Sequential Sampling for Incidence of *Phomopsis* Leaf Blight of Strawberry

W. W. Turechek, M. A. Ellis, and L. V. Madden

First author: Department of Plant Pathology, Cornell University, New York State Agricultural Experiment Station, Geneva 14456; and second and third authors: Department of Plant Pathology, The Ohio State University, Ohio Agricultural Research Development Center, Wooster 44691.

Accepted for publication 13 December 2000.

ABSTRACT

Turechek, W. W., Ellis, M. A., and Madden, L. V. 2001. Sequential sampling for incidence of *Phomopsis* leaf blight of strawberry. *Phytopathology* 91:336-347.

Sequential sampling models for estimation and classification were developed for the incidence of strawberry leaflets infected by *Phomopsis obscurans*. Sampling protocols were based on a binary power law analysis of the spatial heterogeneity of *Phomopsis* leaf blight in commercial fields in Ohio. For sequential estimation, samples were collected until mean disease incidence could be estimated with a preselected coefficient of variation of the mean (C). For sequential classification, samples were collected until there was sufficient evidence to classify mean incidence as being below or above a threshold (p_t) based on the sequential probability ratio test. Monte-Carlo simulations were used to determine the theoretical average sample number (ASN) and probability of classifying mean incidence as less than p_t (operating characteristic) for any true

value of incidence. Estimation and classification sampling models were both tested with bootstrap simulations of randomly selected data sets and validated by data sets from another year that were not utilized in developing the models. In general, achieved (or calculated) C after sequentially sampling for estimation was close to the preselected C of 0.2, and mean incidence was estimated with little bias. Achieving a C of 0.1 with less than 75 sampling units (the nominal value for many original data sets) was more problematical, especially with true incidence less than 0.2. ASN for classification was only 9 to 18 at disease incidence values near p_t , and approximately five or less at incidence values far from p_t . Correct classification decisions were made in over 88% of the validation data sets. Results indicated that it is possible to estimate *Phomopsis* leaf blight with high precision and with high correct classification probabilities.

Additional keywords: beta-binomial distribution, integrated pest management, theoretical epidemiology.

Strawberry leaf blight, caused by *Phomopsis obscurans*, is a common foliar disease of strawberry in Ohio and in many parts of the United States. Most of the commonly grown cultivars are susceptible to the disease (5–7). Significant yield losses are possible when disease incidence is high due to a reduction in the photosynthetic capacity of the plant. It is speculated that high foliar disease intensity could lead to a decrease in the winter hardiness of the plant, potentially reducing yield of the following years' crop (27). The disease is primarily controlled through the application of pesticides, typically applied on a calendar-based schedule as part of a comprehensive disease management program (25,26). However, intensity of the disease varies greatly within and among farms and from year to year in Ohio (27), indicating that routine fungicide spraying is not needed. Scheduling of fungicide applications can be based on weather conditions and information on mean disease intensity at a particular time (13). At present, economic thresholds and economic injury levels (23) for *Phomopsis* leaf blight have not yet been determined. However, any use of disease thresholds fully depends on being able to determine if mean disease intensity is above or below a hypothesized threshold based on observations of a reasonably low number of sampling units.

Until recently, very little was known about the incidence and distribution of *Phomopsis* leaf blight in commercial farms with perennial strawberry production (27). Recent investigations have shown that the disease is spatially aggregated in relatively small patches within fields and that the degree of aggregation is primarily

dependent on mean incidence (27). It is still not known which environmental and crop production factors are associated with high disease incidence. To relate these factors to disease development under natural conditions requires that sampling of commercial fields be undertaken to precisely estimate mean disease incidence with a small number of samples.

If the spatial pattern of a disease is known, or can be assumed based on prior information, then a sequential sampling plan can be designed (1,16,21). Two types of sequential sampling plans are sequential estimation and sequential classification. Sequential estimation is a procedure in which sampling units are collected, one at a time, and the number of diseased individuals are tallied for each unit until sufficient numbers of diseased individuals are collected so one can estimate the mean with a preselected level of precision (18). A common measure of precision is the coefficient of variation of the mean (C), although the standard error (SE) can also be used. With sequential sampling for classification (or hypothesis testing or decision making), sampling progresses as in sequential estimation. However, it differs in that sampling ends when sufficient numbers of individuals are collected so one can state that disease incidence is either significantly higher or lower than a user-defined disease threshold (the hypothesized value) (1,16). The classification is made with preselected error-rate probabilities. The precision of the estimated mean is not controlled at any specific level, but depends on how far the mean is from the hypothesized value.

Ideally, a sampling strategy should be validated before it is used in practice (11). Validation refers to the "quantitative comparison of output from the (sampling) model with empirical results (data) not used in the development of the model" (2). Usually, however, all data sets collected are used in the development of the sampling plan, and statistical bootstrapping or some other procedure is performed as a partial validation of the plan (18,21,28). It is impor-

Corresponding author: L. V. Madden; E-mail address: madden.1@osu.edu

Publication no. P-2001-0112-01R

© 2001 The American Phytopathological Society

tant to note that validation is a statistical procedure used, in this context, to determine if the sampling model performed as intended. Thus, for sequential estimation, validation primarily determines if the estimate of mean incidence is as precise and accurate as intended (2). For sequential classification, validation determines whether a field was classified correctly, in terms of mean incidence, with the prespecified error probability. Validation does not guarantee that the practical use of the sampling model will result in adequate disease control, especially if an improper disease threshold was used (22). If the application of a sampling model does not meet expectations, then model parameters need to be adjusted. This may require additional sampling.

In this paper, we use previously determined results (27) on the relationship between mean incidence and spatial heterogeneity of *Phomopsis* leaf blight to develop sequential sampling models or plans for (i) estimating mean incidence of strawberry leaf blight with a predetermined level of precision; and (ii) classifying mean incidence above or below various disease threshold values with predetermined error probabilities. We used bootstrapping to test the sequential sampling plans with data sets used in developing the plans, and more fully validated the sampling plans with independent data sets.

MATERIALS AND METHODS

Data collection. *Model construction data set.* Naturally occurring epidemics of strawberry leaf blight were monitored at two commercial strawberry farms in 1996 and 1997. Farm one was located near Wooster, OH (40°49'06"N, 81°55'58"W) and farm two was located approximately 151 km southwest of Wooster in Darbydale, OH (39°51'13"N, 83°10'59"W). Each farm had several perennial plantings (fields) comprised of 10 to 20 rows of strawberry. Three or four fields were randomly selected for disease assessment from a total of eight or nine fields at each sampling date over 2 years. Not all fields were sampled the same number of times. The same leaves were not necessarily sampled in different assessments of the same field. In total, 121 data sets were collected over the 2 years, and information obtained from the spatial analysis of this data was used for the construction of the sequential sampling plans. Additional details for the model construction data are given in Turechek and Madden (27).

Validation data set. In 1998, a statewide survey of *Phomopsis* leaf blight of strawberry was conducted from 16 June through 23 June. This time was chosen to coincide with harvesting. In total, 26 farms were sampled; farms selected for the survey were chosen in a stratified random sample. In this survey, the state was divided into nine regions or strata (Fig. 1 in literature citation 28), and commercial strawberry farms within these regions were identified from the member directories of the National Association of Strawberry Growers and the Ohio Small Fruit Growers Association. Additional farms were identified through personal communications with growers. From this listing, three or four farms were randomly chosen from each of the nine regions or strata (except in regions eight and nine where only two and one strawberry farm could be identified, respectively). For each farm, three fields were randomly chosen from the fields identified for disease assessment. All three fields on the 26 farms surveyed had leaflets with leaf blight. In total, 78 independent data sets from 1998 were used as the validation data.

Sampling. Fields were sampled almost identically over the 3 years. For each field and assessment time, a single row was randomly chosen as the primary transect. In 1996 and 1997, a short secondary transect was arbitrarily chosen to approximately bisect the field; this transect was not included in 1998 samplings. Sampling units consisted of 15 leaflets (five leaves with three leaflets each). Each leaflet was rated for the presence or absence of leaf blight symptoms. The initial sampling unit was indiscriminately chosen within the first 20 steps of the primary transect, and each

subsequent sampling unit was located at a distance of three steps (≈ 2.5 m) from the previous sampling unit, assuring approximately evenly spaced sampling units. This is a form of cluster sampling, with clusters being equivalent to sampling units (10). The number of sampling units, N , varied from 27 to 87, depending upon the field size, but was typically between 65 and 75.

Spatial pattern analysis. A spatial pattern analysis for the model construction data set has been published (27). Because the same methodology was applied to the validation data set (1998), only a brief description of the methods used is given here.

Distribution analysis. To assess the degree of heterogeneity, the beta-binomial and the binomial distributions were fitted to the data with the program BBD (14) for each individual field assessment. The binomial distribution has a single parameter representing the probability of selecting a diseased leaflet (π). The beta-binomial distribution has two parameters, p , which is the expected probability of selecting a diseased leaflet, and θ , which is a measure of the variation (heterogeneity) in disease incidence among the sampling units. A good fit of the binomial distribution to the data is suggestive of a random spatial pattern of disease incidence, whereas a good fit of the beta-binomial distribution is suggestive of an aggregated pattern of disease incidence at the spatial scale of the sampling unit or smaller (15). At $\theta = 0$, the beta-binomial reduces to the binomial distribution. A log-likelihood ratio test statistic (LRS) was calculated for each data set to determine if the beta-binomial fitted the data better than the binomial. LRS has a χ^2 distribution with 1 degree of freedom under the null hypothesis of no difference between the two log-likelihoods (i.e., when the beta-binomial does not characterize the frequency distribution of diseased leaflets better than the binomial) (31).

Binary power law analysis. The binary form of Taylor's power law was used to assess the degree of heterogeneity for a collection of data sets. The binary power law can be expressed as

$$\ln(v_{\text{obs}}) = \ln(A_x) + b \ln[np(1-p)] \quad (1)$$

in which v_{obs} is observed sample variance, $np(1-p)$ is the theoretical variance for the binomial distribution (i.e., the binomial variance), n is the number of individuals per sampling unit ($n = 15$ leaflets), and $\ln(A_x)$ and b are the intercept and slope of a straight line (15,27). In fitting equation 1 to the data by regression analysis, we used estimated p from the beta-binomial distribution ($\hat{p}$) and the moment estimate of v_{obs} for each data set. Use of the observed proportion of diseased leaflets for p in equation 1 had no effect on the regression results (L. V. Madden, unpublished data). When A_x and b are both equal to 1, a random distribution of diseased leaflets is indicated across all data sets. When $b = 1$ and $A_x > 1$, spatial heterogeneity or overdispersion is constant and nonzero over the range of p . When b and A_x are both greater than 1, overdispersion or heterogeneity as measured by θ of the beta-binomial distribution changes systematically with p (15).

Sequential sampling curves. *Sequential estimation.* In order to estimate incidence with a fixed level of precision, sequential estimation sampling curves were developed using the results of the binary power law analysis for the model construction data sets. In our situation, precision was determined with C , defined as the SE of estimated p ($\hat{p}$) divided by $\hat{p}$ (10,17). When heterogeneity of incidence is adequately described with the binary power law, the SE of $\hat{p}$ can be written as

$$\sqrt{a[\hat{p}(1-\hat{p})]^{b-1}/N}$$

in which $a = A_x n^{b-2}$ (10). In practice, estimates of a and b ($\hat{a}$ and $\hat{b}$) from a regression analysis are used in the formula for the SE.

Sequential estimation begins by tallying the total number of diseased leaflets, T_N , as each sampling unit is collected. If x_i is the number of diseased leaflets in the i^{th} sampling unit, then $T_N = \sum x_i$

is the tally of the total (cumulative) number of diseased leaflets at the N^{th} unit ($i = 1, \dots, N$). When T_N exceeds some threshold (critical) value, called the “stop limit”, determined by C , n , N , and the binary power law parameters (a and b) in our case, sampling ceases; p can then be easily determined by $\hat{p} = T_N/(nN) = \Sigma x_i/(nN)$. When $b > 1$ in the binary power law, one cannot write T_N as a simple function of a , b , n , N , and C (18). However, it can be written as function of T_N (γ_N), which can then be solved numerically for T_N . The function is written as

$$\gamma_N = T_N^{b-2} (nN - T_N)^b = (C^2 / a) n^{2b-2} N^{2b-1} \quad (2)$$

A Mathcad (Mathsoft Inc., Cambridge, MA) worksheet was written to iteratively solve the γ_N equation for T_N at $N = 1$ to 500 and $C = 0.1$ and 0.2 .

Sequential classification. The binary power law results for the model construction data were also used to develop a sequential classification sampling plan. A common approach for classification, developed by Wald (30) is the sequential probability ratio test (SPRT). Detailed reviews of this general methodology can be found in Binns (1) and Nyrop and Binns (21). Madden and Hughes (16) showed how to use the SPRT for overdispersed (aggregated) disease incidence data collected as a cluster sample. Only a brief overview is given here.

We begin with a critical or threshold value, p_r , which represents some level of disease deemed important for hypothesis testing, and possibly tied to some action, such as initiating a control procedure. The null hypothesis of interest is $H_0: p \leq p_r$, versus the alternative hypothesis $H_1: p > p_r$. A type I error occurs when one rejects the null hypothesis when it is true (that is, one falsely declares that $p > p_r$); a type II error occurs when one accepts the null hypothesis when it is false (that is, one falsely declares that $p \leq p_r$).

To evaluate a hypothesis, the probability of these errors at any value of p can be summarized in the operating characteristic (OC). OC is defined as the probability of accepting the null hypothesis, given the true value of p . Below p_r , the OC is the probability of correctly accepting the null hypothesis (1-type I error), and above p_r , the OC is the probability of incorrectly accepting the null hypothesis (type II error). Ideally, the OC is 1 when $p \leq p_r$, and 0 when $p > p_r$; however, this is never achieved. Another means of evaluating a hypothesis test applicable for sequential sampling is a plot of the average sample number (ASN) versus p . The ASN is the mean number of sampling units (of n individuals each for cluster sampling) required to accept or reject the null hypothesis for any true value of p . The OC and ASN provide the average (expected) properties of the test over a large number of sampling bouts.

Testing a hypothesis in sequential sampling for classification requires one to choose from three possible decisions at sample size N : (i) reject the null hypothesis; (ii) accept the null hypothesis; or (iii) continue sampling because not enough information has been collected to make either decision one or two. To accomplish this with the SPRT requires that the composite hypotheses above be converted into simple hypotheses. The new null hypothesis is $H_0': p = p_0$, and the new alternative hypothesis is $H_1': p = p_1$, where p_0 and p_1 are new constants such that $p_0 < p_r$ and $p_1 = (p_0 + p_r)/2$. Accepting H_0' is interpreted as $p \leq p_r$, whereas accepting H_1' is interpreted as $p > p_r$.

Unlike sequential estimation, two stop limits or lines are calculated with sequential classification. If observed $T_N (= \Sigma x_i)$ is above the upper stop line, the alternative hypothesis is accepted; if the observed T_N is below the lower stop line, the null hypothesis is accepted. Otherwise, sampling continues. There is not an exact expression for the theoretical stop lines when the data have a beta-binomial distribution, with either fixed or variable θ (16). Approximation formulas for stop lines were recently developed by Madden and Hughes (16) for the beta-binomial distribution under

cluster sampling when the binary power law (equation 1) describes the variance in relation to p . The general expression is

$$i_0 + snN < T_N < i_1 + snN \quad (3)$$

where the expression on either side of T_N represent straight lines with common slope, sn , and intercepts of i_0 and i_1 . The line on the left of the inequality represents the lower boundary for T_N and the line on the right of the inequality represents the upper boundary for T_N . The intercept terms depend on the degree of overdispersion, p_0 , p_1 , and parameters related to type I (α) and type II (β) errors. Formulas for s , i_0 , and i_1 are provided in the Appendix.

Sequential sampling procedures were designed for four levels of p_r : $p_r = 0.1$ ($p_0 = 0.05$, $p_1 = 0.15$); $p_r = 0.15$ ($p_0 = 0.05$, $p_1 = 0.25$); $p_r = 0.2$ ($p_0 = 0.1$, $p_1 = 0.3$); and $p_r = 0.25$ ($p_0 = 0.1$, $p_1 = 0.4$), at each of the four combinations of α (0.05 or 0.10) and β (0.05 and 0.10). In total, 16 sequential sampling plans were designed and evaluated by comparing OC and ASN values. Although formulas for OC and ASN are known for some statistical distributions if several assumptions are met, it is generally considered more accurate to use Monte-Carlo simulations (19) to determine these statistics for selected values of p_0 , p_1 , α , and β (21). Monte-Carlo method is described in the Appendix. From the simulation results, we selected four cases for further evaluation with field data. The parameters for the four selected cases were $p_r = 0.1$ with $\alpha = \beta = 0.05$ or 0.1 , and $p_r = 0.2$ with $\alpha = \beta = 0.05$ or 0.1 .

Sampling evaluation. Sequential estimation–bootstrapping. As a partial validation of the sequential estimation model, sequential sampling was simulated (mimicked) from a subset of the 121 model construction (1996 and 1997) data sets through statistical bootstrap methods (4,19). The 121 data sets collected during 1996 to 1997 were divided into six categories based on their mean incidence of leaf blight. The following categories were constructed: $0.01 \leq \hat{p} < 0.05$; $0.05 \leq \hat{p} < 0.1$; $0.1 \leq \hat{p} < 0.2$; $0.2 \leq \hat{p} < 0.4$; $0.4 \leq \hat{p} < 0.6$; and $\hat{p} \geq 0.6$. From each category, three data sets were randomly chosen. The predicted stop limit curve (i.e., values of T_N) was determined with equation 2 with chosen C and values of a and b estimated for the pooled model construction data. The bootstrap simulation was done for both $C = 0.1$ and $C = 0.2$.

For a single bootstrap simulation with any data set, sampling units (of $n = 15$ leaflets each) were sampled randomly one at a time, with replacement, from a list of observed sampling units, and the cumulative number of diseased leaflets (Σx_i) was determined. Sampling ceased when the observed cumulative number of diseased leaflets exceeded the T_N prediction of the sampling model (equation 2) or the sample size equaled the size of the data set being sampled. An additional sampling rule imposed was that a minimum of three sampling units must be collected before sampling ceased. The moment estimate of p (i.e., achieved p [$\hat{p}$]), the SE of $\hat{p}$, and the achieved (calculated) C were calculated. A 95% confidence interval for p was constructed to determine if the interval included the assumed true p (equal to estimated p for the full data set). The bootstrap simulation was repeated for a total of 100 times for each of the 18 data sets and specified value of C . Because sampling is done with replacement, the bootstrap sample generally has the same statistical properties of the original data set but does not duplicate the original data. To summarize the results of the simulation, “box plots” and nonparametric confidence intervals based on the sign test (18) were constructed for achieved C , the difference between (assumed) true p of the data set and achieved p after sampling ($p - \hat{p}$), and achieved N for each of the 18 data sets. Methods were based on, and fully described in, Madden et al. (18) and Turechek and Madden (28).

Sequential estimation–validation data sets. Sequential estimation sampling of the validation data sets (from 1998) was performed with a Minitab macro (Minitab version 12, State College, PA) designed by the authors. In this procedure, the order of the

sampling units, as originally collected, was preserved. For each of the 78 data sets, sequential sampling began by tallying the total number of diseased leaflets in each sampling unit, beginning with the first and continuing until the stop limit was crossed. Stop limits (equation 2) used in this procedure were identical to the stop limits used in the bootstrap simulations. That is, the limits were based on the analysis of the model construction data and not the validation data. Each data set was sampled only once. For each data set, p was estimated from the number of sampling units collected in the sequential procedure and compared with p estimated from the entire transect (the true value of p). Also, achieved C was calculated, and the number of sampling units collected was recorded.

SPRT-bootstrap evaluation. Using the parameters (p_0 , p_1 , α , and β) from the four sampling designs selected after the Monte-Carlo evaluation, sequential classification sampling was simulated (mimicked) by statistical bootstrap methods as described previously. This was done for the six incidence categories previously identified, plus a seventh category, $\hat{p} < 0.01$. Thus, 21 data sets were assessed. The stop lines were generated by equation 3 with slope and intercept terms defined in equations A4 through A6. For a single bootstrap simulation with any data set, sampling units (of $n = 15$ leaflets each) were sampled randomly, one at a time, with replacement, from a list of observed sampling units, and the cumulative number of diseased leaflets was determined. Sampling ceased when the observed cumulative number of diseased leaflets crossed either the lower or upper stop lines or the sample size equaled the size of the data set being sampled. An additional sampling rule imposed was that a minimum of four sampling units must be collected before sampling ceased. The simulation was repeated a total of 100 times for each data set and specified values of the parameters.

SPRT-validation data sets. For the validation procedure, simulated sampling of the 1998 data progressed as with the sequential estimation validation, except sampling stopped when observed T_N crossed either the upper or the lower stop line or all sampling units were collected. Stop lines were based on either the binomial distribution (equations 3 and A1 to A3) or the beta-binomial approximation (equations 3 and A4 to A6). For each of the 78 data sets, the number of sampling units collected until a decision could be made (i.e., the stop line was crossed) was recorded. To determine if the correct decision was made, it was assumed that estimated p for the full transect represented the true value of p for that field. This true value of p was compared with the hypothesized p_i under evaluation to decide whether the null or alternative hypothesis was accepted. Then, the decision based on the sequential sample was compared with the assumed "correct" decision to determine the type I and II error rate.

RESULTS

Spatial heterogeneity summary. For the model construction data set, the beta-binomial distribution fit the data better than the binomial distribution in 92% of the 121 data sets based on LRS (27). The binary power law fitted the data well for both 1996 and 1997, and a covariance analysis indicated that the parameter estimates did not significantly differ from each other between 1996 and 1997; therefore, the data were pooled and the parameters re-estimated here to provide greater precision of the estimates. The pooled estimates of $\ln(A_x)$ and b were 0.91 (SE = 0.023) and 1.18 (SE = 0.015), respectively. With these parameters, A_x is estimated as $e^{0.91} = 2.48$, a is estimated as $2.48 \cdot 15^{1.18-2} = 0.27$, and A for the SPRT is estimated as $2.48/15^{2-2 \cdot 1.18} = 6.57$.

For the validation data of 1998, the beta-binomial fit the data better than the binomial distribution in 91% of 78 data sets based on a LRS. Mean disease incidence of leaflets ranged from 0.0027 to 0.81 (Fig. 1A), and the $\hat{\theta}$ s of the beta-binomial distribution ranged from 0 to 0.39 (Fig. 1B). The estimates of $\ln(A_x)$ and b

from the binary power law fitted to the 1998 data were 0.89 (SE = 0.035) and 1.20 (SE = 0.033), respectively (Fig. 1C). The validation data sets, therefore, were very similar to the model construction data sets in terms of the heterogeneity analysis.

Fig. 1. A and B, Frequency distribution of the estimated beta-binomial parameters $\hat{p}$ and $\hat{\theta}$, respectively, and C, relationship between the natural logarithm of the observed variance and the natural logarithm of the theoretical variance for a binomial distribution for incidence of *Phomopsis* leaf blight of strawberry. A and B, Vertical broken lines indicate the median (also shown numerically in the figure). C, The solid line represents the binomial (random) line, the dash-dot line represents the least squares fit to the validation data (equation 1), the circles are the individual observations for the validation data, and the broken line (barely distinguishable from the dash-dot line) represents the least squares fit to the model construction data of 1996 to 1997 (27). Validations data are from 78 commercial strawberry fields in Ohio in 1998.

Sequential estimation. The cumulative number of diseased (T_N ; the stop limit) versus the number of sampling units (N) based on equation 2, and the corresponding level of disease incidence at the critical T_N , are shown in Figure 2A and B, respectively. Curves were generated for the two C (0.2 and 0.1) values discussed previously and, for comparative purposes, $C = 0.3$. As desired precision increased or as incidence decreased, a greater number of diseased individuals must be found before sampling ceased (18).

Bootstrap evaluation. Figure 3A and B show box plots of the achieved (calculated) C values for the bootstrap evaluation of 18 data sets at two prespecified C values (0.1 and 0.2). As mean incidence increased, achieved C approached the nominal (prespecified) value and eventually fell below this value for $C = 0.2$ and, to a lesser extent, for $C = 0.1$. At low mean incidence (e.g., data sets one to three), the prespecified C could not be achieved even when the number of sampling units in the sequential procedure equaled the number of sampling units for the entire data set. For instance, N of the sequential sample equaled N of the full data set in all 100 bootstrap simulations for each of the first nine data sets when $C = 0.1$, and for the first three data sets (except for one bootstrap simulation replication in data set two) when $C = 0.2$ (Fig. 3E and F). Conversely, at high mean incidence (e.g., data sets 16 to 18), desired C (especially $C = 0.2$) could be achieved with less than the specified minimum number of sampling units ($N = 3$) in the sequential procedure (Fig. 3E and F). Thus, achieved C was less than the desired C with $N = 3$ sampling units.

Fig. 2. Sequential-estimation stop limits for estimating mean disease incidence ($\hat{p}$) of strawberry leaf blight with coefficient of variation of the mean (C) of 0.1, 0.2, and 0.3. **A**, Cumulative number of diseased leaflets (T_N) versus the number of sampling units (N) based on equation 2 with estimated b of 1.18 and estimated a of 0.27. **B**, Mean disease incidence at critical T_N (i.e., the point where the observed cumulative number of diseased leaflets crosses the model T_N curve in **A**). Mean disease incidence is estimated by $T_N/(nN)$ at the critical point.

For most data sets, achieved or estimated p ($\hat{p}$) based on sequential sampling was very close to the assumed true p (based on the estimated p for the full data set). For instance, the 95% confidence interval for the median of the difference between the true and achieved p included zero for both levels of desired C (Fig. 3C and D). At low $\hat{p}$, there was little difference in the estimated p and the true p values. This was primarily due to the large sample sizes for the sequential samples at low mean incidence values. It should be noted that because bootstrap simulation is based on sampling with replacement, even when N for a (simulated) sequential sample equals N for the full data set, the observed sample of diseased leaflets is not identical to the original population of diseased leaflets. Thus, $\hat{p}$ does not have to equal the true p in any bootstrap simulation case. The difference between the median $\hat{p}$ for the bootstrap simulations and the true p of each data set, as well as the variation among the bootstrap simulation replicates within each data set, increased as true p increased (Fig. 3C and D). The former confirms previous work showing that sequential sampling can result in slightly biased estimates of means (18,28).

The median or average achieved number of sampling units based on sequential sampling decreased as $\hat{p}$ increased for both levels of C , as required. Furthermore, achieved N was larger at $C = 0.1$ than at $C = 0.2$, especially for lower $\hat{p}$ (Fig. 3E and F). There was substantial variability in achieved N at both levels of C , with variability decreasing with increasing $\hat{p}$. Where comparisons were possible, that is, when N for the sequential sample was less than N for the entire data set for both tested values of C , variability in N was larger at $C = 0.1$ than at $C = 0.2$ (Fig. 3E and F).

Validation. Across the 78 validation data sets, median value for the difference between true p and estimated p with sequential sampling ($p - \hat{p}$) was nearly zero for both desired values of C (Fig. 4A and B). In fact, the difference between the 75th and 25th percentiles of $p - \hat{p}$ (the inter-quartile range) for this collection of data sets was less than 0.05 for both $C = 0.2$ and $C = 0.1$, indicating that estimates of p were accurate for this sequential estimation procedure. However, for low levels of mean incidence ($\hat{p} < 0.1$) and desired $C = 0.1$, this apparent high accuracy gives a somewhat misleading impression of the overall performance of the procedure. This is because the entire transect was sampled, resulting in a $\hat{p}$ equaling the true p (by definition). For the validation, bootstrapping was not used, so each observation in a data set could only be sampled once. The precision of the estimate of p was less than desired because sampling was forced to stop before the desired C of 0.1 was reached. This effect could be seen by the median value of achieved C (across all 78 data sets) equaling 0.15 when preselected or desired C was 0.1 (Fig. 4D). When desired precision was lower ($C = 0.2$), median-achieved C equaled desired C to two decimal places (Fig. 4C). Even at this desired level of precision, the entire transect sometimes was sampled according to the sequential procedure for data sets with small $\hat{p}$.

The number of sampling units collected before sampling ceased decreased as the “true” incidence increased (Fig. 4E and F). The rate of decrease in N was greater for $C = 0.2$ than for $C = 0.1$. As required, fewer sampling units were needed before sampling ceased for $C = 0.2$ than for $C = 0.1$. As with the bootstrap results, the degree of variability in N also was greater for $C = 0.1$ than for $C = 0.2$. The theoretical expected N (calculated with equation 17b of Madden and Hughes literature citation 16), using a and b estimates of the binary power law analysis (equation 1), was in general agreement with achieved N values, except at low p ($p \leq 0.2$ for $C = 0.1$ and $p \leq 0.05$ for $C = 0.2$) (Fig. 4E and F). As previously described, the disagreement in N values at low p was due to data sets with fewer sampling units than required to obtain the desired level of precision.

Sequential classification. Pairs of stop lines based on equation 3 (with slope and intercept terms given by equations A4 to A6) for the SPRT are shown in Figure 5A for a single p_i at different α and β values and in Figure 6A for four different p_i values at fixed α

and β . Changing α or β between 0.05 and 0.10 had small effects on the height of the stop lines (and difference between the upper and lower lines in each pair) (Fig. 5A). Equivalently, the height of the stop lines also could be controlled by altering p_0 and p_1 for a fixed p_i (e.g., $p_0 = 0.1$ and $p_1 = 0.3$, or $p_0 = 0.17$ and $p_1 = 0.23$, both for $p_i = 0.2$), but this is not shown here. The slope of the stop lines was directly proportional to p_i , by definition (16).

As expected (1), beta-binomial based Monte-Carlo simulation (described in Appendix) showed that decreasing the α and β parameters resulted in a greater correct decision rate; that is, there was a somewhat higher OC curve at $p < p_i$, a somewhat lower OC at $p > p_i$, and a steeper OC curve in the vicinity of p_i (Fig. 5B). The increased accuracy that occurred with the lowering of α and β was, however, associated with an increase in ASN, especially between p_0 and p_1 (Fig. 5C). Changing only α or β had a smaller effect on ASN and OC than the simultaneous change in both (W. W. Turechek, unpublished data).

Changing the critical value (p_i) of the SPRT had a large influence on the position of the OC curve and height of the ASN curve (Fig. 6B and C), as also shown by Nyrop et al. (22) for an Iwao-type (1) sequential sampling model. Increasing p_i resulted in a shift of the OC to the right as well as a slight flattening of the OC.

The flattening indicated a slight reduction in correct decision rates and was caused by the generally increasing difference between the chosen p_1 and p_0 parameter values with increasing p_i in these specific tests of the SPRT. Moreover, ASN decreased with increasing p_i , especially between p_i of 0.10 and 0.15 (Fig. 6C), as the OC simultaneously became flatter. If the difference between p_1 and p_0 was held fixed, however, the ASN increased in the vicinity of p_i , and the OC became steeper, with increasing p_i (L. V. Madden, unpublished data).

Bootstrap evaluation. Based on the Monte-Carlo results, the following four conditions were tested more extensively with bootstrap simulation: $p_i = 0.1$ with $\alpha = \beta = 0.05$; $p_i = 0.1$ with $\alpha = \beta = 0.1$; $p_i = 0.2$ with $\alpha = \beta = 0.05$; and $p_i = 0.2$ with $\alpha = \beta = 0.1$. Achieved OC curves based on the 21 data sets randomly selected from the model construction data agreed with the theoretical OC curves derived from Monte-Carlo simulations at the two p_i values and two α/β pairs (Fig. 7A and B). When $p_i = 0.1$, achieved ASNs followed closely to those obtained via Monte-Carlo simulation for both α/β choices (Fig. 7C). When $p_i = 0.2$, achieved ASNs were also close to the Monte-Carlo derived ASNs except for one field (Fig. 7D). This field had a $\hat{p}$ (0.202) very close to the p_i being tested. For this single data set, the ASNs from bootstrapping were

Fig. 3. Box plots of: **A and B**, the achieved (calculated) coefficient of variation of the mean disease incidence (C); **C and D**, the difference of the “true” value of disease incidence (p) and estimated p from the sequential sample ($\hat{p}$); and **E and F**, the achieved sample size (N) from 100 bootstrap simulation samplings of each of 18 *Phomopsis* leaf blight data sets. Horizontal axis refer to the numerical labels for the data sets. Stop limits were based on the use of equation 2. Data sets are arranged in increasing order of $\hat{p}$, so that data sets 1 through 3 correspond to $0.01 \leq \hat{p} < 0.05$, sets 4 to 6 correspond to $0.05 \leq \hat{p} < 0.1$, sets 7 to 9 correspond to $0.1 \leq \hat{p} < 0.2$, sets 10 to 12 correspond to $0.2 \leq \hat{p} < 0.4$, sets 13 to 15 correspond to $0.4 \leq \hat{p} < 0.6$, and sets 16 to 18 correspond to $\hat{p} \geq 0.6$. Preselected values of C for **A, C, and E**, 0.2 and **B, D, and F**, 0.1. Box plots show the median, middle 50% of data (open box), 95% confidence interval for median (solid bar inside box, based on sign test), “whiskers” (vertical straight lines) that extend to highest and lowest observations that are still within the region defined by lower and upper limits, and outliers (solid circles) beyond the whiskers. Lower limit is defined as $Q_1 - 1.5 \cdot (Q_3 - Q_1)$, and upper limit is defined as $Q_1 + 1.5 \cdot (Q_3 - Q_1)$, where Q_1 and Q_3 are the first and third quartiles of the data (of either $p - \hat{p}$, C , or N). The 18 data sets are a randomly chosen subset, within each of six incidence categories, from 121 independent assessments of leaf blight epidemics observed in commercial strawberry fields in Ohio in 1996 and 1997. “True” p is equated to the estimated p from the full data set.

considerably higher than those obtained via Monte-Carlo simulation. This was because the $\hat{\theta}$ of the beta-binomial distribution for the field (0.071) was less than that predicted by the binary power law (0.158; described in Appendix).

Validation. Table 1 shows the results for the evaluation of the four selected SPRT sampling protocols with the 1998 validation data. Correct and incorrect decision rates were determined across all data sets for both the binomial (equations 3 and A1 through

A3) and the beta-binomial stop lines, with heterogeneity defined by the binary power law (equations 3 and A4 through A6). As expected, the beta-binomial stop lines resulted in higher correct decision rates than those obtained for the binomial-based lines. Greater than 92% of the decisions were correct for $p_t = 0.1$ at both α/β pairs with the beta-binomial based stop lines. Correct decisions were made in 88.5% of the data sets for $p_t = 0.2$ at both error parameter pairs with the beta-binomial based lines. Of the incor-

Fig. 4. **A and B,** Frequency distribution of the difference of the assumed “true” value of disease incidence (p ; expected probability of a leaflet being diseased), and estimated p from the sequential sample ($\hat{p}$); **C and D,** the achieved coefficient of variation of the mean (C); and **E and F,** the relationship between number of sampling units (N) collected and p , based on sequential sampling for estimation of mean *Phomopsis* leaf blight incidence of strawberry in 78 data sets in 1998. “True” p is equated to the estimated p from the full data set. Preselected values of C are **A, C, and E,** 0.2 and **B and D, and F,** 0.1. **A through D,** The vertical solid line represents the median of the 78 validation data sets, and the broken lines represent the first and third quartiles of the data sets. The difference between the two broken lines is the interquartile range (middle 50% of the data sets). **E and F,** The solid line represents expected N given the true level of incidence (equation 4; equation 17b in literature citation 16). The data sets in this figure were not used for developing the sampling model.

rect decisions, type I errors were made more frequently than type II errors for both p_i values, although α and β parameters were set as equal in this specific evaluation.

Mean N , i.e., the average sample size over the 78 data sets (different ps), which is not to be confused with ASN, the average sample size at a given p , was slightly lower for the binomial than for the beta-binomial stop lines (Table 1). Mean N was also lower for $\alpha = \beta = 0.1$ than for $\alpha = \beta = 0.05$. For the binomial lines, over 75% of the N s were equal to 4, the prespecified minimum number of sampling units for the sequential sampling bouts (data not shown). The lower number of sampling units, however, resulted in somewhat lower correct decision rates (Table 1).

DISCUSSION

Sequential sampling is a method that allows one to obtain information about a population with minimal sampling effort (21). Depending on the goals of the investigator, sequential sampling can be performed either to precisely estimate the mean or to test a hypothesis about the mean (1,11,21,22). Both sampling goals require that prior information be available on either the discrete statistical distribution of the data (including the value of the heterogeneity parameter of the distribution) or the relationship between the mean and the variance. These statistical properties of data are functions of the spatial heterogeneity of the observations (2,15). With previously determined results characterizing the spatial heterogeneity of the incidence of *Phomopsis* leaf blight of strawberries in commercial fields (27), we developed sequential sampling plans for both estimation and hypothesis testing (classification) using the formulas derived by Madden et al. (18) and Madden and Hughes (16). The foundation for both of these sampling models was the binary power law relationship between the variance and mean disease incidence and the formal link between the binary power law and the beta-binomial distribution for aggregated or overdispersed disease incidence data (15). Bootstrapping (4,19) revealed that the sampling plans generally performed well on data with the same statistical properties as the original model construction data (i.e., the bootstrap simulation data). Moreover, validation testing showed that the sampling models could be used reliably with data not used in model construction. This study represents the first independent testing of the sequential sampling model proposed by Madden and Hughes (16) for classifying the mean of overdispersed disease incidence data, and the most comprehensive evaluation, to date, of the model developed by Madden et al. (18) for estimating mean incidence through sequential sampling.

Although the development and validation of sequential sampling procedures is a time consuming and statistically rigorous task, the implementation of these designs is not. To implement either the sequential estimation or classification procedure, one begins the sampling bout by randomly choosing a strawberry row and indiscriminately selecting the first sampling unit ($n = 15$ leaflets) within the first 15 m of the row. Subsequent sampling units are approximately every 2.5 m (i.e., three steps) down the row. The total number of sampling units observed depends on the incidence of disease and the parameters of the sampling model. For sequential estimation, one determines the cumulative number of diseased leaflets (T_N) after each sampling unit is observed and compares this with the T_N prediction (i.e., stop limit) of equation 2 calculated for the selected level of precision (C) (Fig. 2A). Sampling ends when observed T_N exceeds the stop limit of Figure 2A; estimated p is then given by $T_N / (nN)$. The same general sampling protocol is implemented for sequential classification, except that one compares the observed T_N with a pair of stop lines (Fig. 6) determined by equation 3, for the critical or threshold level of disease incidence (p_i) and the model parameters (p_0 , p_1 , α , β). Sampling stops when observed T_N falls either above or below the pair of lines. When observed T_N falls below the stop line, this is

interpreted as disease incidence being less than p_i ; when observed T_N falls above the stop line, this is interpreted as disease incidence being greater than p_i .

Except for low disease incidence ($\hat{p} \leq 0.05$), it was possible to estimate the mean with a C of 0.2 with (substantially) fewer sampling units than for the complete data sets (Figs. 3 and 4). Although sequential sampling can result in biased estimates of the

Fig. 5. A, Stop lines (cumulative number of diseased individuals [T_N] versus number of sampling units [N]), B, operating characteristic (OC) curves, and C, average sample number (ASN) curves for a sequential probability ratio test of disease incidence from a sequential cluster sample (of $n = 15$ individuals per sampling unit), with parameters $p_0 = 0.05$ and $p_1 = 0.15$ and four combinations of α and β parameters. Line symbols are consistent for all sub-figures, but the labels are only shown in A. Stop lines were determined by equation 3 with slope and intercept terms given in equations A4 through A6. OC and ASN curves were determined via Monte-Carlo simulation of beta-binomial variables, as described in the Appendix. Estimates of a , A , and b are 0.27, 6.57, and 1.18, respectively. B and C, Vertical lines represent p_0 and p_1 .

mean (18), there was only slight evidence of bias for most data sets, either with the bootstrap or validation testing. Achieving a C of 0.1 was more problematical than a C of 0.2. For $\hat{p} \leq 0.2$, it was generally not possible to obtain a C of 0.1 with $N < 75$ (Fig. 4F), the nominal number of sampling units of the original data. This

difficulty was not a property of the sequential sampling method per se; rather, this reflected the relatively high heterogeneity of the observed data. For instance, required N for a given prespecified C can be predicted (10) for nonsequential sampling based on the estimated binary power law parameters (equation 1) as

$$N = \hat{a}\hat{p}^{\hat{b}-2}(1-\hat{p})^{\hat{b}}C^{-2} \quad (4)$$

For $C = 0.1$, N is predicted as 1,163, 296, 157, and 77 for $\hat{p}$ values of 0.01, 0.05, 0.10, and 0.20, respectively. In practice, 75 sampling units collected three steps apart usually allowed one to traverse an entire row in a strawberry field in approximately 1 h. To obtain much larger N s would require sampling from more rows or more intensely within individual rows. It is doubtful that an investigator typically will be willing to invest the time or labor resources to obtain this very precise estimate of mean incidence for Phomopsis leaf blight of strawberry. However, a C of 0.2 is a reasonable level of precision for most purposes (2,16,17).

In addition to the magnitude of N , variation in N at a given $\hat{p}$ was also higher with a preselected C of 0.1 than with a C of 0.2 (Figs. 3 and 4). This can be understood by calculating an approximate large-sample variance of N [$V(N)$] when the regression estimates of the binary power law (equation 1) are used in equation 4 (12). The concept behind $V(N)$ is that each field has a (somewhat) different relationship between the variance of disease incidence (v_{obs}) and mean incidence or (equivalently) between θ of the beta-binomial distribution and p , because the coefficient of determination of the fit of equation 1 to the data was not 100%. This indicates that different N s are required to achieve a desired C for any given p , depending on the field. Because the regression analysis results in a direct estimate of $\ln(A_x)$ ($=\hat{t}$), and not of A_x or a , $\exp(\hat{t}n^{\hat{b}-2})$ was used for $\hat{a}$ in equation 4 to calculate $V(N)$. At any given value of p (and assuming, for convenience, that the estimate of p is not correlated with the binary power law parameter estimates), $V(N)$ can be written as a (very complicated) function of $\hat{b}$, $\hat{t}$, and their variances and covariance from the regression analysis (L. V. Madden, *unpublished data*). Additional complexities regarding the statistical distribution of N with sequential sampling (24) are ignored here. Among other things, $V(N)$ increases with increasing N , indicating that as p decreases both N and the uncertainty of N increases. Furthermore, $V(N)$ is directly proportional to C^{-4} ; thus, $V(N)$ for $C = 0.1$ is 16 times larger than $V(N)$ for $C = 0.2$ ($[0.1 / 0.2]^{-4} = 16$). One can generally expect both a larger number of samples and a wider range of sample sizes across fields as the required level of precision is increased for sequential sampling for estimation.

When sequentially sampling for classification purposes, a considerably smaller number of samples generally is required compared with sequentially sampling for estimation (16). This can be seen in Figures 5C and 6C for the Phomopsis leaf blight with stop lines based on equation 3, with slope and intercepts defined by equations A4 through A6 (Appendix). SPRTs are characterized by small expected sample sizes (i.e., ASNs) for a given level of heterogeneity and classification error probability (21). ASNs were very low at incidence values far from the threshold (i.e., the hypothesized value, p_i), typically less than five sampling units with the parameter values (p_0 , p_1 , α , and β) used in the evaluations. Even when observed disease incidence was close to $p_i = 0.1$, ASN was only approximately 18; at higher p_i threshold values, ASNs were less than 10 when observed incidence was near p_i for the chosen parameters. Because of the observed relationship between v_{obs} and p for Phomopsis leaf blight and the inherent properties of binary data (16), the stop lines would become farther apart (equations A5 and A6) as p_i increases to 0.5 at fixed α and β parameters and a constant difference between p_0 and p_1 ($p_1 - p_0$). This would result in increasing ASN with p_i up to 0.5, and a decline as p_i approaches 1. In order to limit the mag-

Fig. 6. A, Stop lines (cumulative number of diseased individuals [T_N] versus number of sampling units [N]), B, operating characteristic (OC) curves, and C, average sample number (ASN) curves for a sequential probability ratio test of disease incidence from a sequential cluster sample (of $n = 15$ individuals per sampling unit) with parameters $\alpha = \beta = 0.05$ and four pairs of p_0 and p_1 parameters. The parameter pairs are: $p_0 = 0.05$ and $p_1 = 0.15$ ($p_i = 0.10$; solid line), $p_0 = 0.05$ and $p_1 = 0.25$ ($p_i = 0.15$; broken line), $p_0 = 0.10$ and $p_1 = 0.30$ ($p_i = 0.20$; dash-dot line), and $p_0 = 0.10$ and $p_1 = 0.40$ ($p_i = 0.25$; dash-dot-dot line). Stop lines were determined by equation 3 with slope and intercept terms given in equations A4 through A6. OC and ASN curves were determined via Monte-Carlo simulation of beta-binomial variables as described in the Appendix. Estimates of a , A , and b are 0.27, 6.57, and 1.18, respectively.

nitude of ASN in this study, we generally increased the difference between p_0 and p_1 as we increased p_i (with one exception). This essentially led to decreasing ASNs for decision making as p_i increased over the range of p_i tested. However, a general property of sequential sampling is that reducing the ASN results in a less desirable OC curve (1), that is, one that is not as high below p_i or as low above p_i . This can be seen by the curves for $p_i = 0.1$ and 0.2 in Figure 6B.

Although there was some undesirable flattening of the Monte-Carlo derived OC curves for classifying *Phomopsis* leaf blight incidence as ASN decreased with increasing p_i , bootstrap simulations with the model construction data showed that a high percentage of means were correctly classified across a wide range of p values using a small number of sampling units (Fig. 7). Validation testing confirmed that the SPRT based on the beta-binomial approximation (equations 3 and A4 through A6) resulted in correct decisions in most cases (Table 1), and that the SPRT based on the binomial distribution (equations 3 and A1 through A3) resulted in somewhat fewer correct decisions. The higher correct decision rate for the beta-binomial over the binomial was achieved for the validation data with an increase, on average, of only one to

three sampling units. The overall type II error rate for the beta-binomial model with the validation data was (considerably) less than the chosen β parameter; the type I error rate also was less than the α parameter, except for one case with $p_i = 0.2$. The asymmetry in errors was probably due to the fact that equations A4 through A6 are only approximations for the slope and intercepts of the stop limits when data have a beta-binomial distribution (16). In particular, because equations A4 through A6 rely on the Central Limit Theorem for their derivation (16), there is no theoretical reason why this sampling model would be applicable at small sample sizes. However, the high correct decision rates under conditions with low N , as found with Monte-Carlo simulation, bootstrap evaluations of the model construction data sets, and the validation testing, demonstrate the robustness of the sampling model to violations of some simplifying assumptions. Different error rates (i.e., different OCs) could be achieved by choosing different α and β parameters (Fig. 5) or different p_0 and p_1 parameters for the given p_i (1).

Because a specific disease incidence threshold for management purposes has not yet been determined for *Phomopsis* leaf blight, we tested several different p_i values in this study. Due to the gen-

Fig. 7. **A and B,** Operating characteristic (OC), and **C and D,** average sample number (ASN) curves for a sequential probability ratio test of disease incidence from a sequential cluster sample ($n = 15$ individuals per sampling unit), with parameters **A and C,** $p_0 = 0.05$ and $p_1 = 0.15$ ($p_i = 0.10$), and **B and D,** $p_0 = 0.10$ and $p_1 = 0.30$ ($p_i = 0.20$). OC and ASN curves were determined via Monte-Carlo (MC) simulation of beta-binomial variables, with θ parameter of the beta-binomial distribution predicted with the parameters of the binary power law (equation 1) as described in the text. Estimates of a , A , and b are 0.27, 6.57, and 1.18, respectively. Error probability parameters were either $\alpha = \beta = 0.05$, or $\alpha = \beta = 0.10$. Solid symbols represent **A and B,** achieved OC and **C and D,** ASN values from 100 bootstrap simulations of sequential sampling from 21 independent *Phomopsis* leaf blight data sets (Ach = achieved). **A through D,** Vertical lines represent p_0 and p_1 . The data sets are a randomly chosen subset, within each of seven incidence classes, from 121 independent assessments of leaf blight epidemics observed in commercial strawberry fields in Ohio in 1996 and 1997. Results not shown for data sets in which $\hat{p} > 0.25$ (for $p_i = 0.1$) or $\hat{p} > 0.50$ (for $p_i = 0.2$) because OC was ≈ 0 and ASN was ≈ 4 in all cases.

eral symmetry of SPRT results with p_t for binary data, it was not necessary to conduct evaluations for $p_t > 0.5$. That is, the ASN for $p_t = 0.8$ roughly is a mirror image of the ASN for $p_t = 0.2$, and the OC for $p_t = 0.8$ is shifted 0.6 units to the right of the one for $p_t = 0.2$. Thus, our results show that it is possible to correctly classify mean disease incidence over a wide range of disease incidence values based on relatively few sampling units.

In conclusion, sequential sampling models for estimation and classification of mean incidence of Phomopsis leaf blight were developed and evaluated by disease observations from commercial strawberry fields over 3 years. Results indicated that it was possible both to estimate p with acceptable precision ($C = 0.2$) and also classify p relative to various incidence thresholds with low ASN and high correct decision rates. Once a threshold is established, it should be possible, in part, to make management decisions based on field observations of disease intensity, as done for some other pathosystems (3,8,29).

APPENDIX

Stop lines for sequential sampling for classification purposes. For the binomial distribution and data collected as a cluster sample, the intercept and slope terms (i_0 , i_1 , and s) of the stop lines of the SPRT (equation 3) are given by

$$s = \ln \left(\frac{1-p_0}{1-p_1} \right) / \ln \left[\frac{p_1(1-p_0)}{p_0(1-p_1)} \right] \quad (A1)$$

$$i_0 = \ln \left[\frac{\beta}{(1-\alpha)} \right] / \ln \left[\frac{p_1(1-p_0)}{p_0(1-p_1)} \right] \quad (A2)$$

$$i_1 = \ln \left[\frac{(1-\beta)}{\alpha} \right] / \ln \left[\frac{p_1(1-p_0)}{p_0(1-p_1)} \right] \quad (A3)$$

Here, α and β are parameters related to, but not the same as, the probabilities of making a type I and type II error, respectively, as

TABLE 1. Correct decision and error rates (%), and mean number of sampling units for sequential probability ratio tests (SPRTs) of the incidence of Phomopsis leaf blight in 78 commercial strawberry fields in Ohio in 1998

Rates ^a	$p_t = 0.1^b$		$p_t = 0.2^c$	
	$\alpha = \beta = 0.05^d$	$\alpha = \beta = 0.10^d$	$\alpha = \beta = 0.05^d$	$\alpha = \beta = 0.10^d$
Binomial				
Correct decision	88.4	87.2	84.6	83.3
Type I error	5.1	5.1	11.5	12.8
Type II error	6.5	7.7	3.9	3.9
Mean N	4.4	4.2	4.2	4.1
Beta-binomial				
Correct decision	93.6	92.3	88.5	88.5
Type I error	3.8	5.1	9.0	9.0
Type II error	2.6	2.6	2.5	2.5
Mean N	7.5	5.8	5.2	5.0

^a SPRT stop lines for the binomial distribution are derived using equations A1 through A3. Correct decision = the percentage of fields that were classified correctly as having a mean incidence either higher or lower than the threshold p_t . It was assumed that the observed incidence for the entire data set was the true p . A type I error indicates that mean disease incidence was falsely declared greater than p_t . A type II error indicates that mean disease incidence was falsely declared less than p_t . Mean N is based on all 78 validation data sets. Because $\hat{p}$ varied with the data set, the ASN term is not used here because ASN refers to the mean or expected N for populations at individual (known) p values. SPRT stop lines for the beta-binomial distribution were derived using the normal approximation to the beta-binomial (equations A4 through A6), with the variance a function of the mean, as specified by the binary power law (equation 1).

^b $p_t = (p_0 + p_1) / 2$; $p_0 = 0.05$ and $p_1 = 0.15$.

^c $p_t = (p_0 + p_1) / 2$; $p_0 = 0.10$ and $p_1 = 0.30$.

^d α and β are parameters used, in part, to control type I and type II error rates of the SPRT (1), respectively.

shown by Madden and Hughes (16). Hoffman et al. (9) and Nault and Kennedy (20) used the binomial-based stop lines for aggregated plant infestation data collected with cluster sampling. For the beta-binomial approximation with the variance described by the binary power law (equation 1) (15), the slope and intercept terms of equation 3 are

$$s = p_t \quad (A4)$$

$$i_0 = \frac{\ln[\beta/(1-\alpha)] A n^{1-b} [p_t(1-p_t)]^b}{p_1 - p_0} \quad (A5)$$

$$i_1 = \frac{\ln[(1-\beta)/\alpha] A n^{1-b} [p_t(1-p_t)]^b}{p_1 - p_0} \quad (A6)$$

Here, $A = A_x / n^{2-b}$, and A_x and b are the parameters of the binary power law. Equations A4 through A6 were derived assuming that the estimated p parameter for a data set ($\hat{p}$), but not the frequency of diseased leaflets per sampling unit (x_i), can be approximated by a normal distribution (16).

Monte-Carlo simulation. OC and ASN curves were calculated for the SPRT by Monte-Carlo simulation based on a modified algorithm of Hoffman et al. (9). In short, for each of 29 different values of p (the true disease incidence), data from a binomial distribution or a beta-binomial distribution were generated (with $n = 15$ individuals per sampling unit). For the beta-binomial cases, the θ heterogeneity parameter at each value of p was determined as

$$\theta = [a - f(p) / n] / [f(p) - a]$$

in which $f(p) = [p(1-p)]^{1-b}$, $a = A_x n^{b-2}$, and A_x and b are the binary power law parameters of equation 1 (15,27). The estimates of A_x and b from the regression analysis of the combined data of 1996 and 1997 (i.e., model construction data set) were used in the expression for θ . For each of the 29 values of p , sampling was simulated by sequentially drawing random values from the distribution. After each selection, the cumulative number of diseased individuals (T_N) was calculated and sampling (i.e., the random selections) ceased when T_N crossed either the upper or lower stop line. Stop lines were constructed by equation 3 with slope and intercept terms defined in equations A1 through A3 for the binomial and equations A4 through A6 for the beta-binomial distribution. This was repeated 1,000 times for each selection of p . Simulation results are only presented for the beta-binomial data. The general theoretical effects of binomial data or binomial-based stop lines on OC and ASN are described in Madden and Hughes (16).

ACKNOWLEDGMENTS

Salaries and research support were provided by state and federal funds appropriated to the Ohio Agricultural Research and Development Center (OARDC), The Ohio State University. We thank L. Wilson, M. Lipps, and B. Ellis for technical assistance. The Mathcad worksheet, Minitab macro, and BBD program are distributed by the authors for free.

LITERATURE CITED

- Binns, M. R. 1994. Sequential sampling for classifying pest status. Pages 137-174 in: Handbook of Sampling Methods for Arthropods in Agriculture. L. P. Pedigo and G. D. Buntin, eds. CRC Press, Boca Raton, FL.
- Campbell, C. L., and Madden, L. V. 1990. Introduction to Plant Disease Epidemiology. John Wiley & Sons, New York.
- Dillard, H. R., and Seem, R. C. 1990. Use of an action threshold for common maize rust to reduce crop loss in sweet corn. Phytopathology 80:846-849.
- Efron, B. 1982. The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia.
- Ellis, M. A. 1995. Integrated Pest Management (IPM). Disease Management Guidelines for Strawberries in Ohio. Ohio State Univ. Plant Pathol.

- Dep. Ser. 93. Ohio State University, Columbus.
6. Ellis, M. A. 1998. Strawberry Leaf Spot Diseases. Ohio State University Extension Fact Sheet HYG-3015-95. Ohio State University, Columbus.
 7. Funt, R. C., Ellis, M. A., and Welty, C. 1997. Midwest Small Fruit Pest Management Handbook. The Ohio State Univ. Extension Bull. 861, Columbus.
 8. Gaunt, R. E., and Cole, M. J. 1992. Sequential sampling for wheat stripe rust management. *Crop Prot.* 11:138-140.
 9. Hoffman, M. P., Nyrop, J. P., Kirkwylund, J. J., Riggs, D. M., Gilrein, D. O., and Moyer, D. D. 1996. A sequential sampling plan for use in scheduling control of Lepidopterous pests for fresh market sweet corn. *J. Econ. Entomol.* 89:386-395.
 10. Hughes, G., Madden, L. V., and Munkvold, G. P. 1996. Cluster sampling for disease incidence data. *Phytopathology* 86:132-137.
 11. Hutchinson, W. D. 1994. Sequential sampling to determine population density. Pages 208-243 in: *Handbook of Sampling Methods for Arthropods in Agriculture*. L. P. Pedigo and G. D. Buntin, eds. CRC Press, Boca Raton, FL.
 12. Lawless, J. F. 1982. *Statistical Models and Methods for Lifetime Data*. Wiley-Interscience, New York.
 13. Madden, L. V., and Ellis, M. A. 1988. How to develop plant disease forecasters. Pages 191-208 in: *Experimental Techniques in Plant Disease Epidemiology*. J. Kranz and J. Rotem, eds. Springer-Verlag, Berlin.
 14. Madden, L. V., and Hughes, G. 1994. BBD—computer software for fitting the beta-binomial distribution to disease incidence data. *Plant Dis.* 78:536-540.
 15. Madden, L. V., and Hughes, G. 1995. Plant disease incidence: Distributions, heterogeneity, and temporal analysis. *Annu. Rev. Phytopathol.* 33:529-564.
 16. Madden, L. V., and Hughes, G. 1999. Sampling for plant disease incidence. *Phytopathology* 89:1088-1103.
 17. Madden, L. V., Hughes, G., and Ellis, M. A. 1995. Spatial heterogeneity of the incidence of grape downy mildew. *Phytopathology* 85:269-275.
 18. Madden, L. V., Hughes, G., and Munkvold, G. P. 1996. Plant disease incidence: Inverse sampling, sequential sampling, and confidence intervals when mean incidence is zero. *Crop Prot.* 15:621-632.
 19. Manly, B. F. J. 1997. Randomization, Bootstrap, and Monte Carlo Methods in Biology, 2nd ed. Chapman and Hall, London.
 20. Nault, B. A., and Kennedy, G. G. 1996. Sequential sampling plans for use in timing insecticide applications for control of European corn borer (Lepidoptera: Pyralidae) in potato. *J. Econ. Entomol.* 89:1468-1476.
 21. Nyrop, J. P., and Binns, M. 1991. Quantitative methods for designing and analyzing sampling programs for use in pest management. Pages 67-132 in: *Handbook of Pest Management in Agriculture*, Vol. 2. D. Pimentel and A. A. Hanson, eds. CRC Press, Boca Raton, FL.
 22. Nyrop, J. P., Binns, M. R., and van der Werf, W. 1999. Sampling for IPM decision making: Where should we invest time and resources? *Phytopathology* 89:1104-1111.
 23. Pedigo, L. P. 1989. *Entomology and Pest Management*. Macmillan Publishing Co., New York.
 24. Sampford, M. R. 1962. *An Introduction to Sampling Theory*. Oliver and Boyd, Edinburgh.
 25. Turechek, W. W., Ellis, M. A., and Madden, L. V. 1998. Evaluation of fungicides for control of strawberry leaf blight, 1997. *Fungi. Nemat. Tests* 53:126.
 26. Turechek, W. W., Ellis, M. A., and Madden, L. V. 1999. Evaluation of fungicides for control of strawberry leaf blight. Pages 8-10 in: *Proc. Ohio Fruit Vegetables Growers Congr.*, Toledo, OH.
 27. Turechek, W. W., and Madden, L. V. 1999. Spatial pattern analysis of strawberry leaf blight in perennial production systems. *Phytopathology* 89:421-433.
 28. Turechek, W. W., and Madden, L. V. 1999. Spatial pattern analysis and sequential sampling for the incidence of leaf spot on strawberry in Ohio. *Plant Dis.* 83:992-1000.
 29. Verreet, J. A., Klink, H., and Hoffmann, G. M. 2000. Regional monitoring for disease prediction and optimization of plant protection measures: The IPM wheat model. *Plant Dis.* 84:816-826.
 30. Wald, A. 1947. *Sequential Analysis*. John Wiley & Sons, New York.
 31. Williams, D. A. 1975. The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics* 31:949-952.