated $r$ value, which in Vanderplank's (11) terminology is either an "apparent" or "absolute" infection rate. In this type of analysis one should ask whether the choice of model is really valid, ie, is the disease increase really typical of true "simple interest disease" (SID, monomolecular) or "compound interest disease" (CID, logistic) types, or do the data fail to support either of the two models? In some cases investigators find that their data do not conform to either of these disease progress models. When this happens it is desirable to use a single mathematical function that will depict the full range of shapes taken on by disease progress curves. Because of its simplicity and flexibility of application, the Weibull probability density function (pdf) and cumulative distribution is such a disease progress model (8,12). Examples of disease progress curves fitted with the Weibull function for both simulated and real data are presented.

## MATHEMATICS OF THE MODEL

The Weibull pdf may be expressed in the form:

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

in which: $y$ = disease proportion; $t$ = time (relative units); $e$ = base of the natural log system, and $t>a$, $b>0$, $c>0$.

The pdf provides an absolute rate, $dy/dt$, of disease increase and the parameters ($a$, $b$, and $c$) of the density function vary with the characteristics of the epidemic under investigation. In this function, $a$ is the location parameter which represents the earliest possible occurrence of disease, $b$ is the scale parameter which is inversely related to the rate of disease increase, and $c$ is the shape parameter which characterizes the manner in which disease progressed and may be used to select the Vanderplank model to which the data

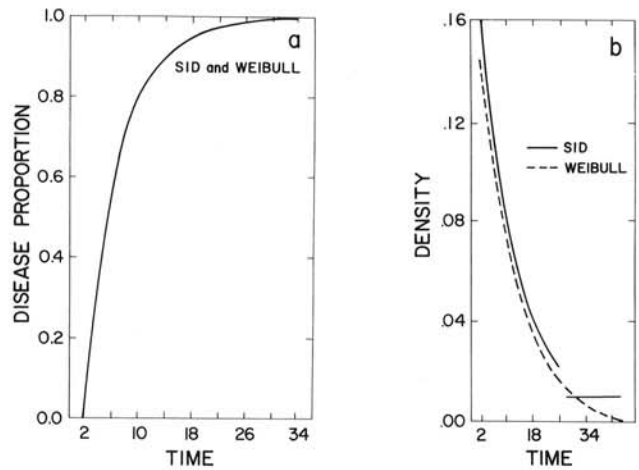


**Fig. 1.** Disease progression data generated by using the simple interest disease (SID) model with an $r$ value of 0.1 and $y_0$ equal to 0.01. Estimated Weibull parameters were: $a = 1.9$, $b = 10.0$, $c = 1.0$. Error sums of squares equaled $4.6 \times 10^{-8}$. **A)** Disease proportion ($y$); **B)** probability density function ($dy/dt$).

0031-949X/80/03023204/$03.00/0

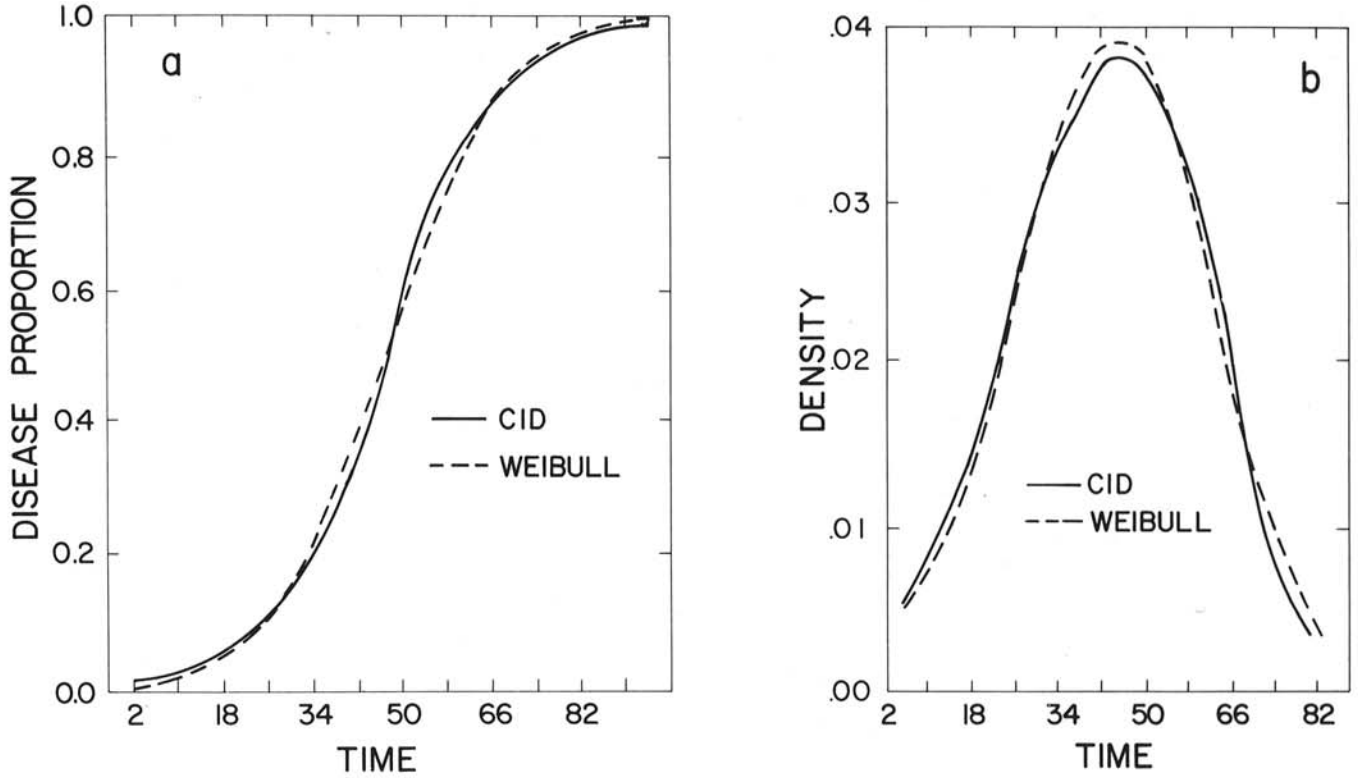


**Fig. 2.** Disease progression data generated by using the compound interest disease (CID) model with an $r$ value of 0.1 and $y_0$ equal to 0.01. Estimated Weibull parameters were: $a = -6.8$, $b = 60.7$, $c = 3.6$. Error sums of squares equaled $5.0 \times 10^{-3}$. **A)** Disease proportion ($y$); **B)** probability density function ($dy/dt$).

more nearly conform.

The Weibull cumulative distribution function is derived by integrating the pdf and is presented in the following expression:

$$y = 1 - \exp\{-[(t-a)/b]^c\}$$

Because the distribution includes the two-parameter exponential as a special case ($c = 1.0$), the Weibull model can be thought of as a generalization of the exponential. Therefore, the Weibull function may be of value to those who have been using the exponential distribution but feel the need for a more flexible model.

A FORTRAN computer program was written to estimate the Weibull parameters. A maximum likelihood technique was used to derive initial estimates of $b$ and $c$. These estimates were refined, and $a$ also was estimated with a scale-invariant, iterative, least squares procedure within the program.

Disease progression data were generated for the SID and CID models by a FORTRAN program. The SID model is:

$$y = 1 - (1-y_0)\, e^{-rt}$$

which is derived by integrating:

$$dy/dt = r(1-y)$$

The variables $t$ and $y$ are defined as before; $y_0$ equals the initial amount of disease and $r$ is the rate parameter. In Vanderplank's terminology, $r$ is replaced by $QR$ and is called the absolute infection rate (11). The CID model equation is:

$$y = \frac{1}{1 + \exp\{-\ln[y_0/(1-y_0)] + rt\}}$$

which is one formulation derived by integrating:

$$dy/dt = ry(1-y)$$

All variables are defined as before. The rate parameter ($r$) of the CID model was termed the apparent infection rate by Vanderplank (11). The $r$ values used in the simulations ranged from 0.05 to 0.70. 0.70.

## APPLICATION OF THE MODEL

Data generated by Vanderplank's (11) equation for SID having an $r$ value of 0.1 are described well with values predicted by the Weibull model (Fig. 1-A). The predicted cumulative values fall on the generated values, and for SID the resulting shape parameter ($c$) equaled 1.0. The corresponding pdf exhibited the expected exponential, and there was close agreement between the generated and the predicted values (Fig. 1-B).

Data generated from Vanderplank's (11) CID model having an $r$ value of 0.1 yielded the typical S-shaped curve (Fig. 2-A). The data fitted according to the Weibull model closely approximated the generated values; for CID the shape parameter was equal to 3.6. The pdf (Fig. 2-B) for CID was symmetrical and approximated a normal distribution.

In comparing the estimated Weibull parameters for increasing $r$ values (Table 1), several characteristic results should be noted. First,

TABLE 1. Estimated Weibull parameters for simple interest (SID) and compound interest (CID) type diseases increasing at selected values of the rate parameter $r$

| | SID | | | CID | | |
|---|---|---|---|---|---|---|
| $r$ | $a^x$ | $b^y$ | $c^z$ | $a^x$ | $b^y$ | $c^z$ |
| 0.05 | 1.9 | 20.0 | 1.0 | −14.0 | 117.9 | 3.6 |
| 0.1 | 1.9 | 10.0 | 1.0 | −6.8 | 60.7 | 3.6 |
| 0.3 | 2.0 | 3.3 | 1.0 | −1.0 | 20.2 | 3.6 |
| 0.5 | 2.0 | 2.0 | 1.0 | 0.2 | 12.2 | 3.6 |
| 0.7 | 2.0 | 1.4 | 1.0 | 0.7 | 8.7 | 3.6 |

[x] Weibull location parameter.
[y] Weibull scale parameter.
[z] Weibull shape parameter.

for rates less than 0.7, shape parameters for CID and SID are 3.6 and 1.0, respectively. Second, the shape paramter, $c$, is independent of the location and scale parameters ($a$ and $b$, respectively). Third, when $c$ equals 1.0, $r$ equals $1/b$.

An example of the unimodal, continuous shapes which may be depicted by the Weibull pdf is illustrated in Fig. 3. When the shape parameter, $c$, is less than 3.6, the function is skewed to the right; when $c = 3.6$, the pdf is symmetrical; when $c$ is greater than 3.6 (as in epidemics in which diseased individuals contribute toward new disease to a greater extent than that approximated by the CID relationship) the pdf is negatively skewed. Ray blight of chrysanthemum (*Chrysanthemum morifolium*[Ramat.] Hemsl.), which is caused by *Mycosphaerella ligulicola* Baker, Dimock, and Davis, is an example of a disease of this third type. The Weibull function was fitted to disease progression data generated by the computer simulator MYCOS (7), and the estimate of the shape parameter was equal to 9.2 (Fig. 4-A). For this value of $c$, the pdf was skewed to the left (Fig. 4-B).

Investigators that choose to model data by using one of Vanderplank's transformations can utilize the Weibull pdf to determine whether the proper model was selected. For Bald's (2) data for spotted wilt of tomato, Vanderplank (11) suggested the SID model because the "spotted wilt virus was entering fields from without" and plant-to-plant spread within the tomato fields could not be detected. Therefore, he classified the disease as SID and used the transformation, $\ln[1/(1-y)]$. Our analysis of the data revealed an estimated shape parameter of 3.4. This indicated that the data were nearly normally (not exponentially) distributed as expected from biological evidence. The data do not support the original assumption of simple interest increase. The disease actually was increasing more like a CID than a SID. Its effect on an estimated $r$ value is readily apparent if the disease increase between times 24 and 31 is considered. For the SID

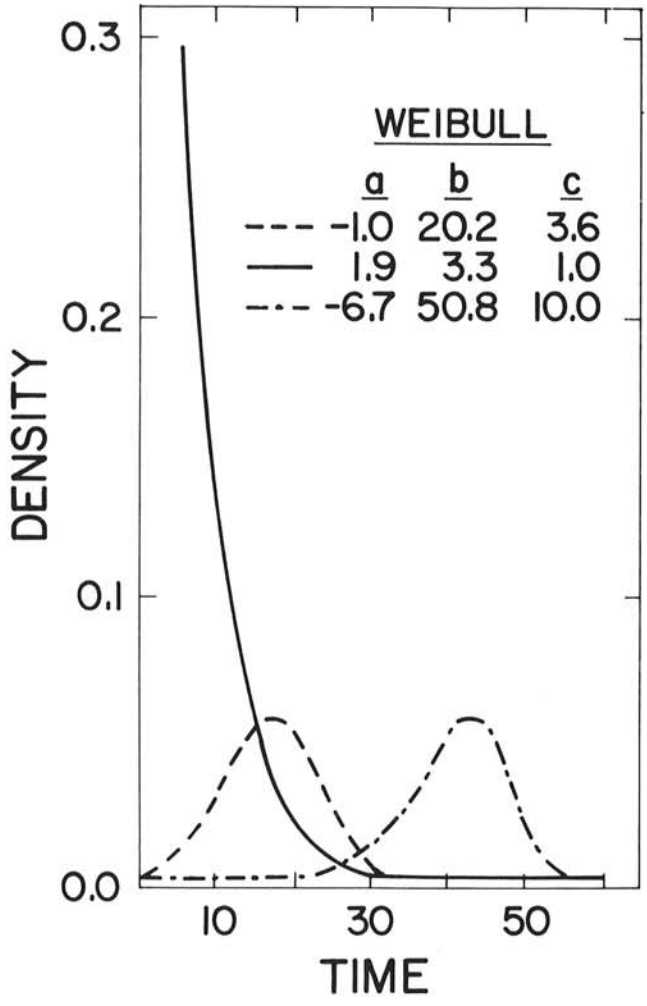


**Fig. 3.** Unimodal continuous shapes of the Weibull probability density function for various values of *a*, *b*, and *c*.

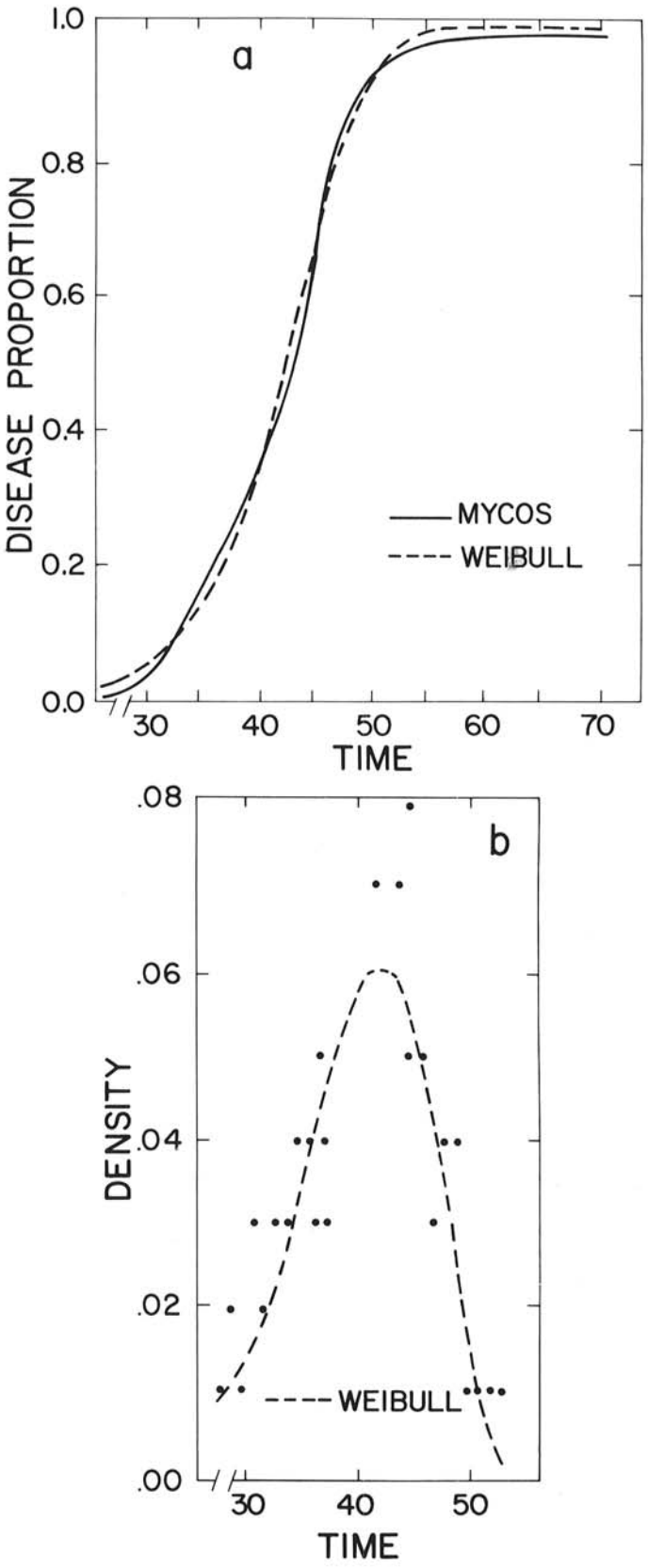


**Fig. 4.** Disease progression data generated by the computer simulator MYCOS (7). Estimated Weibull parameters were: $a = -6.7$, $b = 50.8$, $c = 9.2$. Error sums of squares equaled $1.2 \times 10^{-2}$. **A)** Disease proportion ($y$); **B)** probability density function ($dy/dt$).

model, $r$ equaled 0.051/day; for the CID model, $r$ equaled 0.138/day. This was a difference of approximately 170% in the estimate of $r$.

## DISCUSSION

The Weibull function appears to be a very useful disease progress model: it enables analysis of almost all shapes of disease progress curves, and in the Vanderplank type of analysis, the Weibull parameters can be used to identify the type of disease progress model which most nearly approximates the observed data. It also allows calculation of a rate function at any time, $t$ (by the first derivative of the Weibull distribution [equation 1]) and associated values that provide additional information on the disease progression, namely, parameters that describe location, scale, and shape. An interesting property is that the sum of the location and scale parameters for a given data set is the 63rd percentile of the Weibull distribution, ie, the time, $t$, at which 63% of the plants are infected or when 63% of the surface area is covered by lesions.

"Simple interest" and "compound interest" disease progress models, although the most popular, are only two of several models proposed to describe disease progress curves. Several workers (1,3,5) have used more flexible growth models to analyze epidemics (eg, Richard's function). Fitting data points to these growth models is a nontrivial matter and most techniques are unreliable when final disease level (asymptote) is not known, data sets are small, and estimated parameters are highly correlated ([6] and Knoble and Madden, *unpublished*). These limitations are less severe in estimating the Weibull parameters.

The Weibull cumulative distribution and pdf have been used extensively in life-testing and time-to-failure investigations (9,10,12). The many statistical investigations and developments of the Weibull model have led to its use, not only as a model in the sense of curve fitting, but also to its application to censored data for the prediction of expected future occurrences with associated confidence intervals (4)—a most desirable feature in epidemiological studies.

## LITERATURE CITED


1. ANALYTIS, S. 1973. Zur Methode der Analyse von Epidemien dargestellt an Apfelschorf (*Venturia inaequalis* [Cooke] Aderh.). Acta. Phytomed. 1:1-76.
2. BALD, J. G. 1937. Investigations on "spotted wilt" of tomatoes. III. Infections in field plots. Bull. Council Sci. Ind. Res. Austr. 106:1-32.
3. BERGER, R. D., and J. W. MISHOE. 1976. CSMP simulation of several growth functions to describe epidemic progress. (Abstr.) Proc. Am. Phytopathol. Soc. 3:217.
4. BILLMAN, B. R., C. E. ANTLE, and L. J. BAIN. 1972. Statistical inference from censored Weibull samples. Technometrics 14:831-840.
5. JOWETT, D., J. A. BROWNING, and B. C. HANING. 1974. Non-linear disease progress curves. Pages 115–136 in: J. Kranz, ed. Epidemics of Plant Disease. Mathematical Analysis and Modeling. Springer, Berlin-Heidelberg-New York. 170 pp.
6. KNOBLE, H. D., and F. Y. BORDEN. 1972. Nonlinear estimation of parameters of a flexible growth model. The Pennsylvania State Agri. Exp. Stn. Prog. Rep. 325:1-41.
7. McCOY, R. E. 1971. Epidemiology of chrysanthemum ascochyta blight. PhD Thesis. Cornell University, Ithaca, NY. 177 pp.
8. PENNYPACKER, S. P., C. E. ANTLE, and H. D. KNOBLE. 1974. A flexible model for studying disease progression. (Abstr.) Proc. Am. Phytopathol. Soc. 1:48.
9. PETO, R., and P. LEE. 1973. Weibull distributions for continuous carcinogenesis experiments. Biometrics 29:457-470.
10. ROCKETTE, H. E., C. E. ANTLE, and L. A. KLIMKO. 1974. Maximum likelihood estimation with the Weibull model. J. Am. Statist. Assoc. 69:246-249.
11. VANDERPLANK, J. E. 1963. Plant Diseases: Epidemics and Control. Academic Press, New York. 349 pp.
12. WEIBULL, W. 1951. A statistical distribution function of wide applicability. J. Appl. Mech. 18:293-297.