Using the Beta-Binomial Distribution to Describe Aggregated Patterns of Disease Incidence

G. Hughes and L. V. Madden

Institute of Ecology and Resource Management, University of Edinburgh, West Mains Road, Edinburgh EH9 3JG, Scotland, UK; and Department of Plant Pathology, Ohio Agricultural Research and Development Center, Ohio State University, 1680 Madison Avenue, Wooster 44691-4096.

Computer software for the calculation of parameter values and expected frequencies for the beta-binomial distribution is available from the authors on request.

Accepted for publication 22 February 1993.

ABSTRACT

Hughes, G., and Madden, L. V. 1993. Using the beta-binomial distribution to describe aggregated patterns of disease incidence. Phytopathology 83:759-763.

We discuss the use of the beta-binomial distribution for the description of plant disease incidence data, collected on the basis of scoring plants as either "diseased" or "healthy". The beta-binomial is a discrete probability distribution derived by regarding the probability of a plant being diseased (a constant in the binomial distribution) as a beta-distributed variable. An important characteristic of the beta-binomial is that its variance is larger than that of the binomial distribution with the same mean. The beta-binomial distribution, therefore, may serve to describe aggregated disease incidence data. Using maximum likelihood, we estimated beta-binomial parameters p (mean disease incidence) and θ (an index of aggregation) for four previously published sets of disease incidence data in which there were some indications of aggregation. Goodness-of-fit tests showed that, in all these cases, the beta-binomial provided a good description of the observed data and resulted in a better fit than did the binomial distribution. The relationship between the parameters of the beta-binomial distribution and those of variance-mean relationships for aggregated disease-incidence data is shown.

The use of probability distributions to characterize spatial patterns of disease is now a well-established technique in plant disease epidemiology (4,10). For example (7,19,20,23), counts of lesions per sampling unit can be grouped into frequency classes and tested for goodness-of-fit to expected frequencies based on discrete distributions such as the Poisson and the negative binomial. Given certain assumptions (10), the Poisson distribution provides expected frequencies based on the supposition of spatial randomness. Typically, however, the negative binomial distribution provides a better description of observed count data than the Poisson. Because the negative binomial has a larger variance than the Poisson distribution with the same mean, this is taken as an indication of an aggregated (clustered, heterogeneous, patchy) spatial pattern of disease units.

Alternatively, disease incidence (the proportion of plants or plant units diseased [3]) may be assessed. For example, Cochran (5) analyzed the incidence of tomato spotted wilt virus (TSWV) infection of tomato plants in field trials in Australia. If the location of a diseased plant is independent of the location of other diseased plants and there is a constant probability, π, of a plant being diseased, then the number of diseased plants, X, out of n in a sample unit (such as a quadrat) has the binomial distribution

$$Prob\,(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x} \qquad (1)$$

in which $Prob(\cdot)$ represents probability and x takes the values 0, 1, 2, ..., n. The mean and variance of X are then $n\pi$ and $n\pi(1 - \pi)$, respectively. A test of goodness-of-fit of observed frequencies to expected frequencies based on the binomial distribution provides an indication of the homogeneity, or otherwise, of the pattern of disease incidence. For the TSWV data in Table 1, Cochran (5) gave $\chi^2 = 7.97$, with 3 df, which is just significant at $P = 0.05$.

Cochran (5) noted that the observed series differs from the binomial in having too many groups with no diseased plants and four or more diseased plants and too few groups with one, two, or three diseased plants. This outcome is typical of an aggregated pattern of disease incidence. The purpose of this article is to introduce to the plant pathology literature a probability distribution that may be appropriate for the description of such patterns. We provide illustrations based on Cochran's (5) data and some other previously published analyses of virus disease incidence (12–14). The results are discussed in relation to the analysis of aggregated patterns of disease incidence based on variance-mean relationships, suggested recently by Hughes and Madden (9).

MATERIALS AND METHODS

The beta-binomial distribution. Suppose that π in equation 1 is not constant but has the beta density

$$\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi^{\alpha} (1 - \pi)^{\beta}$$

in which $0 \leq \pi \leq 1$, α and β are positive constants, and $\Gamma(\cdot)$ represents the gamma function. A great diversity of shapes may be taken on by this probability density function (16). When either or both of α and β are <1, the graph of the function (Fig. 1) may be J-, L-, or U-shaped. When $\alpha = \beta = 1$, a uniform distribution is obtained and when α and β are both >1, the graph (Fig. 1) is unimodal. For $\alpha = \beta$, the graph of the function (Fig. 1) is symmetrical about $\pi = 0.5$, otherwise it is skewed. Thus, little limitation is being placed on the way that π may fluctuate by the above assumption of a beta density function. Skellam (21) showed that compounding the binomial distribution with a beta density function for π leads to

$$Prob\,(X = x) = \binom{n}{x} \frac{\Gamma(\alpha + \beta)\,\Gamma(\alpha + x)\,\Gamma(\beta + n - x)}{\Gamma(\alpha)\,\Gamma\,(\beta)\,\Gamma(\alpha + \beta + n)} \qquad (2)$$

in which x takes the values 0, 1, 2, ..., n. This is often called the beta-binomial distribution (6). It is one of a number of models available for the analysis of binary data such as disease incidence (1).

Clearly, the derivation of the beta-binomial is analogous to

the derivation of the negative binomial distribution by compounding the Poisson with a gamma distribution (15). The Poisson distribution is the limiting form of the binomial distribution when n is large and π is small (17). The negative binomial is, in fact, the limiting form of the beta-binomial distribution when n and $\alpha + \beta$ are large (17,21).

For the purpose of parameter estimation, it is convenient to write

$$p = \alpha(\alpha + \beta)^{-1} \quad (3a)$$

and

$$\theta = (\alpha + \beta)^{-1} \quad (3b)$$

p being mean disease incidence (i.e., the expected value of the now variable binomial parameter π) and θ a measure of the variation in π. In this parameterization, the mean and variance of X are np and $np(1 - p)\,(1 + n\theta)\,(1 + \theta)^{-1}$, respectively (8). Thus, when $\theta > 0$, the variance of the beta-binomial distribution is larger than the variance of the binomial with the same mean, and the "pure binomial" is obtained when $\theta = 0$. On this basis, the parameter θ can be thought of as an index of aggregation. A published algorithm (22) enables the estimation of p and θ by maximum likelihood.

Data. Cochran's (5) data comprise the number of TSWV-infected tomato plants out of $n = 9$ in each of 160 quadrats. A similar example is provided by the data of Marcus et al (14). Their Figure 1 shows the location of orange trees infected with citrus tristeza virus (CTV) in an orchard in central Israel. We divided the field map into 84 "quadrats" of 4 rows × 3 columns and counted the total number (1981 + 1982) of infected trees out of a maximum of $n = 12$ in each quadrat.

In addition to these two single disease assessments, we examined data from two disease epidemics. The first was reported by Madden et al (13) in a study of the incidence of tobacco etch virus (TEV) and tobacco vein mottling virus (TVMV) in experimental tobacco fields in Kentucky. At each of 18 disease assessments made in field A-N-1985, the number of infected plants out of $n = 40$ in each of 75 quadrats was recorded. The second epidemic from which data were examined was reported by Madden et al (12) in a study of the incidence of maize dwarf mosaic virus (MDMV) in experimental maize field plots in Ohio. At each of six disease assessments made in field P-3, the number of infected plants out of $n = 100$ in each of 36 quadrats was recorded.

Analysis. Using the algorithm of Smith (22), we calculated maximum likelihood estimates of p and θ for each disease assessment. The moment estimates $\hat{p} = m/n$, $\hat{\theta} = [s^2 - n\hat{p}(1 - \hat{p})]/[n^2\hat{p}(1 - \hat{p}) - s^2]$, in which m and s^2 are the observed mean and variance of the number of diseased plants per quadrat, respectively (8), were used as starting values for the maximum likelihood estimation procedure.

We then calculated the expected frequencies for the beta-binomial distribution with a method based on that suggested by Skellam (21). The expected frequencies for the binomial distribution were calculated from equation 1 with $\pi = \hat{p}$. After pooling frequency classes so that expected frequencies were >5, we calculated, wherever possible, the χ^2 goodness-of-fit statistic for the observed data and both the binomial and beta-binomial distributions in turn. For the former, the number of degrees of freedom is the number of frequency classes, after pooling, minus two; for the latter, number of classes, after pooling, minus three.

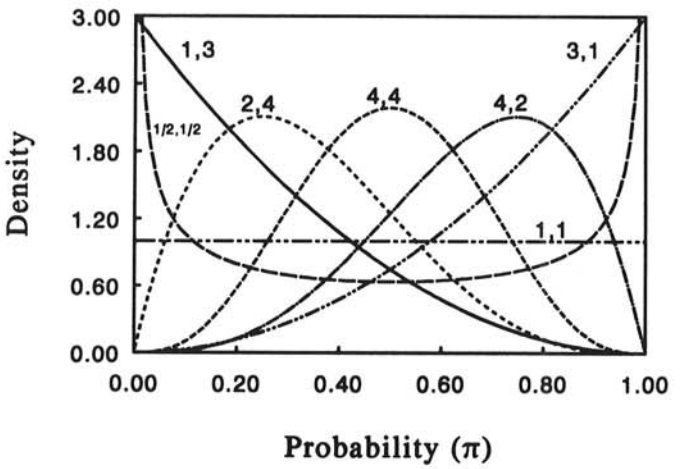

Fig. 1. Graphs of the beta density function for various values of parameters α,β.

RESULTS

Tomato spotted wilt virus. The maximum likelihood estimates of the beta-binomial parameters were $\hat{p} = 0.181$ (standard error [SE] = 0.0119) and $\hat{\theta} = 0.053$ (SE = 0.0204). It is clear that the description of the frequency distribution of diseased plants provided by the beta-binomial was a significant improvement over that of the binomial distribution (Table 1). For the beta-binomial the goodness-of-fit $\chi^2 = 0.10$ with 3 df ($P > 0.99$). For similar data presented by Bald (2), the binomial distribution appeared to provide an adequate description of the frequency distribution of disease incidence, suggesting a random pattern of diseased plants. In such a case, the use of the beta-binomial distribution would be superfluous.

TABLE 1. Observed and expected (binomial and beta-binomial) frequencies for the tomato spotted wilt virus data reported by Cochran (5)

Number of diseased plants per quadrat	Observed frequency	Expected binomial frequency[a]	Expected beta-binomial frequency[b]
0	36	26.45	36.56
1	48	52.70	47.69
2	38	46.67	37.51
3	23	24.11	22.16
4	10	8.00	10.48
5	3	1.77	4.02
6	1	0.25	1.23
7	1	0.03	0.29
8	0	0.00	0.05
9	0	0.00	0.00

[a] Frequency classes 4–9 were pooled for calculation of the χ^2 statistic (described in text).
[b] Frequency classes 5–9 were pooled for calculation of the χ^2 statistic (described in text).

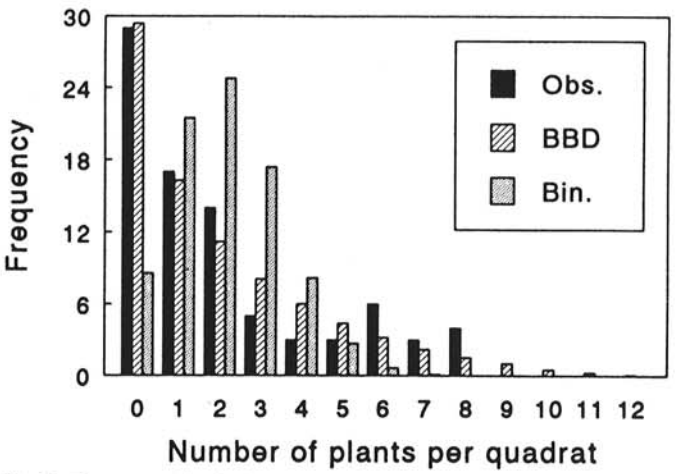

Fig. 2. Frequency distribution of orange trees infected by citrus tristeza virus in 84 quadrats with 12 trees per quadrat. Observed (Obs.) and expected frequencies for the binomial (Bin.) and beta-binomial (BBD) distributions are shown. Estimated parameters and goodness-of-fit statistics are given in the text.

Citrus tristeza virus. The maximum likelihood estimates of the beta-binomial parameters were $\hat{p} = 0.173$ (SE = 0.0203) and $\hat{\theta} = 0.271$ (SE = 0.0628). The observed frequencies and the expected frequencies for both the binomial and beta-binomial distributions are shown in Figure 2, which shows that the binomial distribution tended to underestimate the observed frequencies of diseased plants in the tails of the distribution and overestimate those in the center. The beta-binomial distribution again provided a much better description of the frequency distribution of diseased plants than the binomial. For the former, the goodness-of-fit $\chi^2 = 4.04$, with 4 df ($P = 0.40$), and for the latter, $\chi^2 = 63.1$, with 3 df ($P < 0.001$). We also found that dividing up the field map into 42 "quadrats" of 4 rows × 6 columns (so that $n = 24$) had very little effect on the estimates of the beta-binomial parameters, which were, in this case, $\hat{p} = 0.173$ (SE = 0.0259) and $\hat{\theta} = 0.261$ (SE = 0.0694).

Tobacco etch virus and tobacco vein mottling virus. We made separate maximum likelihood estimates of p and θ for each disease assessment, wherever possible (at disease assessments 1–3, the data comprise mostly zeros, resulting in failure of the damped Newton-Raphson procedure used in the maximum likelihood estimation subroutine). Mean disease incidence for both viruses combined ($\hat{p}$) increased over time, in a sigmoid fashion (found also in [13]), while $\hat{\theta}$ increased to a peak at around $p = 0.5$, then decreased (Table 2). The χ^2 goodness-of-fit statistics (Table 2) indicated the improved description of the frequency distribution of diseased plants provided by the beta-binomial, compared to the binomial distribution, over the entire course of the epidemic. Figure 3 shows the observed and expected frequencies at two levels of mean disease incidence. Comparison of the observed frequencies with the binomial expected frequencies clearly indicates the aggregated nature of disease incidence at both assessments. In both cases, a good description of the observed data was provided by the beta-binomial expected frequencies.

Maize dwarf mosaic virus. We again made separate maximum likelihood estimates of p and θ for each disease assessment (Table 3). Mean disease incidence ($\hat{p}$) increased over time, in a sigmoid fashion (found also in [12]), while $\hat{\theta}$ increased to a peak and then decreased (Table 3). In this case, the peak value of $\hat{\theta}$ appeared to occur at a value of p below 0.5. The χ^2 goodness-of-fit statistics (Table 3) indicated, as before, that the beta-binomial distribution described the observed frequency distributions of disease incidence better than the binomial distribution.

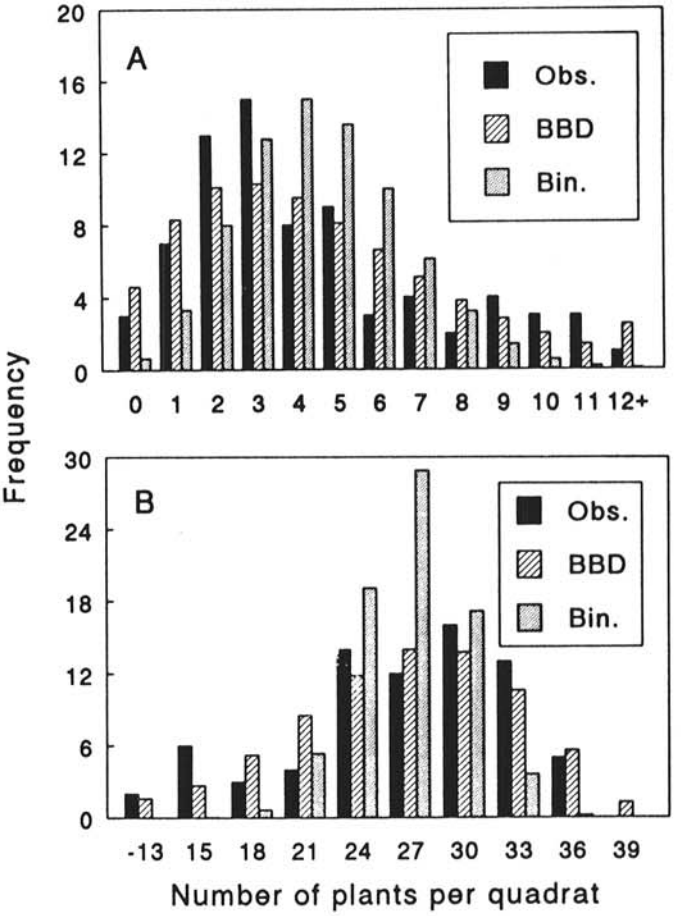

Fig. 3. Frequency distribution of tobacco plants infected by tobacco etch virus and tobacco vein mottling virus at **A,** 23 and **B,** 42 days in 75 quadrats with 40 plants each. Observed (Obs.) and expected frequencies for the binomial (Bin.) and beta-binomial (BBD) distributions are shown. For presentation purposes, frequency classes 12–40 (12+) were pooled at day 23 and 0–13 (−13) at day 42. Furthermore, for day 42, frequencies are given for groups of three classes, and the midpoints are shown (15 for classes 14–16, 18 for classes 17–19, etc.). Estimated parameters and goodness-of-fit statistics are given in Table 2.

TABLE 2. Beta-binomial parameter estimates and goodness-of-fit statistics for the tobacco etch virus and tobacco vein mottling virus epidemic in tobacco field A-N-1985 reported by Madden et al (13)

Disease assessment	Day	Parameter estimates[a]				Goodness-of-fit[b]					
						Binomial[c]			Beta-binomial		
		$\hat{p}$	SE	$\hat{\theta}$	SE	χ^2	df	P	χ^2	df	P
1	0	0.0007		-							
2	2	0.0013		-							
3	7	0.0020		-							
4	9	0.0040	0.00124	0.0056	0.00667	-			-		
5	12	0.0086	0.00205	0.019	0.0102	-			-		
6	14	0.012	0.0023	0.010	0.0071	-			-		
7	16	0.023	0.0035	0.021	0.0098	5.2	1	0.022	0.4	1	0.52
8	19	0.063	0.0060	0.023	0.0092	14.5	4	0.006	4.1	4	0.40
9	21	0.081	0.0070	0.026	0.0099	11.0	4	0.027	1.8	4	0.77
10[d]	23	0.112	0.0088	0.042	0.0101	31.5	5	<0.001	6.2	5	0.29
11	26	0.156	0.0107	0.049	0.0113	49.5	6	<0.001	7.0	7	0.43
12	28	0.235	0.0142	0.067	0.0156	71.5	6	<0.001	7.6	7	0.37
13	30	0.323	0.0170	0.082	0.0179	83.6	7	<0.001	7.2	6	0.30
14	33	0.443	0.0204	0.113	0.0230	127.9	7	<0.001	3.9	7	0.79
15	35	0.550	0.0211	0.123	0.0249	133.9	7	<0.001	5.8	7	0.56
16	37	0.628	0.0192	0.103	0.0222	144.7	7	<0.001	8.4	7	0.30
17	40	0.648	0.0183	0.092	0.0207	69.4	7	<0.001	6.5	6	0.37
18[d]	42	0.668	0.0176	0.086	0.0197	77.6	7	<0.001	9.9	6	0.13

[a] Parameter estimates could not be made from the data for disease assessments 1–3.
[b] Degrees of freedom (df) were determined by pooling frequency classes so that expected frequencies were >5. There were insufficient df for the goodness-of-fit test at disease assessments 4–6.
[c] Calculated from equation 1 with $\pi = \hat{p}$.
[d] Described in Figure 3.

DISCUSSION

When virus infection occurs in an aggregated pattern, the binomial distribution provides an inadequate description of the frequency distribution of diseased plants per quadrat. On the basis of results from the pathosystems used as examples, it is clear that the description provided by the beta-binomial distribution is a significant improvement. This is true over a range of combinations of quadrat size and number of quadrats and at different levels of mean disease incidence. The cost of this improvement is that, by comparison with the binomial distribution, an additional parameter must be estimated at each disease assessment.

In a previous paper (9), we investigated the relationship between the variance, v [=var(π)], and mean, p ($=\bar{\pi}$), for aggregated disease incidence data. Good descriptions of epidemiological data were provided by the relationships

$$v = a[p(1-p)]^b \tag{4a}$$

and

$$v = ap^{b1}(1-p)^{b2} \tag{4b}$$

in which a and either b (equation 4a) or b_1 and b_2 (equation 4b) are parameters to be estimated. The curve of equation 4a is symmetrical about a maximum at $p = 0.5$, whereas that of equation 4b may be asymmetrical. Equations 4a and 4b are of interest in the present context because they represent assumptions about the first two moments of the distribution of (the now variable) π that are satisfied by beta distributions $B(\alpha,\beta)$ in which

$$\alpha = \{[f(p)/a] - 1\}p \tag{5a}$$

and

$$\beta = \{[f(p)/a] - 1\}(1-p). \tag{5b}$$

In the case of equation 4a, $f(p) = [p(1-p)]^{1-b}$, whereas in the case of equation 4b, $f(p) = p^{1-b1}(1-p)^{1-b2}$.

For clustered binary data, constant θ is sometimes assumed (e.g., [18]), leading to $v = ap(1-p)$. The parameters of the corresponding beta distribution are now given by equations 5a and 5b, with $f(p) = 1$ (1). This situation is analogous to that envisaged by Waggoner and Rich (24) when they described disease progress by a logistic equation modified to take account of an aggregated pattern of propagules by incorporation of a (constant) negative binomial parameter, k. The equivalent equation incorporating the beta-binomial parameter θ could describe the rate of increase of disease incidence in, for example, a virus-vector-host pathosystem in which the vector had an aggregated pattern. However, on the basis of Tables 2 and 3, it seems as unrealistic to assume constant θ over a wide range of p as it does to assume constant k with changing disease intensity (10).

Empirically, the variation of $\hat{\theta}$ with $\hat{p}$ seems to be characterized by a curve with a single maximum at $0 < p < 1$ (Tables 2 and 3). Substituting equations 5a and 5b into equation 3b gives

$$\theta = \frac{a}{f(p) - a} \tag{6}$$

which has this shape when $b > 1$ in equation 4a, in which case the maximum value of θ occurs at $p = 0.5$, or b_1 and $b_2 > 1$ in equation 4b, in which case the maximum value of θ occurs at $p = (1 - b_1)/[(1 - b_1) + (1 - b_2)]$. Values of b and b_1 are >1 for aggregated disease incidence data, but the value of b_2 may vary widely (9). Values of $b_2 \leq 1$ may serve to provide an empirical description of some variance-mean relationships, but those relationships do not provide useful information in the context of equation 6.

For the tobacco virus disease epidemic described in Table 2, in which θ appears to be at a maximum at about $p = 0.5$, equation 4a provided a good description of the variance-mean relationship ($\hat{a} = 0.16$, $\hat{b} = 1.29$; $r^2 = 0.995$) (found also in [9]). For the maize virus disease epidemic described in Table 3, in which θ appears to be at a maximum at $0 < p < 0.5$, equation 4b ($\hat{a} = 0.08$, $\hat{b}_1 = 1.31$, $\hat{b}_2 = 1.77$; $r^2 = 0.887$) provides a better description of the variance-mean relationship than equation 4a ($\hat{a} = 0.05$, $\hat{b} = 1.20$; $r^2 = 0.753$). The estimated coefficients of equation 4b for the MDMV epidemic correspond to maximum θ at $p = 0.29$, which is consistent with Table 3.

The correspondence between the beta-binomial distribution and the variance-mean relationships described by Hughes and Madden (9) provides a basis for reducing, from two at each disease assessment, the number of parameters required to describe an epidemic. First, temporal variation in p can usually be characterized by a nonlinear disease progress curve requiring the estimation of just three parameters (the initial and maximum [asymptotic] levels of disease and a rate parameter) for an epidemic (11). Second, variation in θ may be described in terms of p (equation 6), requiring the estimation of a further two parameters (a and b) when equation 4a is appropriate or three (a, b_1, and b_2) when equation 4b is appropriate. Thus, a total of only five or six parameters may need to be estimated to describe temporal and spatial variation in disease incidence during an epidemic in terms of the beta-binomial distribution.

Statistical descriptions of spatial pattern do not by themselves provide an explanation of the mechanisms responsible for the pattern (3,7). They do, however, have some important uses. One is in identifying the appropriate statistical model to use for the analysis of data, and another is in the design of sampling procedures. The use of the beta-binomial distribution in these applications will be discussed in future reports.

LITERATURE CITED

1. Anderson, D. A. 1988. Some models for overdispersed binomial data. Aust. J. Stat. 30:125-148.
2. Bald, J. G. 1937. Investigations on "spotted wilt" of tomatoes. III.

TABLE 3. Beta-binomial parameter estimates and goodness-of-fit statistics for the maize dwarf mosaic virus epidemic in maize field P-3 reported by Madden et al (12)

Disease assessment	Day	Parameter estimates				Goodness-of-fit[a]					
						Binomial[b]			Beta-binomial		
		$\hat{p}$	SE	$\hat{\theta}$	SE	χ^2	df	P	χ^2	df	P
1	14	0.042	0.0051	0.017	0.0065	5.2	3	0.16	0.8	3	0.84
2	21	0.252	0.0191	0.066	0.0180	25.1	3	<0.001	5.2	3	0.16
3	28	0.431	0.0156	0.026	0.0087	21.3	3	<0.001	3.4	3	0.33
4	35	0.586	0.0134	0.017	0.0064	12.5	4	0.014	3.2	3	0.37
5	42	0.790	0.0092	0.009	0.0044	8.2	3	0.041	1.9	3	0.60
6	56	0.871	0.0081	0.011	0.0051	9.2	3	0.026	2.1	3	0.55

[a] Degrees of freedom (df) were determined by pooling frequency classes so that expected frequencies were >5.
[b] Calculated from equation 1 with $\pi = \hat{p}$.

Infection in field plots. Commonw. Aust. Counc. Sci. Ind. Res. Bull. 6.
3. Campbell, C. L., and Madden, L. V. 1990. Introduction to Plant Disease Epidemiology. John Wiley & Sons, Inc., New York. 532 pp.
4. Campbell, C. L., and Noe, J. P. 1985. The spatial analysis of soilborne pathogens and root diseases. Annu. Rev. Phytopathol. 23:129-148.
5. Cochran, W. G. 1936. The statistical analysis of field counts of diseased plants. Suppl. J. Roy. Stat. Soc. 3:49-67.
6. Cox, D. R., and Snell, E. J. 1989. Analysis of Binary Data. 2nd ed. Chapman & Hall, London. 236 pp.
7. Gilligan, C. A. 1982. Statistical analysis of the spatial pattern of *Botrytis fabae* on *Vicia faba*: A methodological study. Trans. Br. Mycol. Soc. 79:193-200.
8. Griffiths, D. A. 1973. Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. Biometrics 29:637-648.
9. Hughes, G., and Madden, L. V. 1992. Aggregation and incidence of disease. Plant Pathol. 41:657-660.
10. Madden, L. V. 1989. Dynamic nature of within-field disease and pathogen distributions. Pages 96-126 in: Spatial Components of Plant Disease Epidemics. M. J. Jeger, ed. Prentice Hall, Englewood, NJ.
11. Madden, L. V., and Campbell, C. L. 1990. Nonlinear disease progress curves. Pages 181-229 in: Epidemics of Plant Diseases: Mathematical Analysis and Modeling. 2nd ed. J. Kranz, ed. Springer Verlag, Berlin.
12. Madden, L. V., Louie, R., and Knoke, J. K. 1987. Temporal and spatial analysis of maize dwarf mosaic epidemics. Phytopathology 77:148-156.
13. Madden, L. V., Pirone, T. P., and Raccah, B. 1987. Analysis of spatial patterns of virus-diseased tobacco plants. Phytopathology 77:1409-1417.
14. Marcus, R., Fishman, S., Talpaz, H., Salomon, R., and Bar-Joseph, M. 1984. On the spatial distribution of citrus tristeza virus disease. Phytoparasitica 12:45-52.
15. Moran, P. A. P. 1968. An Introduction to Probability Theory. Oxford Univ. Pr., London. 542 pp.
16. Olkin I., Gleser L. J., and Derman, C. 1980. Probability Models and Applications. Macmillan Publishing Co., Inc., New York. 576 pp.
17. Patil, G. P., and Joshi, S. W., eds. 1968. A Dictionary and Bibliography of Discrete Distributions. Oliver and Boyd, Edinburgh. 268 pp.
18. Paul, S. R. 1979. A clumped beta-binomial model for the analysis of clustered attribute data. Biometrics 35:821-824.
19. Rouse, D. I., MacKenzie, D. R., Nelson, R. R., and Elliot, V. J. 1981. Distribution of wheat powdery mildew incidence in field plots and relationship to disease severity. Phytopathology 71:1015-1020.
20. Shaw, M. W., and Royle, D. J. 1987. Spatial distribution of *Septoria nodorum* and *S. tritici* within crops of winter wheat. Plant Pathol. 36:84-94.
21. Skellam, J. G. 1948. A probability distribution derived from the binomial distribution by regarding the probability of success as variable between sets of trials. J. Roy. Stat. Soc. B 10:257-261.
22. Smith, D. M. 1983. Maximum likelihood estimation of the parameters of the beta binomial distribution. Appl. Stat. 32:192-204.
23. Strandberg, J. 1973. Spatial distribution of cabbage black rot and the estimation of diseased plant populations. Phytopathology 63:998-1003.
24. Waggoner, P. E., and Rich, S. 1981. Lesion distribution, multiple infection, and the logistic increase of plant disease. Proc. Natl. Acad. Sci. USA 78:3292-3295.