Phytopathology • 2020 • 110:734-743 • https://doi.org/10.1094/PHYTO-10-18-0372-R

Analytical and Theoretical Plant Pathology

e-Xtra*

Quantitative Ordinal Scale Estimates of Plant Disease Severity: Comparing Treatments Using a Proportional Odds Model

K. S. Chiang,[1] H. I. Liu,[1] Y. L. Chen,[1] M. El Jarroudi,[2] and C. H. Bock[3,†]

[1] Division of Biometrics, Department of Agronomy, National Chung Hsing University, Taichung, Taiwan
[2] Department of Environmental Sciences and Management, Université de Liège, 6700 Arlon, Belgium
[3] Southeastern Fruit and Tree Nut Research Laboratory, U.S. Department of Agriculture Agricultural Research Service, Byron, GA 31008, U.S.A.

Accepted for publication 14 December 2019.

ABSTRACT

Studies in plant pathology, agronomy, and plant breeding requiring disease severity assessment often use quantitative ordinal scales (i.e., a special type of ordinal scale that uses defined numeric ranges); a frequently used example of such a scale is the Horsfall-Barratt scale. Parametric proportional odds models (POMs) may be used to analyze the ratings obtained from quantitative ordinal scales directly, without converting ratings to percent area affected using range midpoints of such scales (currently a standard procedure). Our aim was to evaluate the performance of the POM for comparing treatments using ordinal estimates of disease severity relative to two alternatives, the midpoint conversions (MCs) and nearest percent estimates (NPEs). A simulation method was implemented and the parameters of the simulation estimated using actual disease severity data from the field. The criterion for comparison of the three approaches was the power of the hypothesis test (the probability to reject the null hypothesis when it is false). Most often, NPEs had superior performance. The performance of the POM was never inferior to using the MC at severity <40%. Especially at low disease severity (≤10%), the POM was superior to using the MC method. Thus, for early onset of disease or for comparing treatments with severities <40%, the POM is preferable for analyzing disease severity data based on quantitative ordinal scales when comparing treatments and at severities >40% is equivalent to other methods.

Keywords: ecology and epidemiology, disease control and pest management

Plant pathologists face many situations in which the measurement of nearest percent estimates (NPEs) of disease severity is time-consuming or impractical. In such situations, researchers have often used an ordinal scale of measurement (Bock et al. 2010b; Shah and Madden 2004). An ordinal scale depicts the order or ranking of measurements, but the difference among the classes is generally neither equal nor equivalent and may be qualitative or quantitative, or some combination thereof. Madden et al. (2007) also suggested that although one reason for using an ordinal scale is for convenience and speed of rating, another reason is that a rater may not be capable of easily distinguishing differences in severity within an ordinal class. Indeed, rater NPEs of disease severity are notoriously variable (Bock et al. 2009; Nita et al. 2003; Nutter et al. 1993). Therefore, NPEs of disease may be of questionable value if severity cannot be determined accurately and reliably.

Ordinal scales are commonly used as an alternative to NPEs when assessing disease severity. Such ordinal scales may be qualitative—that is, the severity of disease obtained with an ordinal rating scale is an ordered numeric variable, but the rating scale is based on descriptions of symptoms (Agresti 2007, 2010; Larrabee et al. 2014; Madden et al. 2007). In contrast, an ordinal scale with classes describing defined, consecutive numeric ranges (or intervals) can be termed a *quantitative* ordinal scale; with plant disease, this special form of the ordinal scale is generally based on the percent area with symptoms. As an example, the Horsfall-Barratt (HB) scale divides the percent scale into 12 consecutive logarithmic-based intervals of severity between 0 and 100% (Horsfall and Barratt 1945). Several quantitative ordinal scales have been developed that subdivide the percent scale into different numbers of classes and varying interval sizes (Bardsley and Ngugi 2013; Chiang et al. 2014; Forbes and Korva 1994; Hartung and Piepho 2007; Hunter 1983; Hunter and Roberts 1978). As another example, Chiang et al. (2014) indicated that a 10% linear scale that emphasizes severities ≤50% disease, and has additional grades at low severities (≤10%), may be a good choice for assessing disease severity when use of a quantitative ordinal scale is preferred. In contrast, qualitative ordinal rating scales are based on descriptions of symptoms (Madden et al. 2007). Although both quantitative and qualitative ordinal rating scales share the same structure (i.e., 1-to-*n* classes), quantitative ordinal scale classes are described by intervals of increasing and consecutively defined numeric magnitude.

In this article, we focus solely on quantitative ordinal scales. In regard to the nature of the particular quantitative ordinal scale used, Snedecor and Cochran (1989) and Madden et al. (2007) stated that for scales with classes of equal interval sizes, standard analyses such as analysis of variance (ANOVA) may be appropriate for the ordinal measurements, but only if the classes used in the scale represent equal intervals on an underlying continuous ratio scale (e.g., the percent scale). However, scales with unequal interval sizes between 0 and 100% severities are often used in visual estimates of plant disease assessment, such as the HB scale and the amended 10% scale (Chiang et al. 2014) mentioned above. If one is using data

[†]Corresponding author: C. H. Bock; clive.bock@usda.gov

Funding: This research was supported by the Ministry of Science and Technology of Taiwan, Republic of China (grant MOST 107-2313-B-005-031 to K. S. Chiang) and the U.S. Department of Agriculture Agricultural Research Service (project 6606-21220-011–00D to C. H. Bock).

This article reports the results of research only. Mention of a trademark or proprietary product is solely for the purpose of providing specific information and does not constitute a guarantee or warranty of the product by the U.S. Department of Agriculture and does not imply its approval to the exclusion of other products that may also be suitable.

*The ***e*-X**tra logo stands for "electronic extra" and indicates that supplementary figures and a supplementary data file are published online.

The author(s) declare no conflict of interest.

based on a scale with unequal sized intervals between 0 and 100%, one should not apply ANOVA directly to the scale values; instead, the recommended approach is to use the midpoint of the severity range for each class (Bock et al. 2010a, b; Chiang et al. 2014; Madden et al. 2007; Nita et al. 2003). However, use of the midpoint might amplify imprecision of quantitative values used in subsequent analysis, especially when the interval size is wider. In such cases, using the midpoint of the severity range for each class may be an unreliable procedure.

A proportional odds model, also known as an ordered logistic model, is an appealing method to choose when analyzing qualitative ordinal rating scale data (Agresti 2007). The use of the proportional odds model has been discussed previously in relation to studies in plant pathology (Fu et al. 2012; Henderson et al. 2007; Landschoot et al. 2013; Paul and Munkvold 2004; Shah and Madden 2004). Indeed, the proportional odds model can be used to analyze directly the ratings obtained from disease scales with unequal class widths (e.g., the HB scale), without the need for conversion of ratings to percentages based on class midpoints (Madden et al. 2007). Moreover, the model can incorporate covariate effects in the same way as ordinary regression analysis. These useful features of the proportional odds model make it possible to compare estimates from studies using different response scales. Potential disadvantages of the proportional odds model include the need for sufficiently large numbers of observations for each experimental unit (Shah and Madden 2004) as well as the assumption that the slopes are the same for all categories (Agresti 2007, 2010; Fu et al. 2012; Schabenberger and Pierce 2002).

In this study, we compare treatments based on ordinal scale data either (i) converted to the midpoint of the corresponding disease severity range of the ordinal class and use of a standard parametric analytical technique (a *t* test in this case) or (ii) using the proportional odds model applied directly to the ordinal class data (without midpoint conversion). This idea is novel and deals with the data using different points of view. Midpoint conversion followed by a *t* test regards the data set as "measurement" data. In contrast, if a proportional odds model is used, the data set is considered as "count" data. To the best of our knowledge, there have been no previous studies that present the advantages and disadvantages of the proportional odds model when used to compare treatments based on quantitative ordinal scale estimates of plant disease severity.

We hypothesize that there may be advantages, in some cases, to using a proportional odds model to compare treatments based on ordinal scale estimates of plant disease severity compared with the midpoint conversion procedure. In addition, we wish to determine how large the sample size should be for each experimental unit when using a proportional odds model. In this article, we focused on severity levels ≤50%, so the data presented here are directly relevant to the ranges of disease severities most often observed in the field for many pathosystems (Kranz 1977) and most often in the range of greatest interest to plant pathologists.

MATERIALS AND METHODS

General approach. Based on estimates of disease severity using a quantitative ordinal scale, the performance of the proportional odds model analyzing ordinal class data was compared with using midpoint conversion of the ordinal class intervals analyzed using a parametric *t* test for comparison of treatments (e.g., varieties, fungicides, etc.). We first considered the characteristics (i.e., structure or widths of the intervals) of the quantitative ordinal scale. Subsequently, a simulation method was employed to execute the study. The parameters of the simulation method were estimated using disease severity estimate data collected previously from the field. Power analysis (the power of the hypothesis test) was used as the criterion for comparison.

Three types of scale were taken into account. The first type was the percentage scale, a continuous ratio scale based on NPEs (with disease severity estimated by the raters to the nearest 1%). The second type was the HB quantitative ordinal scale (using only 5 of the 10 categories of the scale representing 0^+ to 3, 3^+ to 6, 6^+ to 12, 12^+ to 25, and 25^+ to 50% disease severity; assigned the ordinal classes of 1, 2, 3, 4, and 5, respectively) (Horsfall and Barratt 1945). For simplification, we regarded the disease severity as 50% when the disease severity >50%. The third type was an amended 10% quantitative ordinal scale based on 10% linear intervals emphasizing severities ≤50% disease with additional classes at low severities (seven classes representing 0^+ to 1, 1^+ to 4, 4^+ to 10, 10^+ to 20, 20^+ to 30, 30^+ to 40, and 40^+ to 50% disease severity; assigned the ordinal classes 1, 2, 3, 4, 5, 6, and 7, respectively) (Chiang et al. 2014). As with the HB scale, we regarded the disease severity as 50% when the disease severity >50%.

A standard procedure is that NPEs are converted to the appropriate class of the quantitative ordinal scale for data analysis or performing simulations (Bock et al. 2010a; Chiang et al. 2014; Nita et al. 2003). The scale data are subsequently converted to the appropriate interval midpoint value of each class for analysis using a *t* test.

Disease assessment data. We used two previously collected data sets to estimate the required parameters: rater estimates of the severity of citrus canker, caused by *Xanthomonas citri* [Hasse] Gabriel et al. on leaves of grapefruit (*Citrus* × *paradisi* Macfad.), and rater estimates of the severity of Septoria leaf blotch (SLB) caused by *Zymoseptoria tritici* (Desm.) Quaedvlieg & Crous on leaves of winter wheat (*Triticum aestivum* L.). The data sets were described previously (Bock et al. 2008a, b, 2015; Chiang et al. 2017a; El Jarroudi et al. 2015).

Briefly, the citrus canker data set comprised assessments of a sample of 210 diseased leaves by three different raters along with the actual (true) disease values. Actual disease severity was measured on a leaf-by-leaf basis using image analysis (ASSESS; American Phytopathological Society, St. Paul, MN). Rater NPEs of disease severities and image analysis measurements were for each of the 210 leaves on two separate occasions. The SLB data set comprised rater-estimated and actual severity data from samples of leaves of winter wheat from plants both in control plots and plots receiving fungicides in field experiments in the Grand-Duchy of Luxembourg. As for the citrus canker data set, actual disease severity was measured on a leaf-by-leaf basis using image analysis and ASSESS software (Lamari 2002). Four raters assessed the severity of SLB on the wheat leaves and represent different hypothetical rater types used in the study (Bock et al. 2015; Chiang et al. 2017a; El Jarroudi et al. 2015). Raters 3 and 4 overestimated except at extremely high severities, whereas rater 2 underestimated and rater 1 had relatively accurate estimates. This spectrum was considered to provide a fair representation of the rater population. In 2006, 345 leaves from control plots and 240 leaves from fungicide-treated plots were photographed, image analyzed, and assessed; in 2007, 201 leaves from control plots and 171 leaves from fungicide-treated plots were photographed, image analyzed, and assessed (a grand total of 957 leaves).

Simulation study. As in several previous studies (Bock et al. 2010a; Chiang et al. 2014, 2016a, b, 2017b), a two-stage simulation approach was employed to approximate the mechanisms governing sampling of specimens for disease severity estimation in relation to hypothesis testing. This simulation method considered both the variation in symptom severity among individuals in a field plot (an error at stage I) and the error rate in assessment (an error at stage II). The algorithm in the simulation process was as follows.

First, two normally distributed, hypothetical populations of plants with disease (treatments A and B) were compared. An actual severity (Y_{actual}) value for a treatment was selected. A linear model was used to describe the relationship between the mean rater-estimated severity (μ_{rater}) and the actual severity (Y_{actual}) (Bock et al. 2010a; Chiang et al. 2014, 2016a, b). The standard deviation of the rater mean estimate (σ_{rater}) was regarded as a function of Y_{actual} determined by the rater estimates of severity using the real data from the field. The relationships are as follows (equations 1 and 2):

$$\mu_{rater} = \theta Y_{actual} \tag{1}$$

$$\sigma_{rater} = f(Y_{actual}) \tag{2}$$

A linear relationship between estimated disease and actual disease severity is generally observed (Bock et al. 2008a, b; Forbes and Korva 1994; Nita et al. 2003; Nutter and Esker 2006; Sherwood et al. 1983), so we used a generalized rater distribution describing an unbiased rate ($\theta = 1$).

Second, according to previous articles (Bock et al. 2010a; Chiang et al. 2014, 2016a, b; Forbes and Korva 1994), the frequency (y) of NPEs of specific actual disease severities by raters was assumed to follow a log-normal distribution (a positively skewed distribution; equation 3). That is,

$$y \sim Lognormal(\mu, \rho^2) \tag{3}$$

with parameters $-\infty < \mu < \infty$; $0 < \rho < \infty$. This will be denoted log_e (y) ~ *Normal* (μ, ρ^2). Thus, the mean of log_e (y) is μ and the variance of log_e (y) is ρ^2. The two parameters (μ and ρ^2) of equation 3 were acquired from μ_{rater} and σ_{rater} in equations 1 and 2. Their relationships are as follows (equations 4 and 5):

$$\mu = \ln(\mu_{rater}) - \frac{1}{2}\ln\left[1 + \left(\frac{\sigma_{rater}}{\mu_{rater}}\right)^2\right] \tag{4}$$

$$\rho = \sqrt{\ln\left[1 + \left(\frac{\sigma_{rater}}{\mu_{rater}}\right)^2\right]} \tag{5}$$

Third, a simulation value (NPE) based on the distribution of rater-estimated disease severities was obtained using equation 3. An example and further discussion of the simulation process is described in the Appendix.

Relationship between severity and σ_{rater}. To establish the relationship between the standard deviation of the rater mean NPE (σ_{rater}) and the actual disease severity for estimates, the rater estimates from 0 to 100% were divided into consecutive groupings with an approximately equivalent number of estimates in each interval.

Here, we used both data sets and subsets thereof to obtain results based on a range of rater precision. In order to present these contrasting rater abilities, we combined the data for the four raters from the SLB study (data set 1), only raters 3 and 4 from the SLB study (data set 2), all raters from the citrus canker study (data set 3), only rater 2 from the SLB study (data set 4), and only rater 1 from the SLB study (data set 5). For each of the above data sets, the data were subjected to nonlinear regression techniques, and a solution (e.g., hyperbolic or parabolic) was found to be best suited to describe the relationships between the standard deviations of the rater mean NPE and the actual disease severities. The parameters, the corresponding standard error, and the coefficient of determination (R^2) for each of the scenarios were used to evaluate the appropriateness of the model (i.e., hyperbolic or parabolic). These analyses were calculated using SAS software (version 9.4; SAS Institute, Cary, NC).

The proportional odds model. When response categories are ordered, the logits can utilize the ordering. A cumulative probability for Y is the probability that Y falls at or below a particular point (Agresti 2007). For outcome category j, the cumulative probability is as follows (equation 6):

$$P(Y \leq j) = \pi_1 + \pi_2 + \cdots \pi_j, j = 1, 2, \cdots, J \tag{6}$$

The logits of the cumulative probabilities are shown in equation 7:

$$\text{logit}[P(Y \leq j)] = \log\left(\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right) = \log\left(\frac{\pi_1 + \pi_2 + \cdots \pi_j}{\pi_{j+1} + \cdots + \pi_J}\right) \tag{7}$$

$j = 1, 2, \ldots, J - 1$. These are called cumulative logits. For example, for $J = 3$, models use both of the following (equation 8):

$$\text{logit}[P(Y \leq 1)] = \log\left(\frac{\pi_1}{\pi_2 + \pi_3}\right) \text{and}$$

$$\text{logit}[P(Y \leq 2)] = \log\left(\frac{\pi_1 + \pi_2}{\pi_3}\right) \tag{8}$$

Each cumulative logit uses all of the response categories. For only one explanatory variable x, the model shown in equation 9 has parameter β describing the effect of x on the log odds of response in category j or below:

$$\text{logit}[P(Y \leq j)] = \alpha_j + \beta x, j = 1, 2, \cdots, J - 1 \tag{9}$$

Here, let $x = 1$ for those sampled from treatment A and $x = 0$ for those sampled from treatment B. The variable x is called an indicator variable. It indicates categories for the predictors. A model for cumulative logit j looks like a binary logistic regression model in which categories $1 - j$ combine to form a single category and categories $j + 1$ to J form a second category. Thus, it can calculate the degree of the estimated odds that treatment A trends in a lower severity direction rather than a higher severity direction (i.e., $Y \leq j$ rather than $Y > j$) as compared with the estimated odds for treatment B.

The common effect β for each j implies an assumption that the curves have the same shape in order to obtain model parsimony in equation 9. The score test of the proportional odds assumption can be used to test the hypothesis that the effects are the same for each cumulative logit. That is, the score test compares the model with one parameter for x to a more complex model with the different parameters for each j (Agresti 2007, 2010).

Power analysis. The power of the hypothesis test using a simulation procedure was used to compare the performance of each of the rater precision characteristics and for comparing the methods of rating and analyzing the quantitative ordinal scale. Assuming two treatments, A and B, are applied to developing epidemics, the disease severity distribution of treatment A has mean μ_A and that of treatment B has mean $\mu_B = \mu_A + \mu_\Delta$, where μ_Δ represents the difference between the means of the two severity distributions. The standard deviations (φ) of the disease severity distributions of treatments A and B are assumed to be equal. This approach was applied to both the midpoint data analyzed using a *t* test and to the class ratings analyzed using the proportional odds model.

Thus, there were five assessment and/or analysis methods that were compared: NPEs (nearest NPE values analyzed using a *t* test), HB-MC (HB scale and midpoint conversion analyzed using a *t* test), HB-POM (HB scale analyzed using a proportional odds model), AM-MC (amended 10% scale and midpoint analyzed using a *t* test), and AM-POM (amended 10% scale analyzed using a proportional odds model).

To calculate the probability that H_0 is rejected, the simulation procedure outlined above was repeated 5,000 times. A proportional odds model or a *t* test were performed using the HB and AM scale data (the proportional odds model was run first, followed by the *t* test on each set). NPEs were subject only to a *t* test. Thus, using each of the five methods (NPE, HB-MC, HB-POM, AM-MC, and AM-POM), the proportion of occasions that H_0 was rejected in the 5,000 simulations was plotted against sample size (which may be described as the number of replicates in each treatment) for a range of set disease severity population means of 1, 5, 10, 20, 30, and 40%. For each test, the difference between the population means (μ_Δ) was assumed to be either 0 or 5% in order to identify the rate of type I and type II error.

The effect of aggregated distribution of disease severity data. In a previous section, disease severity was assumed to have a normal distribution. However, disease severity patterns are often aggregated (heterogeneous), for example, owing to aggregation of inoculum sources and/or microclimate effects. Under an aggregated distribution, we would expect to have more data points with extreme disease severities. Thus, to test the effect of an aggregated distribution of disease in treatments A and B, additional simulations using the *t* distribution, which has a longer tail than the normal distribution, in place of the normal distribution described above were also run and the results presented.

Software and code used in the analyses. The statistical analyses and simulations were performed in R software (R Core Team 2018). The *rtnorm* function of the *msm* package (Jackson 2016) was used to generate the truncated-normal random variables for the simulation study. The reason for using the truncated-normal distribution, rather than a normal distribution, was based on the fact that the actual severities cannot be negative values. For an aggregated distribution of disease severity data, the *rtt* function of the *crch* package (Messner et al. 2018) was used to generate the truncated *t* random variables for the simulation study. In addition, when the truncated *t* random variables were simulated, the corresponding degrees of freedom (sample size – 1) were adjusted. The log-normal random variables were generated using the *rlnorm* function of the *stats* package, which is stored in R. In order to apply a proportional odds model to the simulation dataset, we used the *vgam* package in R (Yee 2018). The *vglm* function was used to calculate the corresponding parameters and standard errors.

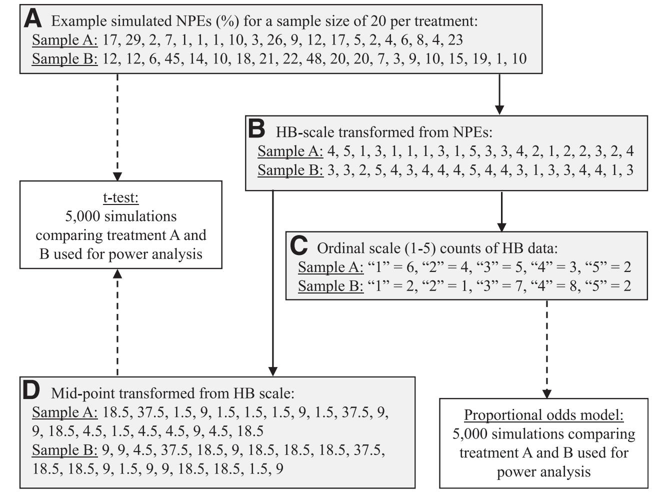
A Example simulated NPEs (%) for a sample size of 20 per treatment:
Sample A: 17, 29, 2, 7, 1, 1, 1, 10, 3, 26, 9, 12, 17, 5, 2, 4, 6, 8, 4, 23
Sample B: 12, 12, 6, 45, 14, 10, 18, 21, 22, 48, 20, 20, 7, 3, 9, 10, 15, 19, 1, 10
B HB-scale transformed from NPEs:
Sample A: 4, 5, 1, 3, 1, 1, 1, 3, 1, 5, 3, 3, 4, 2, 1, 2, 2, 3, 2, 4
Sample B: 3, 3, 2, 5, 4, 3, 4, 4, 4, 5, 4, 4, 3, 1, 3, 3, 4, 4, 1, 3
t-test:
5,000 simulations comparing treatment A and B used for power analysis
C Ordinal scale (1-5) counts of HB data:
Sample A: "1" = 6, "2" = 4, "3" = 5, "4" = 3, "5" = 2
Sample B: "1" = 2, "2" = 1, "3" = 7, "4" = 8, "5" = 2
D Mid-point transformed from HB scale:
Sample A: 18.5, 37.5, 1.5, 9, 1.5, 1.5, 1.5, 9, 1.5, 37.5, 9, 9, 18.5, 4.5, 1.5, 4.5, 4.5, 9, 4.5, 18.5
Sample B: 9, 9, 4.5, 37.5, 18.5, 9, 18.5, 18.5, 18.5, 37.5, 18.5, 18.5, 9, 1.5, 9, 9, 18.5, 18.5, 1.5, 9
Proportional odds model:
5,000 simulations comparing treatment A and B used for power analysis

Fig. 1. Flowchart presenting the analytical methods for comparing three fundamental approaches to analyze quantitative ordinal scale estimates of plant disease severity for comparing two treatments, A and B. **A,** The simulated nearest percent estimates (NPEs) prior to use of the *t* test to compare samples. **B,** The Horsfall-Barratt (HB) scale data transformed from NPEs. **C,** The layout of the count data for analysis using a proportional odds model based on counts of data in classes in B. **D,** The midpoint data converted from the HB scale data in B prior to the use of the *t* test. In this test, the difference between the population means (μ_Δ) is assumed to be 5%. φ (the standard deviation of treatment A or B) = 5%, μ_A (the mean of the disease severity distribution of treatment A) = 10%. Gray boxes present sample severity data using different scales; white boxes represent data analysis method; solid arrows represent data transformation steps; and dashed arrows represent the data used for the analysis (a *t* test or the proportional odds model) and power analysis steps.

RESULTS

Comparison of assessment methods and data sets. In order to demonstrate how the five methods were compared, we provide a simple example (Fig. 1): (i) the assessment methods consist of the simulated NPEs (Fig. 1A); (ii) the HB scale data are transformed from the NPEs (Fig. 1B); (iii) the ordinal scale data (Fig. 1C) and (iv) the midpoint-transformed data (Fig. 1D) are obtained from the HB scale data. Thus, the midpoint approach regards the data as "continuous measurement" data amenable to analysis using a *t* test, whereas the ordinal data are considered as "count" data, and a proportional odds model is used for analysis. In this example, φ, μ_A, and μ_Δ equal 5, 10, and 5%, respectively. The sample size per treatment equals 20.

A comparison of data sets 4 and 5 with data sets 1, 2, and 3 shows that data sets 4 and 5 had smaller standard deviations for the rater mean estimates at severities ≤30%, indicating more precise data over this range (Fig. 2). For data sets 1, 2, and 3, there are larger and more diverse rater errors in estimation than for the actual disease severity at severities ≤30%; specifically, the quality of the estimates by raters 1 and 2 (data sets 4 and 5) from the SLB data set is better.

Effect of rater precision, assessment, and analysis methods on the power of the hypothesis test. At low disease severity (≤10%) (Fig. 3), the results show that when using the standard deviation of the rater mean estimates for data sets 1 and 2 (estimates by less accurate or precise raters), the performance of the proportional odds model is superior to that of the midpoint conversion of the interval method (first and second columns of Fig. 3), regardless of the nature of the quantitative ordinal scale used (HB-POM or AM-POM). With data set 3, no method is noticeably superior (third column of Fig. 3). Similarly, when using the standard deviation of the rater mean estimates for data sets 4 and 5 (fourth and fifth columns of Fig. 3), the performance of the proportional odds model approximates that of the midpoint conversion of the interval method. That is, as the standard deviations of mean estimated severity (using data sets 4 and 5) are lower (more accurate or precise raters), the advantage of the proportional odds model gradually diminishes. Indeed, for raters 1 and 2 (SLB data), the NPEs are slightly superior in performance compared with all other

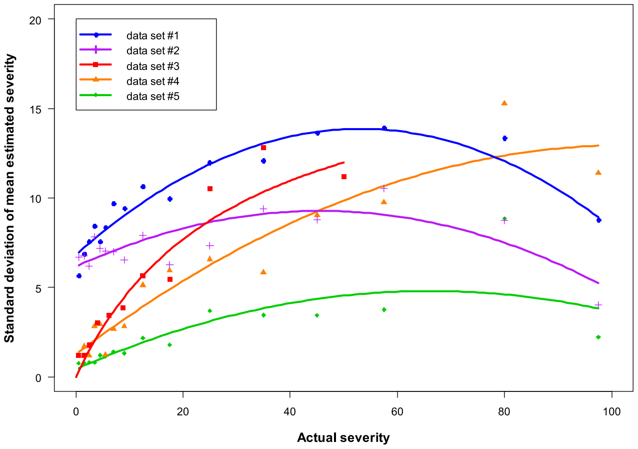

Fig. 2. Relationships between the standard deviations of the estimated means and the actual severity by four raters for severity of Septoria leaf blotch (SLB) (data set 1), raters 3 and 4 only for severity of SLB (data set 2), citrus canker (data set 3), rater 2 only for severity of SLB (data set 4), and rater 1 only for severity of SLB (data set 5). The parabolic regression solution, $\sigma_{rater} = aY_{actual}^2 + bY_{actual} + c$, was the most suitable fit for the standard deviations of severity estimates of SLB and the hyperbolic regression solution, $\sigma_{rater} = (a \times Y_{actual})/(b + Y_{actual})$, was the most suitable fit for the standard deviations of estimates of severity of citrus canker. Coefficients of determination (R^2) were 0.92, 0.61, 0.95, 0.92, and 0.63 for data sets 1 through 5, respectively.

methods. The relationship between the standard deviation of rater mean NPEs and the actual mean disease severities plays a critical role in the performance of the proportional odds model. Especially at low disease severity (≤10%), the proportional odds model is usually superior to and never inferior to the midpoint conversion of the interval method. Thus, for early stages of disease (i.e., low severities), the proportional odds model is preferable for analyzing quantitative disease severity estimation data based on ordinal scales when comparing treatments.

As severity equals or exceeds 20% (Fig. 4), there is not much difference between using the midpoint of the severity range method and using the proportional odds model. Nevertheless, the proportional odds model is not inferior to using the midpoint of the severity range at severities of 20 to 30%, especially when using the AM scale. This accounts for the significant effect that the actual severity has on the power. At a severity of 40% (Fig. 4), there is a much lower power when using HB-MC; and in this case, the proportional odds model is virtually ineffective. This is an artifact attributable to empty cells and data sparseness in the data sets and the simulation, where we imposed a maximum severity of 50%. Thus, in this study, the power using the proportional odds model for analyzing the HB scale data (HB-POM) approaches zero.

When severity is lower or σ_{rater} (the standard deviation of the rater mean estimate) is larger, more data aggregates in the same classes so it is difficult to differentiate between the means by using a *t* test to analyze the midpoint-transformed values. We use graphics to explain this phenomenon (Fig. 5). For the same simulated data presented in Figure 1 ($\varphi = 5\%$, $\mu_A = 10\%$, and sample size = 20), we compared two methods based on the HB scale: (i) the count data for using a proportional odds model (Fig. 5A) and (ii) the midpoint conversion of the interval prior to the use of the *t* test (Fig. 5B). The difference (μ_Δ) between the population means (treatments A and B) is assumed to be 5%, such that μ_A = 10% and μ_B = 15%. The relationship between the standard deviation of the rater mean NPE and the actual disease severity was established by using the data from the four raters estimating severity of SLB (data set 1). For the simulated NPEs and using a *t* test to compare the treatment difference, the null hypothesis is rejected because the *P* value equals 0.048 if the type I error rate is assumed to be 0.05. When the count data for each class are analyzed using the proportional odds model, the null hypothesis is rejected with a *P* value of 0.041 (Fig. 5A). However, the *P* value for the midpoint conversion of the HB data is 0.175 (Fig. 5B); thus, we conclude that the null hypothesis cannot be rejected (here, the difference between the means of the two severity distributions is not significant). In this case, the power is higher when using the proportional odds model. Also, for any fixed scale *j*, the estimated odds that treatment A trends in a lower severity direction rather than a higher severity direction (i.e., $Y \le j$ rather than $Y > j$) are exp(1.216) = 3.374 times the estimated odds for treatment B (the estimates for β and its standard error in equation 9 are 1.216

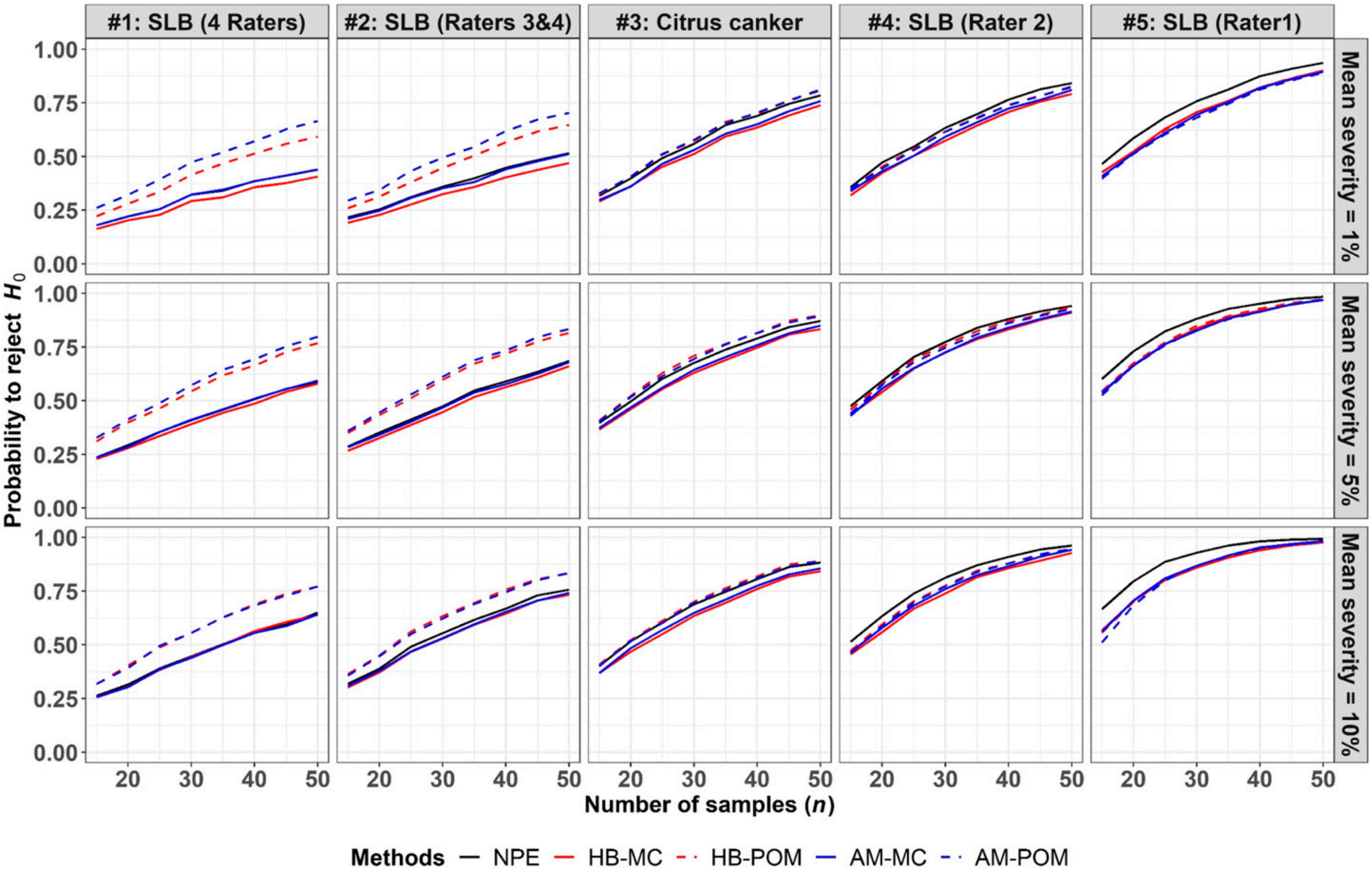

Fig. 3. Relationships between the probability of rejecting H_0 (when this hypothesis is false) and sample sizes (*n* = 15 to 50) for the different scales and analysis methods used at different disease severity means of treatment A (μ_A = 1, 5, and 10% disease severity in rows 1 through 3, respectively). The difference between the population means of treatments A and B (μ_Δ) is assumed to be 5%, φ (the standard deviation of treatment A or B) = 5%, with significance tested at *P* = 0.05. Here, two normally distributed, hypothetical populations of plants with disease (treatments A and B) were compared. The five scales and analysis methods were as follows: NPE = nearest percent estimates and analysis with a *t* test, HB-MC = the Horsfall-Barratt scale with midpoint conversion and analysis with a *t* test, HB-POM = the Horsfall-Barratt scale taking ordinal values and analysis with the proportional odds model, AM-MC = an amended 10% scale with midpoint conversion and analysis with a *t* test, and AM-POM = an amended 10% scale taking ordinal values and analysis with the proportional odds model. Data set 1 is based on all four rater estimates of severity of Septoria leaf blotch (SLB) on leaves of winter wheat. Data set 2 is based on rater 3 and rater 4 estimates for severity of SLB. Data set 3 is based on estimates of severity of citrus canker on leaves of grapefruit by three raters on two occasions. Data set 4 is based on estimates by only rater 2 for severity of SLB. Data set 5 is based on estimates by only rater 1 for severity of SLB.

and 0.595, respectively). An association exists, with lower disease severity when using treatment A compared with treatment B. Accordingly, for this example (and any similar ones), the power is higher when using the proportional odds model compared with the midpoint conversion.

We often found a lower power when using the HB scale for hypothesis testing compared with other methods at severities from 20 to 40%, regardless of whether we used the midpoint conversion of the interval or the proportional odds model. This is attributable to the increasing size of the HB scale intervals (25^{+} to 50%). Therefore, the interval width is also an important factor affecting the power of a hypothesis test.

Test of goodness of fit of the proportional odds assumption. We tested the assumption that the curves in equation 9 have the same shape. That is, is it appropriate to fit a proportional odds model with different slopes for different treatments and subsequently use a nested-model deviance test to assess whether the slopes are the same? We used the boxplots of *P* values versus sample sizes to explore whether using the proportional odds model meets the assumption (Fig. 6). If the *P* value is assumed to be 0.05, all of the analyses show that the proportional odds model meets the requirement (i.e., showed no evidence of lack of fit).

Effect of the difference between sample means (μ_{Δ}) and type I error rates. Increasing the difference between the population means increased the power of the hypothesis test for all assessment methods. When μ_{Δ} is ≥10%, the power is near 1 for all methods. As for the effect of sample standard deviation, when the standard deviation of the severity distribution is large (e.g., φ ≥10%), the hypothesis test has lower power, regardless of assessment method. As μ_{Δ} = 10% and φ = 10%, the tendencies of the power for all methods are the same as already presented in Figures 3 and 4 (data not shown).

Similar to testing type II error rates, the relationships between the probability of rejecting H_0 (when this hypothesis is true) at different sample sizes, the actual severities, and the standard deviations of the rater mean estimates for the different assessment methods were calculated. There was almost no effect of rater method on type I error rate (Supplementary Figs. S1 and S2).

Using the *t* distribution to describe populations of disease severity. We found that using the *t* distribution in place of the normal distribution resulted in patterns that were almost the same (Supplementary Figs. S3 and S4). However, at low disease severities of 1 and 5% in the citrus canker study (data set 3), the proportional odds models (HB-POM or AM-POM) slightly outperformed the NPEs (third column of Supplementary Fig. S3 versus third column of Fig. 3). It was observed that when using the *t* distribution rather than the normal distribution, the overall power of hypothesis testing decreased slightly. This is not surprising because the *t* distribution has more variability than the standard normal distribution.

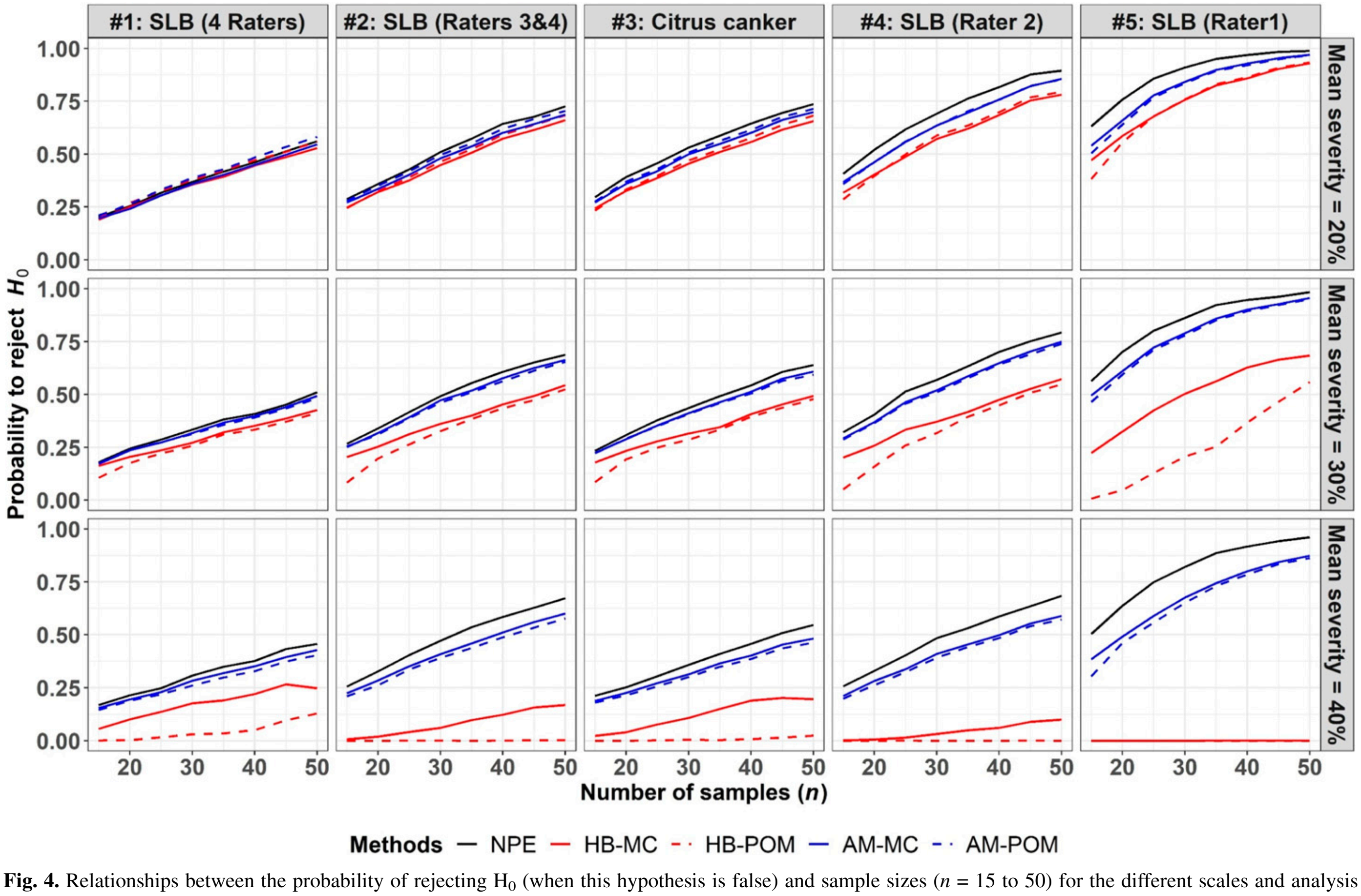

Fig. 4. Relationships between the probability of rejecting H_0 (when this hypothesis is false) and sample sizes (*n* = 15 to 50) for the different scales and analysis methods used at different disease severity means of treatment A (μ_A = 20, 30, and 40% disease severity in rows 1 through 3, respectively). The difference between the population means of treatments A and B (μ_{Δ}) is assumed to be 5%, φ (the standard deviation of treatment A or B) = 5%, with significance tested at *P* = 0.05. Here, two normally distributed, hypothetical populations of plants with disease (treatments A and B) were compared. The five scales and analysis methods were as follows: NPE = nearest percent estimates and analysis with a *t* test, HB-MC = the Horsfall-Barratt scale with midpoint conversion and analysis with a *t* test, HB-POM = the Horsfall-Barratt scale taking ordinal values and analysis with the proportional odds model, AM-MC = an amended 10% scale with midpoint conversion and analysis with a *t* test, and AM-POM = an amended 10% scale taking ordinal values and analysis with the proportional odds model. Data set 1 is based on all four rater estimates of severity of Septoria leaf blotch (SLB) on leaves of winter wheat. Data set 2 is based on rater 3 and rater 4 estimates for severity of SLB. Data set 3 is based on estimates of severity of citrus canker on leaves of grapefruit by three raters on two occasions. Data set 4 is based on estimates by only rater 2 for severity of SLB. Data set 5 is based on estimates by only rater 1 for severity of SLB.

DISCUSSION

Our results show that the performance of the proportional odds model is never inferior to using the midpoint (currently a standard procedure) of quantitative ordinal scale classes at severities <40%. Particularly when considering early stages of disease (≤10% disease severity), using an amended 10% ordinal scale (a linear scale with 10% intervals emphasizing severities ≤50% disease, and with additional intervals at severities <10%) and a proportional odds model to analyze directly the ratings obtained from disease scales is preferable to converting ratings to midpoint percentages of the class range. Kranz (1977) stated that most often plant disease is present at severities <50% (leaves often abscise if disease becomes too severe, making it difficult to obtain data on these samples), so the data presented here are of great practical value for assessments in the range of disease most often observed in the field for many pathosystems. Based on our results, using the proportional odds model to compare treatments based on ordinal estimates of disease severity can be recommended because it reduced the risk of a type II error with such data.

The impact of the standard deviations of the rater mean estimates. We found that less precise raters that have larger standard deviations of the rater mean estimates (e.g., data set 1) at low disease severities (≤10%) have a lower power for any of the methods or scales used (e.g., the comparison between data set 1 and data set 5 for a severity of 5%). The power of NPEs and the midpoint of the interval method decreases dramatically relative to the power of the proportional odds model. Thus, when using the proportional odds model, the power was greater compared with all other methods at low disease severity (≤10%) when data are imprecise (data sets 1 and 2). When visually estimating a mean disease severity, imprecise estimates compared with the actual values are the result of inaccurate individual estimates (Bock et al. 2016). This inaccuracy of individual estimates of the sample and the resulting imprecision is the cause of type II error observed in the analysis (Parker et al. 1995). This imprecision can have a major impact on hypothesis testing at low disease severities and thus not only are proportional odds models preferable in this range, the choice of scale is also critical—selecting a scale with narrow intervals at low severities is important, as is accomplished with the AM scale (Chiang et al. 2014). These observations on the importance of accurate estimates at these low disease severities accentuate the need to utilize assessment aids such as standard area diagrams to maximize accuracy of individual estimates (Del Ponte et al. 2017) to further minimize type II errors.

The influence of the actual mean severity of the samples. Besides the imprecision and resulting standard deviations of the rater mean estimates affecting the power of the hypothesis test, our results show that the actual severity is another significant factor. From the plot of empirical data, we found that the smaller the mean severity, the more skewed the frequency of observations. That is, all actual values or estimates at low mean disease severities (at least up to 5 to 10%) experience an invisible "barrier" at zero (Bock et al. 2010b). Thus, as the severity approaches 20%, the frequencies of the severity data tend to resemble a normal distribution as compared with low disease severity (≤10%) frequencies. In these cases, the advantage of the proportional odds model gradually diminishes. There is no advantage at severities >20% (but no disadvantage, either). Therefore, these two factors (σ_{rater} and the actual severity) are related to the power of hypothesis testing of the resulting samples.

The characteristics of the quantitative ordinal scale. It is clear that the nature of the quantitative ordinal scale used (i.e., the structure or widths of the intervals) is also critical to the outcome of the analysis of the results. This has been demonstrated in the discipline of plant pathology in previous studies (Chiang et al. 2014, 2016a; Hartung and Piepho 2007; Liu et al. 2019), and these findings are entirely consistent with the results of analyses based on ordinal scales used in other disciplines (Spilker 1996; Svensson 2000). The number of levels of the ordered categorical responses

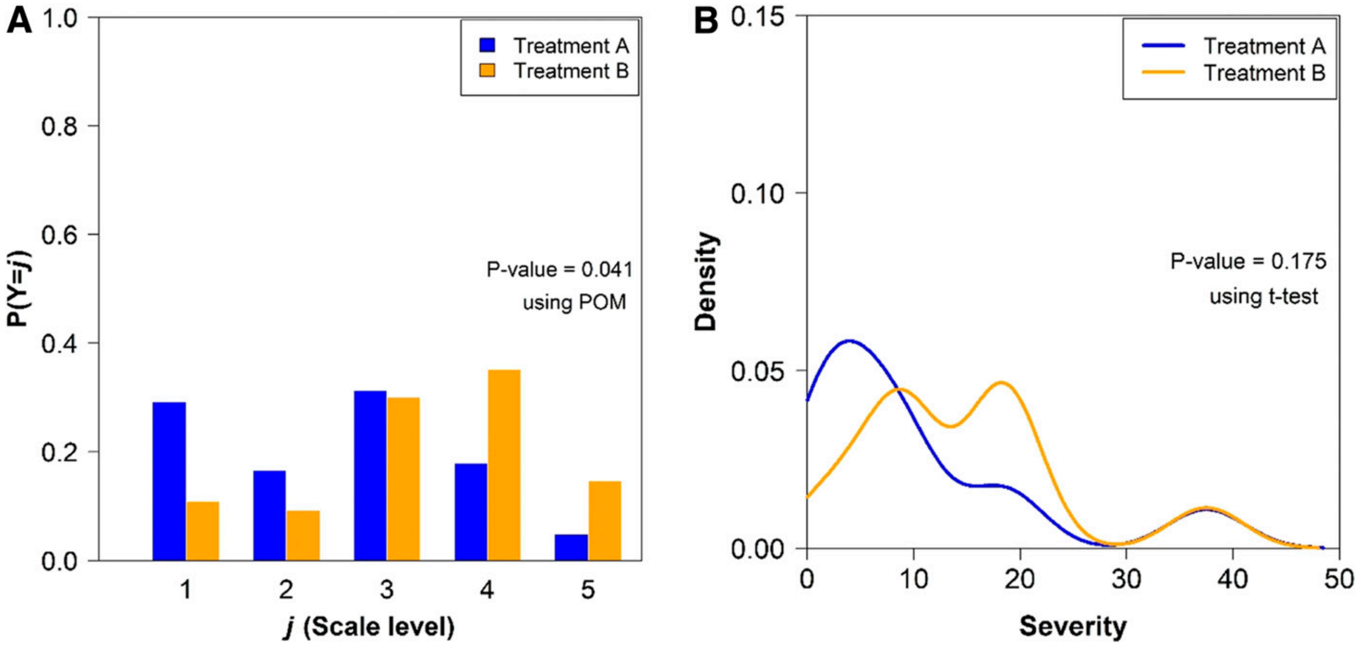

Fig. 5. Comparison of two methods of analysis using quantitative ordinal scale data of disease severity based on the Horsfall-Barratt (HB) scale: **A,** count data for each scale category used for analysis with the proportional odds model (POM); and **B,** midpoint conversion of the interval prior to the use of the *t* test to compare two treatments, A and B. For the same simulated data in Figure 1 (the standard deviation of treatment A or B [φ] = 5%, the mean of the disease severity distribution of treatment A [μ_A] = 10%, and sample size = 20), the difference (μ_Δ) between the population means for treatments A and B (μ_A and μ_B, respectively) is assumed to be 5%. The relationship between the standard deviation of the rater mean nearest percent estimate (NPE) and the actual disease severity was established by using the data from four raters who estimated the severity of Septoria leaf blotch on leaves of winter wheat (data set 1). The graph in A shows a *P* value of 0.041; hence, H_0 is rejected. However, the graph in B shows that the *P* value is 0.175, in which case we conclude that H_0 cannot be rejected. Accordingly, for this case, the power is higher when using the proportional odds model.

and asymmetry of the frequency distributions (Abreu et al. 2008; Javaras and Ripley 2007) have been reported to play a role in the quality of the resulting analysis. Our results demonstrate these same characteristics.

Pros and cons of using a proportional odds model for comparing treatments. Although NPEs were almost invariably superior, the results of this study indicate that, when comparing treatments, the performance of the proportional odds model is never

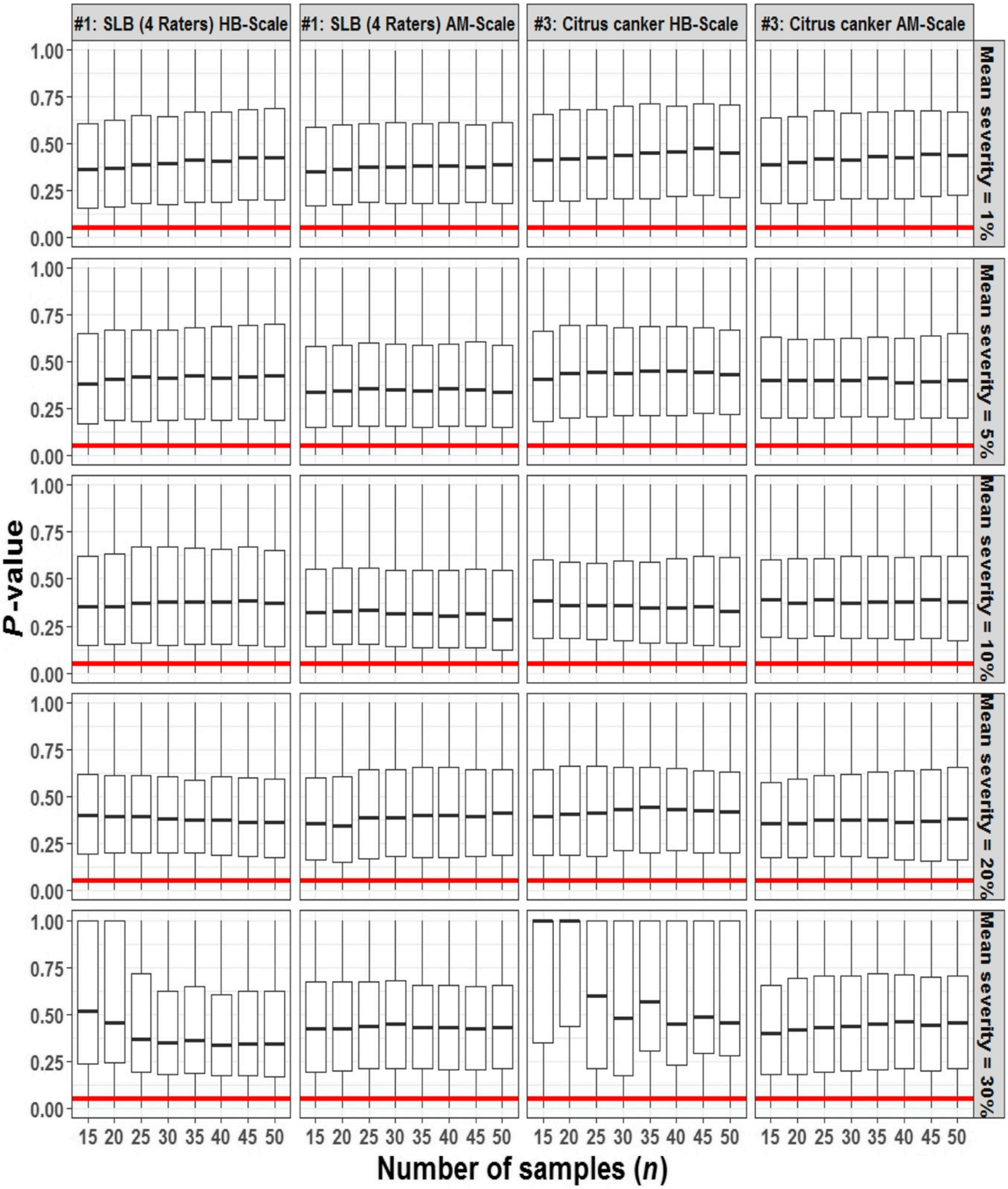

Fig. 6. Boxplots of the *P* values versus sample sizes to present a score test of the proportional odds assumption that the effects are the same for each cumulative probability. The five rows correspond to different levels of mean disease severity, from 1 to 30%. The horizontal line at the bottom of each graph indicates $P = 0.05$. The relationship between the standard deviation of the disease rater mean nearest percent estimate and the actual disease severity was established using the data from the four raters who assessed the severity of Septoria leaf blotch (SLB) on leaves of winter wheat (data set 1, the first and second columns) and the data from the three raters who assessed the severity of citrus canker on leaves of grapefruit (data set 3, the third and fourth columns). HB = Horsfall-Barratt scale and AM = amended 10% scale.

inferior to that of the midpoint method at severities <40%. Especially at low disease severity (≤10%), the proportional odds model is clearly superior to the midpoint conversion method. So, the proportional odds analysis actually improves differentiation precisely where it is most needed in many studies in plant pathology. Differentiating treatments to determine the one with the least severe disease can be important, for example, when comparing fungicide treatments or comparing disease resistance among genotypes. However, the proportional odds method has limitations; it works well only if there are sufficiently large numbers of observations for each experimental unit (Shah and Madden 2004). The question remains how large a number of observations must be collected before a proportional odds model can be successfully applied? Our simulation of the proportional odds model required that the number of observations for each experimental unit be at least 15 before this model could be applied for most practical uses.

In this article, to simplify the information offered by the intervals in the ordinal scale, we regarded the disease severity as 50% when the disease severity >50%. The reason is that most often plant disease is present at severities <50%. However, this simplification will lead to more data aggregating at the same grade in the scale, rendering the proportional odds model virtually ineffective. Thus, in the case of this study, we observed an artifact attributable to limiting the maximum disease severity to 50% in conjunction with using the HB scale (as seen with the severity of 40% in Fig. 4). If the maximum imposed severity of 50% is removed (the number of grades in the scale will be increased), the power will elevate for any methods discussed (data not shown). However, too many divisions represented in an ordinal scale might negate the assumed advantage of simplicity offered by these scales. If the structure or width of the intervals is carefully chosen (e.g., an amended 10% interval ordinal scale emphasizing severities ≤50% disease with additional grades at low severities), the limitation of the proportional odds model will be minimized, even at a severity of 40%.

The effect of aggregation of disease severity data. The better performance of the proportional odds model is seen mainly at low disease severity (≤10%) and when estimates are less precise. Aggregated distribution of disease severity often occurs in data taken in the field. Thus, it is appropriate for the *t* distribution to be used in place of the normal distribution to account for the aggregated distribution. In such cases, the proportional odds model is superior to the midpoint conversion of the interval method regardless of whether the normal distribution or the *t* distribution is used. This further confirms the merits of using the proportional odds model in conjunction with quantitative ordinal scale data, rather than using a midpoint conversion.

Conclusions. Although a proportional odds model improves differentiation precisely in the range of disease most often observed in the field for many pathosystems where quantitative ordinal scales are chosen, most plant pathologists are not familiar with this analytical method. We therefore include a detailed description of the computer software used to implement the proportional odds model as a supplementary file (Supplementary Data File S1). We believe that the results and tools of our study will be helpful in improving the outcome of treatment comparisons in botanical epidemiology and related areas of research.

APPENDIX

Simulation studies for rater distribution of severity estimates. We provide an example to demonstrate the algorithm for simulation studies of a rater estimation distribution. For μ_A = 40% at treatment A, an actual severity value (Y_{actual}) of 45% could be selected. Moreover, the difference between μ_A (40%) and Y_{actual} (45%) is the variation in area affected among individuals in a field plot. We designated this as the error at stage I. According to equation 1, the mean of a generalized rater estimation distribution (a log-normal distribution) equals 45% if $\theta = 1$. Subsequently, a certain simulation value (e.g., 47, 41, 37%, etc.) could be drawn from this log-normal distribution. The difference between the above simulated value (e.g., 47, 41, 37%, etc.) and Y_{actual} (45%) is the error rate in assessment. We designated this as the error at stage II. Furthermore, an actual severity (Y_{actual}) of 30% or even 20% (or any other value) might be selected during the simulation process, because the population (A or B) is normally distributed to mimic variation in infection among individuals in a field plot population. Thus, both the variation in infection among individuals in a field plot and the error rate in assessment were considered in the simulation process.

In this study, the severity as assessed by a generalized rater is assumed to be a log-normal distribution. A beta-distribution could be an alternative to the log-normal distribution. A beta-distribution is bounded at both ends (0, 1), whereas the log-normal is only bounded at the lower end (0, ∞) and the log-normal distribution has the advantage that the tails do not tend to infinity (i.e., the probability is small when the value of the *x*-axis is beyond 100%). Based on simulations of the random variables representing the rater log-normal and beta-distributions of severity estimates, the two distributions look very much alike except that the log-normal distribution has a longer tail than the beta-distribution. In general, there is a positively skewed distribution for disease at low severities. As severities approach midrange severity measurements, the symmetrical bell-shaped distribution resembles a normal distribution. The two distributions can be made very flexible by choosing different shape parameters based on empirical data. So, the two distributions are realistic for estimation of disease severity on the percent scale (0 to 100%) (Supplementary Fig. S5).

To describe the frequency of NPEs of specific actual disease severities by raters, a log-normal distribution is assumed in this study. The reason is that the two parameters (μ and ρ^2) of a log-normal distribution can be directly expressed by equations 4 and 5 (presented in the Materials and Methods). However, the two parameters (α and β) of a beta-distribution cannot be expressed in a closed form; hence, they must be addressed using specific software to calculate these values of α and β (Bain and Engelhardt 1992). Furthermore, it is unlikely that using a beta-distribution would make any difference to the conclusions of this study. That is, the analysis using either distribution will give comparable results (K. S. Chiang, *unpublished data*). Thus, it is reasonable that a log-normal distribution is assumed to describe the frequency of NPEs of specific actual disease severities by raters.

ACKNOWLEDGMENTS

We appreciate the thoughtful comments and suggestions made by the Editor-in-Chief, Senior Editor, and the Reviewers.

LITERATURE CITED

Abreu, M. N. S., Siqueira, A. L., Cardoso, C. S., and Caiaffa, W. T. 2008. Ordered logistic regression models: Application in quality of life studies. Cad. Saude Publica 24(suppl. 4):S581-S591.

Agresti, A. 2007. An Introduction to Analysis of Ordinal Categorical Data, 2nd ed. Wiley, Hoboken, NJ.

Agresti, A. 2010. Analysis of Ordinal Categorical Data, 2nd ed. Wiley, Hoboken, NJ.

Bain, L. J., and Engelhardt, M. 1992. Introduction to Probability and Mathematical Statistics, 2nd ed. Duxbury Press, Belmont, CA.

Bardsley, S. J., and Ngugi, H. K. 2013. Reliability and accuracy of visual methods to quantify severity of foliar bacterial spot symptoms on peach and nectarine. Plant Pathol. 62:460-474.

Bock, C. H., Chiang, K.-S., and Del Ponte, E. M. 2016. Accuracy of plant specimen disease severity estimates: Concepts, history, methods, ramifications and challenges for the future. CAB Rev. 11(32):1-21.

Bock, C. H., El Jarroudi, M., Kouadio, A. L., Mackels, C., Chiang, K. S., and Delfosse, P. 2015. Disease severity estimates–effects of rater accuracy and assessment methods for comparing treatments. Plant Dis. 99:1104-1112.

Bock, C. H., Gottwald, T. R., Parker, P. E., Ferrandino, F., Welham, S., van den Bosch, F., and Parnell, S. 2010a. Some consequences of using the

Horsfall-Barratt scale for hypothesis testing. Phytopathology 100: 1030-1041.

Bock, C. H., Parker, P. E., Cook, A. Z., and Gottwald, T. R. 2008a. Visual rating and the use of image analysis for assessing different symptoms of citrus canker on grapefruit leaves. Plant Dis. 92:530-541.

Bock, C. H., Parker, P. E., Cook, A. Z., and Gottwald, T. R. 2008b. Characteristics of the perception of different severity measures of citrus canker and the relationships between the various symptom types. Plant Dis. 92: 927-939.

Bock, C. H., Parker, P. E., Cook, A. Z., Riley, T., and Gottwald, T. R. 2009. Comparison of assessment of citrus canker foliar symptoms by experienced and inexperienced raters. Plant Dis. 93:412-424.

Bock, C. H., Poole, G., Parker, P. E., and Gottwald, T. R. 2010b. Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. Crit. Rev. Plant Sci. 29:59-107.

Chiang, K. S., Bock, C. H., El Jarroudi, M., Delfosse, P., Lee, I. H., and Liu, H. I. 2016a. Effects of rater bias and assessment method on disease severity estimation with regard to hypothesis testing. Plant Pathol. 65:523-535.

Chiang, K. S., Bock, C. H., Lee, I. H., El Jarroudi, M., and Delfosse, P. 2016b. Plant disease severity assessment - how rater bias, assessment method and experimental design affect hypothesis testing and resource use efficiency. Phytopathology 106:1451-1464.

Chiang, K. S., Liu, H. I., and Bock, C. H. 2017a. A discussion on disease severity index values. Part I: Warning on inherent errors and suggestions to maximize accuracy. Ann. Appl. Biol. 171:139-154.

Chiang, K. S., Liu, H. I., Tsai, J. W., Tsai, J. R., and Bock, C. H. 2017b. A discussion on disease severity index values. Part II: Using the disease severity index for null hypothesis testing. Ann. Appl. Biol. 171:490-505.

Chiang, K. S., Liu, S. C., Bock, C. H., and Gottwald, T. R. 2014. What interval characteristics make a good categorical disease assessment scale? Phytopathology 104:575-585.

Del Ponte, E. M., Pethybridge, S. J., Bock, C. H., Michereff, S. J., Machado, F. J., and Spolti, P. 2017. Standard area diagrams for aiding severity estimation: Scientometrics, pathosystems, and methodological trends in the last 25 years. Phytopathology 107:1161-1174.

El Jarroudi, M., Kouadio, A. L., Mackels, C., Tychon, B., Delfosse, P., and Bock, C. H. 2015. A comparison between visual estimates and image analysis measurements to determine Septoria leaf blotch severity in winter wheat. Plant Pathol. 64:355-364.

Forbes, G. A., and Korva, J. T. 1994. The effect of using a Horsfall-Barratt scale on precision and accuracy of visual estimation of potato late blight severity in the field. Plant Pathol. 43:675-682.

Fu, L. Y., Wang, Y. G., and Liu, C. J. 2012. Rank regression for analyzing ordinal qualitative data for treatment comparison. Phytopathology 102: 1064-1070.

Hartung, K., and Piepho, H. P. 2007. Are ordinal rating scales better than percent ratings? A statistical and “psychological” view. Euphytica 155:15-26.

Henderson, D., Williams, C. J., and Miller, J. S. 2007. Forecasting late blight in potato crops of southern Idaho using logistic regression analysis. Plant Dis. 91:951-956.

Horsfall, J. G., and Barratt, R. W. 1945. An improved grading system for measuring plant disease. [Abstract]. Phytopathology 35:655.

Hunter, R. E. 1983. Influence of scab on late season nut drop of pecans. Plant Dis. 67:806-807.

Hunter, R. E., and Roberts, D. D. 1978. A disease grading system for pecan scab. Pecan Q. 12:3-6.

Jackson, C. 2016. msm: Multi-State Markov and Hidden Markov Models in Continuous Time. R package version 1.6.4.

Javaras, K. N., and Ripley, B. D. 2007. An “unfolding” latent variable model for Likert attitude data. J. Am. Stat. Assoc. 102:454-463.

Kranz, J. 1977. A study on maximum severity in plant disease. Travaux dédiés à G. Viennot-Bourgin 16:9-73.

Lamari, L. 2002. ASSESS: Image Analysis Software for Plant Disease Quantification. American Phytopathological Society, St Paul, MN.

Landschoot, S., Waegeman, W., Audenaert, K., Haesaert, G., and De Baets, B. 2013. Ordinal regression models for predicting deoxynivalenol in winter wheat. Plant Pathol. 62:1319-1329.

Larrabee, B., Scott, H. M., and Bello, N. M. 2014. Ordinary least squares regression of ordered categorical data: Inferential implications in practice. J. Agric. Biol. Environ. Stat. 19:373-386.

Liu, H. I., Tsai, J. R., Chung, W. H., Bock, C. H., and Chiang, K. S. 2019. Effects of quantitative ordinal scale design on the accuracy of estimates of mean disease severity. Agronomy (Basel) 9:565.

Madden, L. V., Hughes, G., and van den Bosch, F. 2007. The Study of Plant Disease Epidemics. American Phytopathological Society, St Paul, MN.

Messner, J., Zeileis, A., and Stauffer, R. 2018. crch: Title Censored Regression With Conditional Heteroscedasticity. R package version 1.0-1.

Nita, M., Ellis, M. A., and Madden, L. V. 2003. Reliability and accuracy of visual estimation of Phomopsis leaf blight of strawberry. Phytopathology 93:995-1005.

Nutter, F. W., Jr., and Esker, P. D. 2006. The role of psychophysics in phytopathology: The Weber-Fechner law revisited. Eur. J. Plant Pathol. 114: 199-213.

Nutter, F. W., Jr., Gleason, M. L., Jenco, J. H., and Christians, N. C. 1993. Assessing the accuracy, intra-rater repeatability, and inter-rater reliability of disease assessment system. Phytopathology 83:806-812.

Parker, S. R., Shaw, M. W., and Royle, D. J. 1995. The reliability of visual estimates of disease severity on cereal leaves. Plant Pathol. 44:856-864.

Paul, P. A., and Munkvold, G. P. 2004. A model-based approach to preplanting risk assessment for gray leaf spot of maize. Phytopathology 94:1350-1357.

R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Schabenberger, O., and Pierce, F. J. 2002. Contemporary Statistical Models for the Plant and Soil Sciences. CRC Press, Boca Raton, FL.

Shah, D. A., and Madden, L. V. 2004. Nonparametric analysis of ordinal data in designed factorial experiments. Phytopathology 94:33-43.

Sherwood, R. T., Berg, C. C., Hoover, M. R., and Zeiders, K. E. 1983. Illusions in visual assessment of *Stagonospora* leaf spot of orchardgrass. Phytopathology 73:173-177.

Snedecor, G. W., and Cochran, W. G. 1989. Statistical Methods, 8th ed. Iowa State University Press, Ames.

Spilker, B. 1996. Quality of Life and Pharmaeconomics in Clinical Trials, 2nd ed. Lippincot-Raven Publishers, Philadelphia, PA.

Svensson, E. 2000. Comparison of the quality of assessments using continuous and discrete ordinal rating scales. Biom. J. 42:417-434.

Yee, T. W. 2018. vgam: Vector Generalized Linear and Additive Models. R package version 1.0-5.