

Sensitivity of Weibull Model Parameter Estimates to Variation in Simulated Disease Progression Data

Wayne M. Thal, C. Lee Campbell, and L. V. Madden

Graduate assistant and assistant professor, Department of Plant Pathology, North Carolina State University, Raleigh 27650, and assistant professor, Department of Plant Pathology, Ohio State University, Wooster 44691.

Journal Series Paper 9093 of the North Carolina Agricultural Research Service, Raleigh.

Accepted for publication 8 August 1984 (submitted for electronic processing).

ABSTRACT

Thal, W. M., Campbell, C. L., and Madden, L. V. 1984. Sensitivity of Weibull model parameter estimates to variation in simulated disease progression data. *Phytopathology* 74:1425-1430.

Sensitivity of estimates of Weibull parameters (a , b , c) to alterations in factors controlled or not controlled by investigators was examined by using simulated disease progression data generated from the monomolecular, Bertalanffy-Richards (with shape parameter m fixed at 0.5), Gompertz, and logistic models. Data were generated using six levels of initial disease proportion ($y_0 = 0.0000001, 0.000001, 0.00001, 0.0001, 0.001, 0.01$); ten levels of ρ —a common weighted mean rate parameter ($\rho = 0.008, 0.017, 0.025, 0.033, 0.041, 0.050, 0.058, 0.067, 0.075, 0.083$); four data point spacings; and four final disease levels. The effect of a fixed a -parameter on estimated c -parameter values was examined. Weibull model parameters were estimated using nonlinear regression techniques; the Marquardt and

Gauss-Newton methods proved most suitable. Estimates of a were sensitive to values of ρ when data were generated using the Gompertz and logistic models and when values of y_0 were high. Estimates of b were inversely related to ρ and estimates of c were insensitive to changes in ρ . Estimated parameter values for models other than the monomolecular were sensitive to changes in y_0 . Reducing the final level of y affected estimates of c ; the magnitude of the effect increased as the inflection point of the generating disease progress model increased. Estimates of c were insensitive to changes in data point spacing but increased as values of a decreased in the two-parameter model. Estimates of a , b , and c were generally highly correlated which may indicate overparameterization of the model.

Additional key words: disease progression models, quantitative epidemiology.

Over the past two decades, disease progress modeling has become an integral part of the quantitative description of plant diseases. These quantitative descriptions have contributed significantly to our understanding of plant pathosystems and their basic components. Disease progress models have also been used by plant pathologists to evaluate the effectiveness of fungicides, disease resistance, and other disease management tactics (13). The use of flexible models has been suggested for describing disease progress. Pennypacker et al (9) proposed the use of the Weibull

function (14) and it was further discussed by Madden (8); Jowett et al (7) and Madden (8) have also discussed the Bertalanffy-Richards function (4,10) for use in disease progress modeling. Each of these models has a parameter that determines inflection point or shape, in contrast to the exponential, logistic, monomolecular and Gompertz models that lack this parameter. The additional parameter allows flexibility in the position of the inflection point and, thus, the shape of the curve.

In the present paper we examine the usefulness and reliability of the Weibull model (Fig. 1.) for describing disease progress as proposed by Pennypacker et al (9). The Weibull model has three parameters designated in this paper as a , b , and c : a is a location parameter that determines the position of the curve along the independent or time axis, b is a scale parameter that is inversely

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. § 1734 solely to indicate this fact.

proportional to the rate parameter of the Bertalanffy-Richards family of models (10), and c is a shape parameter that determines the location of the inflection point.

It has been shown empirically that the value of the Weibull c -parameter can be used to relate a fitted disease progress curve to either the monomolecular or the logistic growth models (9). The effect of final disease level on c -parameter estimates has been discussed briefly (3). The sensitivity of c -parameter estimates to data-point spacing and to specific combinations of parameter values of the logistic and monomolecular models has not been thoroughly investigated.

Our objective was to determine the sensitivity of Weibull parameter estimates to factors under control of the investigator (ie, data point spacing and statistical fitting method), and to factors not controlled by the investigator such as rate parameter, initial disease level, and final disease level. Generated data were utilized to minimize variation that could affect parameter estimation and obscure the relationships being investigated.

MATERIALS AND METHODS

Generation of disease progression data. Disease progression data were generated by using four forms of the Bertalanffy-Richards model: the monomolecular, the Bertalanffy-Richards with shape parameter (m) fixed at 0.5, the Gompertz, and the logistic (Table 1). The monomolecular and logistic models correspond to the simple-interest and compound-interest disease models of Vanderplank (13). The four models have inflection points (ie, point of maximum rate, dy/dt) at disease proportions (y) ranging from 0.0 to 0.5, and thus cover a broad range of possible shapes (Fig. 1).

Since the rate parameters from different models are not directly comparable, a common weighted mean rate parameter, ρ (10), was used in generating disease progress data. This parameter is the average absolute rate weighted by dy/dt and is defined as $\rho = r/(2m + 2)$ in which r is the rate parameter for the Bertalanffy-Richards family of curves. The equation $r = \rho(2m + 2)$ can be

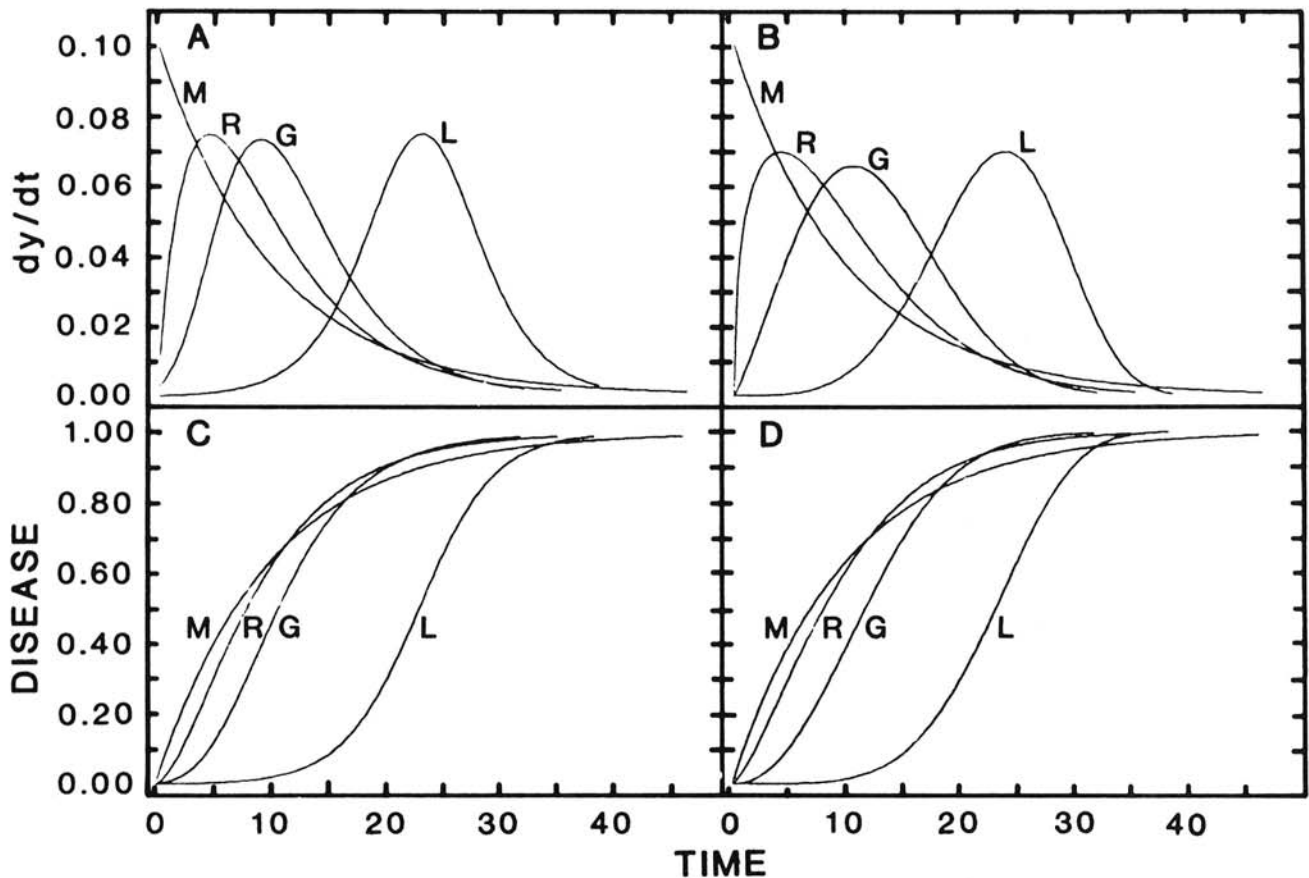


Fig. 1. The A, rate function and C, cumulative distribution (disease progress) function for the Gompertz (G), logistic (L), monomolecular (M), and Bertalanffy-Richards with $m = 0.5$ (R) models which were used in generating disease progress data. Corresponding Weibull B, rate and D, cumulative distribution functions are shown.

TABLE 1. Description of disease progress models

Model	Rate function ($dy/dt =$)	Cumulative distribution function ($y =$)	Inflection point ^f	m^g	n^g	Rate ^h
Weibull	$(c/b)[t - a]/b]^{c-1} e^{-((t-a)/b)^c}$	$1 - e^{-((t-a)/b)^c}$	$1 - e^{[(1/c) - 1]}$	-	-	-
Monomolecular	$r(1 - y)$	$1 - (1 - y_0)e^{-rt}$	0.00	0.0	1.0	$2*\rho$
Bertalanffy-Richards, $m = 0.5$	$2ry(y^{-0.5} - 1)$	$(1 - (1 - \sqrt{y_0})e^{-rt})^2$	0.25	0.5	0.5	$3*\rho$
Gompertz	$ry(-\ln y)$	$e^{(\ln(y_0) + 1)}e^{-rt}$	0.37	~ 1.0	~ 0.0	$4*\rho$
Logistic	$ry(1 - y)$	$[1 + e^{-\ln(y_0)/(1 - y_0) + rt}]^{-1}$	0.50	2.0	-1.0	$6*\rho$

^f Level of y for which dy/dt is a maximum.

^g Value of shape parameter m (9) or n (8) which reduces the Bertalanffy-Richards model to a model of fixed shape.

^h Equation to calculate the rate parameter r based on a fixed value of the weighted mean rate, $\rho = r/(2m + 2)$.

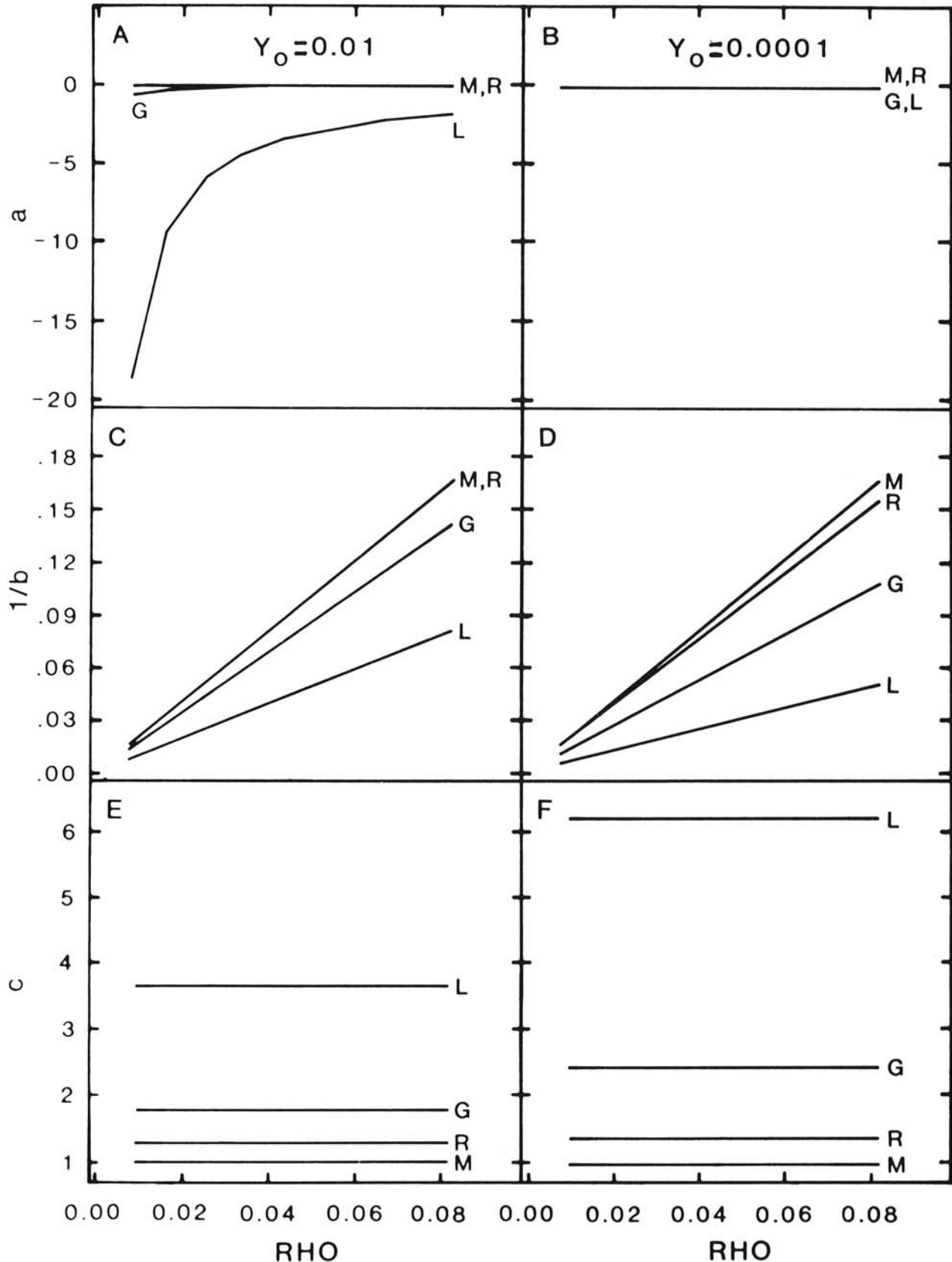


Fig. 2. Relationship between Weibull parameter (a , b , and c) estimates and ρ for the monomolecular (M), Gompertz (G), logistic (L), and Bertalanffy-Richards with $m = 0.5$ (R) models: **A**, a versus ρ , $y_0 = 0.01$; **B**, a versus ρ , $y_0 = 0.0001$; **C**, $1/b$ versus ρ , $y_0 = 0.01$ (regression equations were $G - y = 1.71x$, $L - y = 0.99x$, $M - y = 2.00x$, and $R - y = 2.01x$); **D**, $1/b$ versus ρ , $y_0 = 0.0001$ (regression equations were $G - y = 1.30x$, $L - y = 0.61x$, $M - y = 2.00x$, and $R - y = 1.87x$). **E**, c versus ρ , $y_0 = 0.01$; and **F**, c versus ρ , $y_0 = 0.0001$. Final disease proportion was 0.99.

used to determine the appropriate r for a specified member of the Bertalanffy-Richards family (as determined by m) and a given ρ (Table 1).

Ten values of ρ were used to determine specific rate parameter values for each of the four models (Table 2). Disease progress curves with 21 evenly spaced data points were then generated for each of these rates. Six levels of initial disease, as represented by parameter y_0 , were used in generating disease progress curves to determine the effect of initial disease level on Weibull parameter estimates. Four final levels of disease proportion ($y_f = 0.25, 0.50, 0.75, \text{ and } 0.99$) were used to examine the effect of shortened

epidemics (which could occur when an epidemic is ended due to harvest) on Weibull parameter estimates. The effect of data point spacing was investigated by generating disease progress curves with data points spaced at 5-, 7- and 10-day intervals. The conditions used in generating the disease progress data are summarized in Table 2.

Use of the two-parameter Weibull model. In a separate study, the location parameter, a , was fixed to values of $-10, -5, -3, -2,$ and -1 to test the effect of a fixed location parameter on the estimated c value. This was done because it is easier computationally to fit a nonlinear model with two parameters than with three, and because

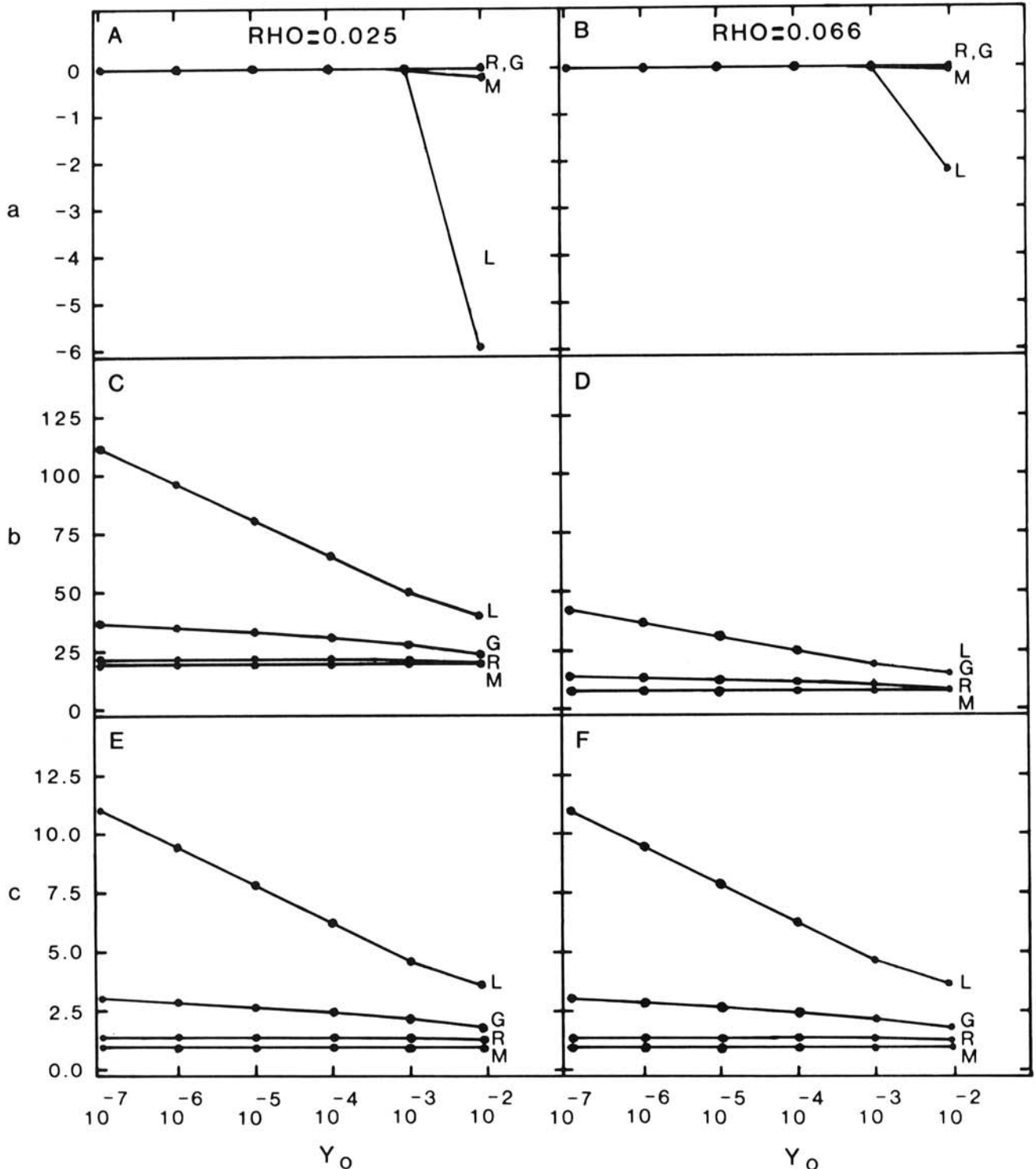


Fig. 3. Relationship between Weibull parameter estimates and y_0 for monomolecular (M), Gompertz (G), logistic (L), and Bertalanffy-Richards with $m = 0.5$ (R). A, a versus y_0 , $\rho = 0.025$; B, a versus y_0 , $\rho = 0.066$; C, b versus y_0 , $\rho = 0.025$; D, b versus y_0 , $\rho = 0.066$; E, c versus y_0 , $\rho = 0.025$; and F, c versus y_0 , $\rho = 0.066$. Final disease proportion was 0.99.

correlations were high among parameter estimates in the three-parameter model, indicating possible overparameterization.

The SAS nonlinear regression procedure NLIN (11,12) provided least squares estimates of the Weibull parameters for the disease progress data. The same initial estimates were used for all regressions. The Marquardt, Gauss-Newton, gradient, and DUD (derivative-free) methods were initially used in computing parameter estimates (11,12). The Marquardt method was used in the reported results unless otherwise indicated.

RESULTS

Effect of rate parameter on Weibull parameter estimates. The relation between the rate parameter of the data-generating model and the estimated Weibull parameters is plotted for two values of y_0 in Fig. 2. The location parameter, a , decreased sharply for low values of ρ with $y_0 = 0.01$ when data generated by the logistic

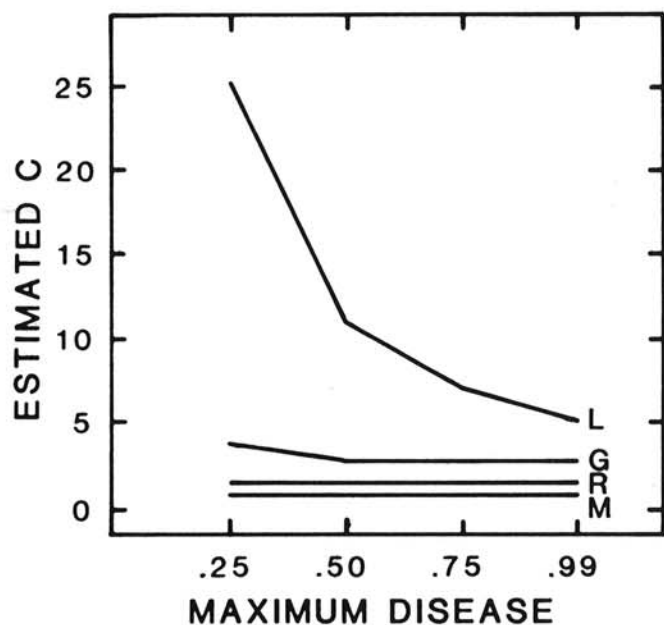


Fig. 4. Effect of final disease level on Weibull c estimates for monomolecular (M), Gompertz (G), logistic (L), and Bertalanffy-Richards $m = 0.5$ (R). Maximum disease is the level at which the epidemics ended. $y_0 = 0.001$. $\rho = 0.50$.

model were used. There was also a decrease in a with $y_0 = 0.0001$; due to the scale, however, it is not visible on the graph. The parameter b was inversely related to ρ , but the relationship generally depended on y_0 . Changes in the rate parameter had little effect on the estimates of c .

Effect of initial disease value on Weibull parameter estimates. The initial level of disease, y_0 , had an effect on estimated parameter values for models other than the monomolecular (Fig. 3). The estimates were most sensitive to changes in y_0 for data from the logistic model and only slightly sensitive for data from the Bertalanffy-Richards model with $m = 0.5$. Values of y_0 near 0.01 had the most influence on a . The effect of y_0 on estimates of a decreased with increasing values of ρ . There were large increases in estimates of both b and c as y_0 values decreased.

Effect of final disease level on Weibull c estimates. The effect of final disease level on estimates of c increased as the inflection point of the generating model increased (Fig. 4.). This effect was similar to that of decreasing y_0 . The c value was insensitive to final disease level for disease progress curves generated from the monomolecular model and changed very little for the Bertalanffy-Richards model with $m = 0.5$. With the Gompertz model and, to a greater extent, with data generated from the logistic model, final disease level affected the value of c .

Effect of data point spacing on Weibull c estimates. Disease progress data generated with points spaced at 5-, 7-, and 10-day intervals gave similar estimates of the Weibull c -parameter. As expected, epidemics were too short in duration to provide enough points for parameter estimation when r was very high, so all comparisons could not be made.

Effect of estimation method. Four estimation methods: Marquardt, Gauss-Newton, gradient, and DUD (derivative-free) were compared for ability to give consistent parameter estimates.

TABLE 2. Parameter values of disease progress models used in generating disease progress data

Parameter	Values
Rate parameter, ρ	0.008, 0.017, 0.025, 0.033, 0.041, 0.050, 0.058, 0.067, 0.075, 0.083.
Initial disease proportion, y_0	0.0000001, 0.000001, 0.00001, 0.0001, 0.001, 0.01.
Data point spacing	5-, 7-, and 10-day intervals
Final disease level, y_f	0.25, 0.50, 0.75, 0.99.
Fixed values of Weibull a parameter	-1, -2, -3, -5, -10.

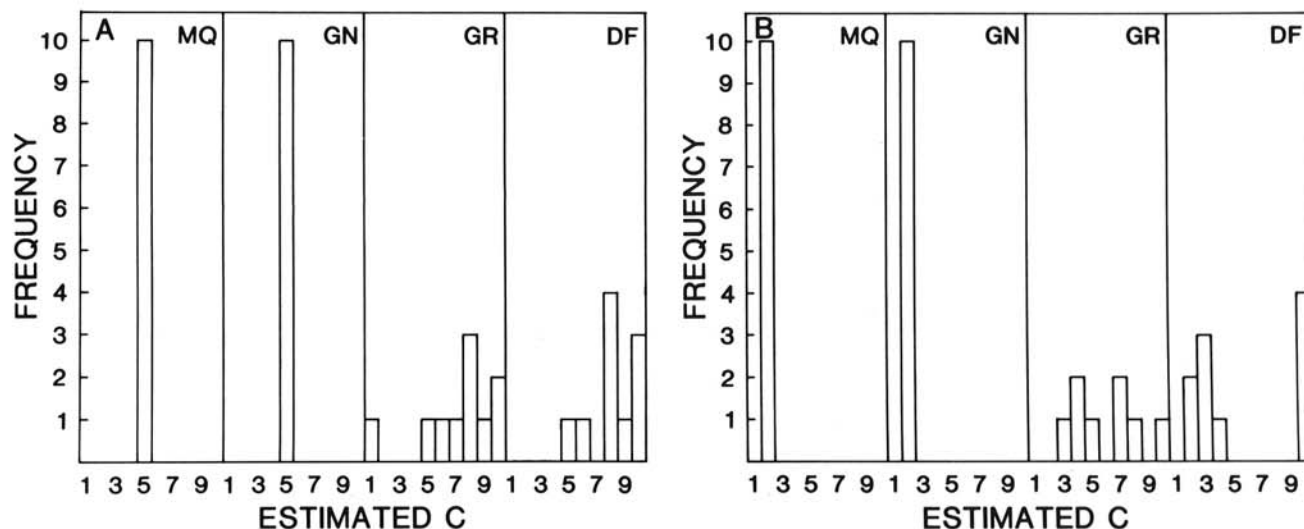


Fig. 5. Frequencies of Weibull c parameter estimates for 10 values of ρ obtained by using the Marquardt (MQ), Gauss-Newton (GN), gradient and derivative-free (DF) estimation methods. $y_0 = 0.001$. A, Estimates for logistic-generated data and B, estimates for Gompertz-generated data. Final disease proportion was 0.99. The gradient method failed to converge for two of the data sets generated from the Gompertz model.

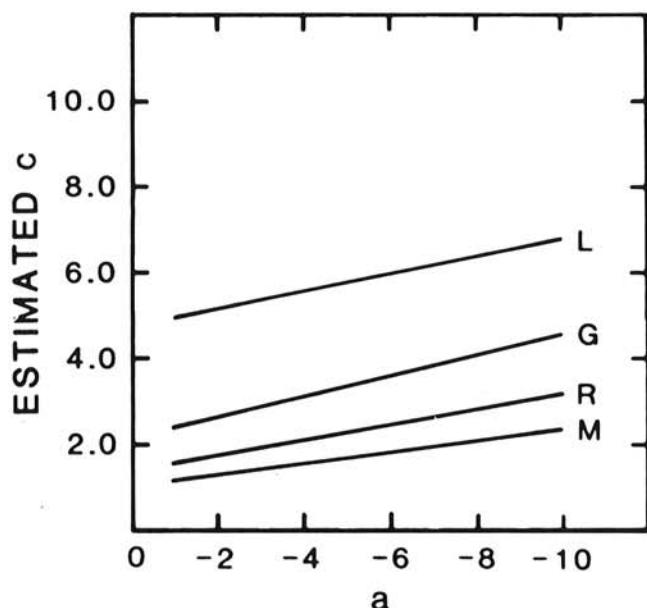


Fig. 6. Effect of fixing Weibull a parameter ($a = -1, -2, -3, -5,$ and -10) on estimates of c of a two-parameter Weibull model. $y_0 = 0.001$. $\rho = 0.050$. Final disease proportion was 0.99.

Histograms of c -values for logistic (Fig. 5A) and Gompertz (Fig. 5B) data with $y_0 = 0.001$ indicate that the Marquardt and Gauss-Newton methods gave the most consistent estimates. These methods gave almost identical estimates for all conditions tested. Estimates obtained using the gradient and DUD methods were not consistent for changes in ρ and sometimes gave estimates with large standard deviations. It may have been possible to get better estimates with these methods if more attention was given to the starting values of the parameters. The Marquardt method was used since it gave good estimates without requiring undue attention to starting values.

Correlation of Weibull parameter estimates. The estimates of a , b , and c were highly correlated, with most correlation coefficients greater than 0.999. Parameters b and c were positively correlated; parameter a was negatively correlated to both b and c .

Estimates of c in the two-parameter Weibull model with fixed a . The value of the fixed location parameter, a , had an effect on c for all data-generating models when the two-parameter Weibull model was fit to the generated disease progress data. There was a linear increase in c as a decreased (Fig. 6).

DISCUSSION

The Weibull model is a flexible model that may offer an improved fit to disease progression data compared with the more commonly used two-parameter models. While two-parameter models are desirable due to their simplicity and ease of interpretation, it is not realistic to assume that all disease progress curves will have the same inflection point. The Gompertz model (inflection point at 0.37) often provides a better fit to disease progress data than the more commonly used logistic model (1,2). Campbell, et al (5) have shown that the best-fitting model for tobacco black shank epidemics varied from one epidemic to another. Evidence of a wide range of inflection points in disease progress data was also indicated for bean hypocotyl rot for which Weibull c -parameter estimates ranged from 1.8 to 9.2 (6). The Weibull model offers a continuous range of inflection points and thus provides an alternative to choosing different models for different epidemics.

The extremely good fits that were consistently obtained by using

the Weibull model on simulated data and on field-collected data (5) make it well-suited for describing disease progress. The parameters correspond to well-defined, epidemiologic concepts, allowing use of the Weibull model for comparative purposes. It also provides a convenient form for representing the data of a disease progress curve and for utilizing these data in models at a higher level of organization, eg, economic or crop loss models incorporating disease information.

The problems of extreme sensitivity to initial disease and to censoring of the data make it necessary to exercise caution when using the Weibull model in a comparative or inferential role since these factors are likely to vary and are generally not controlled by the investigator. The reason for the sensitivity to initial disease appears to be due to the high negative correlation between a and c . Final disease level appears to be most critical when it is below the inflection point.

The high correlations among parameter estimates observed when using the three-parameter Weibull model suggest that it may often be desirable to use the two-parameter form of the model. If the three-parameter form is used, and correlations are high, it must be realized that parameter estimates will have large confidence regions.

Our results should help investigators to assign a disease progress curve to one of the less flexible growth models if this is the intended purpose; however, due to the extreme sensitivity seen under some conditions we feel that investigators should simply use the Weibull model directly as a model of disease progress. Examination and comparison of untransformed biological data (ie, disease progress curves) may give more insight into the biological operations and interactions occurring during epidemics than can be gained by using transformed data. Experience with the Weibull model may lead to a more thorough understanding of the biological significance of the parameters in specific interactions.

LITERATURE CITED

1. Analytis, S. 1973. Methodik der Analyse von Epidemien dargestellt am Apfelschorf (*Venturia inaequalis* (Cooke) Aderh.). Acta Phytomed. 1:1-76.
2. Berger, R. D. 1981. Comparison of the Gompertz and logistic equations to describe plant disease progress. Phytopathology 71:716-719.
3. Berger, R. D. 1982. Problems encountered in fitting disease-progress curves to the Weibull function. (Abstr.) Phytopathology 72:1006.
4. Bertalanffy, L. von. 1938. A quantitative theory of organic growth. Hum. Biol. 10:181-243.
5. Campbell, C. L., Jacobi, W. R., Powell, N. T., and Main, C. E. 1984. Analysis of disease progression and the randomness of occurrence of infected plants during tobacco black shank epidemics. Phytopathology 74:230-235.
6. Campbell, C. L., Pennypacker, S. P., and Madden, L. V. 1980. Progression dynamics of snapbean hypocotyl rot. Phytopathology 70:487-494.
7. Jowett, D., Browning, J. A., and Haning, B. C. 1974. Non-linear disease progress curves. Pages 115-136 in: Epidemics of Plant Disease: Mathematical Analysis and Modeling. J. Kranz, ed. Springer-Verlag, New York. 170 pp.
8. Madden, L. V. 1980. Quantification of disease progression. Prot. Ecol. 2:159-176.
9. Pennypacker, S. P., Knoble, H. D., Antle, C. E., and Madden, L. V. 1980. A flexible model for studying plant disease progression. Phytopathology 70:232-235.
10. Richards, F. J. 1959. A flexible growth function for empirical use. J. Exp. Bot. 10:290-300.
11. SAS Institute Inc. 1982. SAS User's Guide: Basics. 1982 Edition. SAS Institute Inc., Cary, NC. 923 pp.
12. SAS Institute Inc. 1982. SAS User's Guide: Statistics. 1982 Edition. SAS Institute Inc., Cary, NC. 584 pp.
13. Vanderplank, J. E. 1963. Plant Diseases: Epidemics and Control. Academic Press, New York. 349 pp.
14. Weibull, W. 1951. A statistical distribution function of wide applicability. J. Appl. Mech. 18:293-297.