Assessing the Accuracy, Intra-rater Repeatability, and Inter-rater Reliability of Disease Assessment Systems

F. W. Nutter, Jr., M. L. Gleason, J. H. Jenco, and N. C. Christians

First, second, and third authors: associate professors and graduate research assistant, respectively, Department of Plant Pathology, 351 Bessey Hall, Iowa State University, Ames 50011; fourth author: professor, Department of Horticulture, 138 Horticulture Building, Iowa State University, Ames 50011.

Journal paper J-15124-3116 of the Iowa Agriculture and Home Economics Experiment Station, Ames 50011.

We thank Sharon K. Parker, Melanie L. Hockmuth, Mark Carlton, and Karen Shervheim for their technical assistance in conducting this study.

Accepted for publication 30 March 1993.

ABSTRACT

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 systems. Phytopathology 83:806-812.

Dollar spot of bentgrass, caused by *Sclerotinia homoeocarpa*, was chosen as a model pathosystem to evaluate the accuracy and precision of several disease assessment methods. Quadrats were assessed by visually estimating percent disease severity, by measuring percent reflectance of sunlight at 600 and 800 nm, and by image analysis of color photographic slides. Visual and reflectance assessments were performed by four raters and repeated 24-h later to obtain a measure of intra-rater repeatability. Linear regression of the original assessments (Y) versus the repeated assessments (X) revealed that intra-rater repeatability was highest for the 600-nm radiometric method as indicated by slope values not significantly different from 1.0 ($P \leq 0.01$) and intercepts close to and not significantly different from zero ($P \leq 0.01$). Coefficients of determination (R^2) also were highest for this method, ranging from 98.7 to 99.6%, whereas R^2 values relating intra-rater visual assessments ranged from 83.4 to 93.1%. Inter-rater reliability was highest using the 600-nm radiometric method as determined from regression equations relating one rater's visual and radiometric assessments to assessments performed by other raters. Slopes and intercepts among raters using the 600-nm radiometric method did not differ significantly from 1.0 and zero, respectively. However, two of six intercepts and six of six slopes measuring inter-rater reliability of the visual assessment method were significantly different. Slopes significantly different from one indicate the presence of systematic bias among raters, whereas intercepts significantly different from zero indicate the presence of constant sources of error among raters. Radiometric assessments also had a better relationship (R^2) to the true level of dollar spot severity as determined with an acetate sheet-image analysis technique. The radiometric assessment method provided a fast, accurate method to measure dollar spot severity that was more precise than were visual assessments.

Additional keywords: multispectral radiometer, remote sensing.

Plant populations affected by diseases often are assessed to determine the level of disease intensity (1,4,6,7,13,17). Disease assessments provide quantitative data concerning the effectiveness of disease-management strategies (2,9,11), help growers make management decisions (1,4), and provide a means of estimating crop damage (10,13). Shokes et al (18) stated that any method used to assess disease intensity should 1) be easy to use, 2) be applicable over a wide range of conditions, 3) provide reproducible (precise) results, and 4) provide a rapid and accurate measure of disease intensity.

Relatively few studies have been conducted to guide researchers in selecting one disease assessment method over another (14,17) or to evaluate modifications of an existing assessment system to improve the accuracy and precision of estimates. Inter-rater reliability has been operationally defined as the ratio of the true variance (among plots) to the total variance, which includes a variance component for the error among raters (18). Although improved sampling designs and increased sample size can lower true and total variances, limited resources often restrict sample size. In addition, when more than one rater is involved, it is difficult to quantify the bias attributable to any one individual. Shokes et al (18) proposed measuring intra-rater repeatability with a test-retest correlation procedure. The correlation coefficient (r) provides a statistical measure of the relationship between repeated assessments of the same sampling units by the same individual or instrument. However, correlation analysis between two variables cannot be used to infer a cause-and-effect relationship, nor can one variable (repeated assessments) be used to predict the value of another variable (first-time assessments). Least squares regression can be used to determine if there is a significant linear relationship between disease assessments performed by different raters and whether there is a statistical relationship between repeated assessments performed by the same individual. Regression-equation parameters, such as the slope and intercept, could be used to evaluate and compare the accuracy and precision of disease assessment raters and methods (10,13,16).

Dollar spot of bentgrass (*Agrostis palustris* Huds.), caused by *Sclerotinia homoeocarpa* F.T. Bennett, is considered one of the most prevalent and persistent diseases on golf courses in the United States. The disease causes circular, sunken patches on closely mowed turfgrass that rarely exceed 5 cm in diameter. These patches coalesce into larger, irregular patches if the disease becomes severe (19). On closely mowed turfgrass, the disease resembles "leaf spots" on a large two-dimensional "leaf" surface when 1 m^2 quadrats are delineated.

The purpose of the present study was to 1) demonstrate the use of simple linear regression to detect and quantify bias within raters (intra-rater repeatability) and among raters (inter-rater reliability) as affected by visual versus radiometric assessment methods and 2) compare accuracy and precision of visual, radiometric, and image-analysis disease assessment methods on dollar spot of bentgrass. A preliminary report has been published (12).

MATERIALS AND METHODS

Experimental design. To quantify and compare the accuracy and precision of assessment methods and raters, it was first necessary to generate a wide range of dollar spot severity levels. Field plots were established on a creeping bentgrass (*A. palustris*) green located at the Iowa State University Horticulture Farm near Gilbert. This green consisted of cultivars Emerald, Penncross, and Penneagle. Because cultivar/rater and cultivar/assessment method interaction terms were not significant ($P \leq 0.01$), data from these cultivars were pooled, and the experimental design was a randomized complete block with six treatments and 12 replications (initially four per cultivar). To obtain the widest possible range of dollar spot severities, 1-m^2 bentgrass quadrats were inoculated with *Sclerotinia*-infested ryegrass (*Lolium perenne* L.) grain at six inoculum densities: 106.4, 53.4, 26.7, 13.4, 6.7, and 3.3 ml of infested grain per plot. Eight uninoculated, fungicide-treated control plots also were included but were located 5 m from inoculated plots to ensure that nearly disease-free plots could be maintained. These plots were sprayed with chlorothalonil (Daconil 2787) flowable fungicide at a rate of 0.30 L/1,000 m^2 on 9 August and 0.45 L/1,000 m^2 on 20 August. Thus, a total of 80 quadrats was available for disease assessment.

Preparation of inoculum. Inoculum was prepared by autoclaving 225 g of ryegrass grain, 4 g of $CaCO_3$, and 275 ml of water in 1-L Erlenmeyer flasks (15). Five isolates of *S. homoeocarpa* (obtained from P. Sanders, The Pennsylvania State University, University Park) were grown separately on autoclaved ryegrass grain for 10–14 days and then mixed in equal volumes prior to inoculation.

Inoculation procedure. Experimental units were established by creating a grid system of 1-m^2 quadrats. Concrete nails and string were used to demarcate individual quadrats. Treatments were assigned randomly to quadrats within replicate blocks, and the inoculum was spread uniformly within each quadrat. The experimental area was sprinkler-irrigated for 5 days after inoculation to provide environmental conditions favorable for infection. To achieve a greater range of dollar spot severities than the six inoculum levels, quadrats were mowed in North-South, East-West, and diagonal directions on alternate days to facilitate pathogen spread within and among quadrats.

Disease assessment. Prior to each assessment date, concrete nails and string were used to demarcate 1-m^2 experimental units. Percent disease severity (the diseased area of a quadrat divided by the total area of a quadrat multiplied by 100) was assessed visually by four raters on 24 August 1990. The visual disease assessments were repeated by each individual on the following day. The time required for each rater to perform these assessments also was recorded.

Selection of wavelength bands for radiometric assessments was based on differences in percent reflectance signatures as dollar spot severity increased. Percent reflectance of sunlight from bentgrass surfaces at 600 nm increased with increasing disease severity, whereas percent reflectance of sunlight at 800 nm decreased as disease severity increased. Percent reflectance at these two wavelength bands was compared with visual assessments for accuracy and precision. Percent reflectance of sunlight in the 600- and 800-nm wavelength bands was recorded for all quadrats on 24 August 1990 with a CropScan radiometer (CropScan, Inc., Fargo, ND). The radiometer was operated by the same individuals who visually assessed percent dollar spot severity. Radiometric assessments were repeated by each individual on 25 August. Percent reflectance of sunlight was recorded from a height of 2 m above the turf surface with the radiometer centered directly over each quadrat. This enabled the radiometer to measure reflectance from a 1-m-diameter circle. A bubble spirit level was used to consistently orient the radiometer to the appropriate angle and height. Reflectance measurements were recorded during cloud-free periods between 1100 and 1400 each day.

In an attempt to verify the accuracy of visual and radiometric disease assessments, 5- × 5-cm color photographic slides were made of each 1-m^2 quadrat on 24 August. Because of the high cost and length of time required to digitize color slides, 50 slides varying in dollar spot severity were arbitrarily selected for image analysis. Slides were processed and digitized by the Iowa State University (ISU) Image Analysis Facility to obtain estimates of percent dollar spot severity for each quadrat. Slides were placed on a lightbox for even illumination and were viewed from above by a color video camera. Percent diseased area was determined with a Zeiss image-analysis system (Carl Zeiss, Inc., Thornwood, NY). Image analyses of these slides were performed twice (on different days) to determine the repeatability of this method.

In a second method to determine actual percent dollar spot severity from the 5- × 5-cm slides, sheets of clear acetate were placed over projected images of the slides onto a Caramate slide projector screen (Kodak, Rochester, NY). The corresponding diseased areas were blackened on the acetate with a felt-tip pen. Because this method is extremely labor intensive, 10 color slides representing a range of dollar spot severities were arbitrarily selected for this analysis. The acetate images were digitized with an AgVision video image-analysis system (Decagon Devices, Inc., Pullman, WA) to obtain the diseased area of each quadrat expressed in square centimeters. Estimates of percent dollar spot severity were obtained by multiplying the ratio of diseased area (in square centimeters) to total area of the quadrat image (in square centimeters) and multiplying the product by 100. The acetate method was repeated to determine its repeatability.

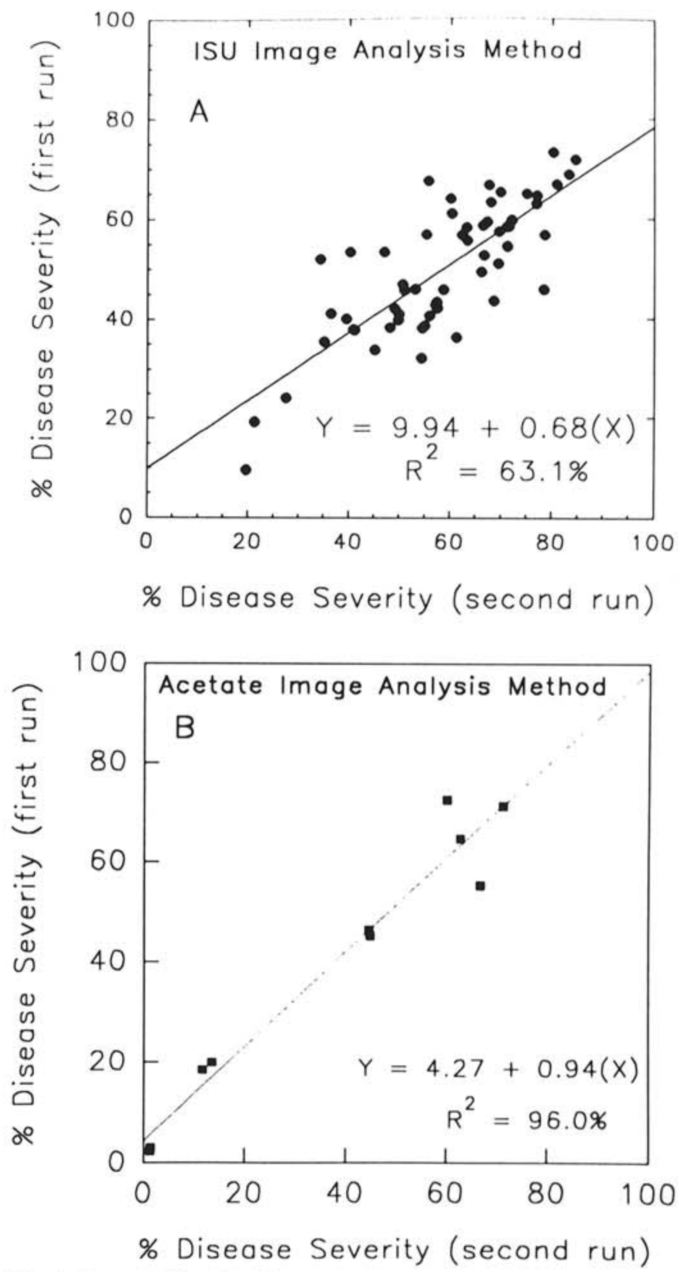

Fig. 1. Repeatability of **A,** image processing and analyses of 5- × 5-cm color slides by the Iowa State University Image Analysis Facility to estimate percent dollar spot severity of 50 bentgrass quadrats arbitrarily selected from a total of 80 quadrats, and **B,** an image-analysis technique employing 5- × 5-cm color slides and acetate sheets to estimate percent dollar spot severity of 10 bentgrass quadrats arbitrarily selected from a total of 80 quadrats.

Data analysis. Regression analysis was used to determine the relative precision of visual, remote-sensing, and image-analysis assessment methods. Precision was defined as the relative measure of the reliability and repeatability of disease assessments performed by different raters. The inter-rater reliability of assessment methods was determined by regressing the assessments made by one individual on those made by other raters. The intra-rater repeatability of assessment methods was determined by regressing the second (repeated) set of assessments obtained by each rater (X) with the first set of assessments (Y) performed by the same individual. The variables used to compare intra-rater repeatability and inter-rater reliability were slope, y-intercept, coefficient of determination (R^2), standard error of the estimate for Y, and the coefficient of variation (CV). Data presented are for the experiments conducted during August 1990. The experiment was repeated on two consecutive days during September 1990, and these results confirmed the results obtained during August.

RESULTS

Comparison of image-analysis methods. The regression coefficient relating repeated image-analysis assessments (Y) with the first image-analysis assessments of 5- × 5-cm color slides (X) was 0.68 and was significantly different from a regression coefficient (slope) of 1.0 that would be expected if this method was highly repeatable (Fig. 1A). The second set of image-analysis assessments explained just 63.1% of the initial image-analysis assessments performed with the same color slides (Fig. 1A). In comparison, the acetate image-analysis method (second run) explained 96.0% of the variation in the first-run assessments, and the slope relating the two assessments (0.94) was not significantly different from 1 (Fig. 1B). Therefore, the acetate method was selected as the best standard (truth) to evaluate visual and radiometric assessment methods for accuracy.

Intra-rater repeatability. In general, there was a good relationship between the first and second visual-assessment ratings performed by each of the four raters (Fig. 2). Although the y-intercepts did not differ significantly from zero ($P \leq 0.01$), regression coefficients ranged from 0.80 to 0.93 and were significantly different from a slope of 1.0 ($P \leq 0.01$). The amount of variation in the first assessments explained by repeated assessments (R^2) ranged from 83.4 to 93.1%.

Coefficients of determination for percent reflectance measurements with the 800-nm wavelength band were higher than the visual assessment method and ranged from 95.9 to 98.9% (Fig. 3). Slopes ranged from 1.02 to 1.12 and were not significantly different from 1. Y-intercepts ranged from −2.01 to −7.57 and were significantly different from zero for all four raters ($P \leq 0.01$).

At 600 nm, coefficients of determination ranged from 98.7 to 99.6% (Fig. 4). Regression coefficients relating the second assessments to the first assessments ranged from 0.93 to 0.97 and were not significantly different from 1. Y-intercepts ranged from 0.24 to 0.43 and were not significantly different from zero.

Inter-rater reliability. The radiometric measurements using either the 600- or 800-nm wavelength bands had higher coefficients of determination than did the visual assessments (Table 1). Coefficients of determination using the visual assessment method ranged from 70.4 (rater 2 vs. rater 3) to 89.2% (rater 1 vs. rater 4). Coefficients of determination for inter-rater reliability using the 600- and 800-nm wavelength bands ranged from 97.8 to 99.2% and from 99.1 to 99.6%, respectively.

Regression coefficients for the 600-nm radiometric assessments obtained by different raters were not significantly different from 1 ($P \leq 0.01$), whereas slopes relating visual assessments among the four raters ranged from 0.74 to 0.88 and were all significantly different from a slope equal to 1 ($P \leq 0.01$). The equations themselves do not show how differently the raters visually

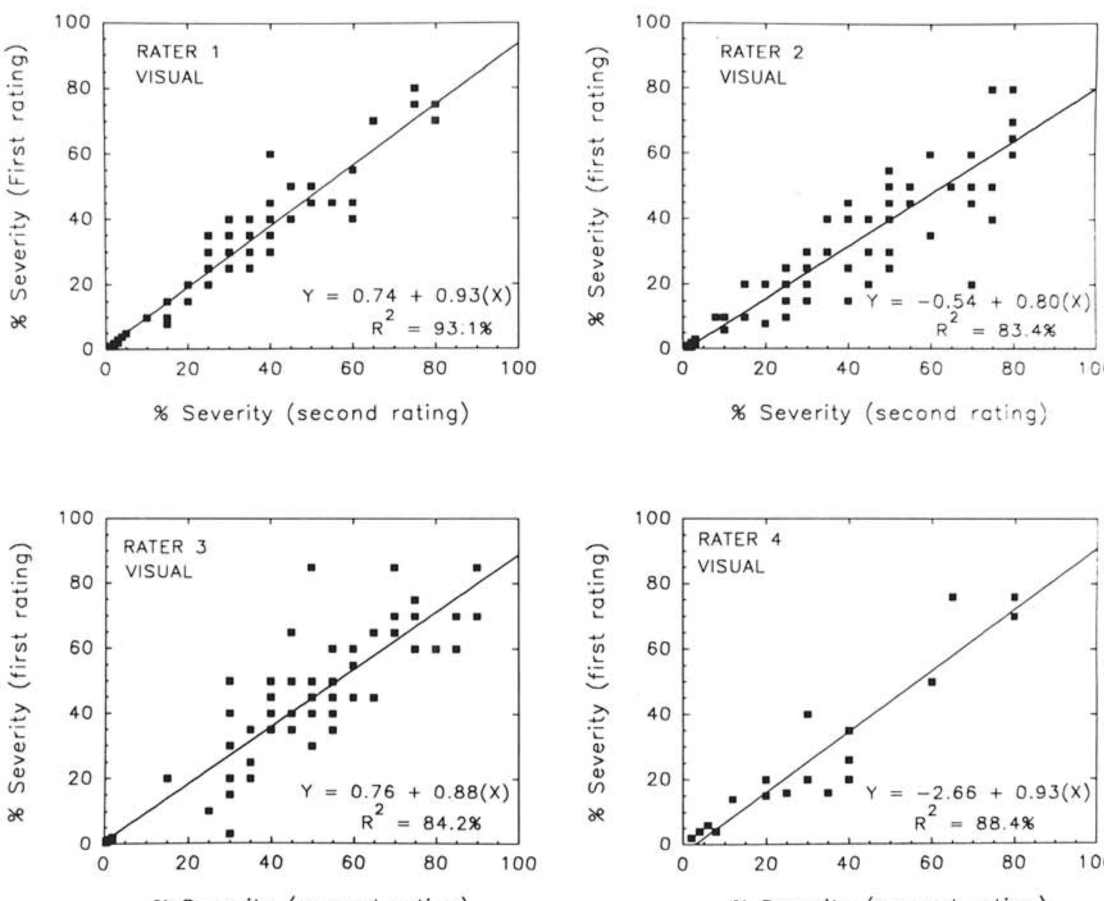

Fig. 2. Intra-rater repeatability of four raters visually assessing percent dollar spot severity of 80 bentgrass quadrats. For rater 4, $n = 20$ because only one replication was reassessed by this individual on 21 August 1990.

perceived different levels of dollar spot severity. For example, raters 1 and 2 (Fig. 5A) show better agreement at the low and high ends of the visual scale, whereas raters 1 and 3 (Fig. 5B) show less agreement as severity increases. Raters 1 and 4 (Fig. 5C) show differences that are largely independent of the level of severity. Slopes relating 800-nm radiometric assessments among raters ranged from 0.96 to 1.09, and two of the six inter-rater reliability comparisons had slopes significantly different from 1 (Table 1).

None of the y-intercepts for inter-rater reliability comparisons using the 600-nm wavelength band were significantly different from zero, whereas four of the six comparisons using the 800-nm wavelength and two of the comparisons using the visual method had intercepts significantly different from zero (Table 1).

Standard errors of the y-estimate for comparisons among raters using the visual method ranged from 6.67 to 11.95% (Table 2). Standard errors for the y-estimate ranged from 0.11 to 0.52% with the 800-nm wavelength and from 0.12 to 0.17% with the 600-nm wavelength. Coefficients of variation (*CV*) were highest with the visual assessment method (21.65 to 40.81%), whereas *CV*s were 0.59 to 1.41% for the radiometric method (Table 2).

Evaluation of assessment methods for accuracy. Because the acetate image-analysis method was used to provide an unbiased measure of dollar spot severity, this method was used to evaluate the accuracy of the visual versus the 600- and 800-nm radiometric assessment methods. Visual disease-severity assessments explained 78.7–96.8% of the variation in the acetate image-analysis disease-severity assessments (Table 3). The 800-nm radiometric assessments explained 79.0–86.1% of the variation in the acetate image-analysis assessments, whereas the 600-nm percent reflectance values accounted for 95.8–98.3% of the acetate image-analysis assessments. For each 1% increase in percent reflectance at 600 nm, dollar spot severity increased by 15.4% (mean slope for the four raters).

Effect of assessment method on time to assess dollar spot severity. Using the visual method, it took an average of 32 ±6.6 min for each rater to assess the 80 quadrats, whereas the radiometric method took approximately two-thirds the time (21.8 ±5.4 min).

DISCUSSION

The expenditure of time and money to develop, evaluate, and compare disease assessment methods can prevent serious flaws in data acquisition (2). Disease assessment methods should provide accurate and precise information that satisfies the goals and needs of the research (18,21). Campbell and Madden (2) defined precision as the lack of variation in disease estimates when the same sampling units are evaluated by other raters. This definition of precision excludes another potential source of error, i.e., the repeatability of individual raters. O'Brien and van Bruggen (14) noted that this type of variation (intra-rater variability) has rarely been measured by plant pathologists. One exception is a study by Shokes et al (18), which employed a test-retest procedure (16,19) to measure intra-rater error as affected by different assessment methods. Shokes et al (18) used the product-moment correlation coefficient (*r*) as an indicator of the overlap between test-retest disease assessments performed by the same rater. A high *r* value close to 1.0 is desirable. In the present study, we defined precision as a measure of both the intra-rater repeatability and inter-rater reliability of a disease assessment method. Intra-rater repeatability of an assessment method was defined as the linear relationship between repeated disease assessments of the same sampling units performed by the same rater or instrument. Inter-rater reliability

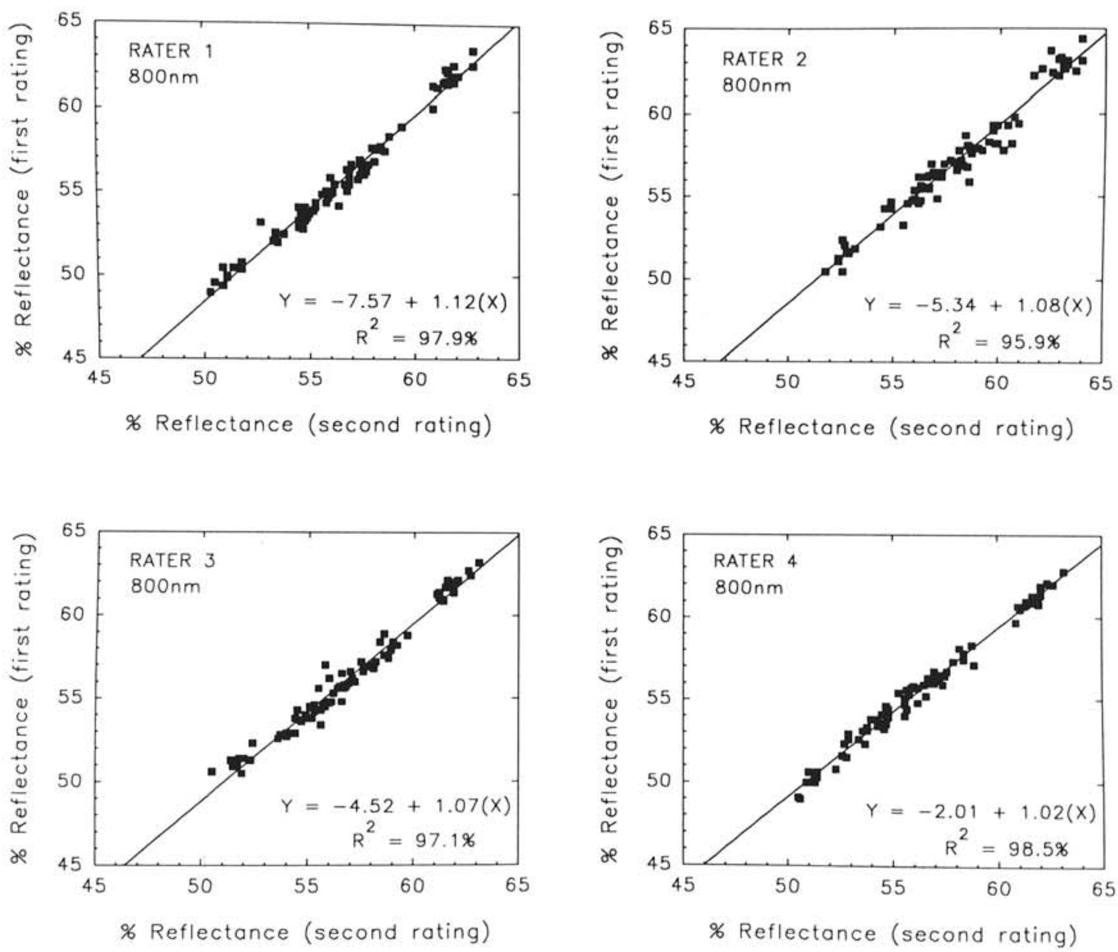

Fig. 3. Intra-rater repeatability of four raters measuring percent reflectance of sunlight (800 nm) from 80 bentgrass quadrats infected with different levels of dollar spot.

was defined as the linear relationship between assessments of the same sampling units performed by different individuals.

In theory, raters and methods that have the highest precision would be those in which the regression coefficients (slopes) measuring intra-rater repeatability or inter-rater reliability are closest to 1.0 and in which y-intercepts are closest to zero (16,20). Slopes deviating significantly from 1.0 indicate the presence of systematic bias, i.e., the difference between estimates made by one rater compared to estimates made by another rater increases as disease severity increases. Intercepts significantly differing from zero indicate the presence of a constant bias. Using regression analysis, both systematic and constant sources of bias (error) were detected among and within raters using the visual assessment method. Systematic and constant sources of bias, however, were not detected among and within raters using the 600-nm wavelength band to assess dollar spot severity. Using the 600-nm wavelength band, all slopes and intercepts were not significantly different from 1.0 for both intra-rater repeatability and inter-rater reliability evaluations. The radiometric method also had higher coefficients of determination than did the visual assessment method, which indicates that the amount of variation explained in one set of disease assessments by another set of assessments is higher with the radiometric method than with the visual method. Another desirable characteristic of an assessment method is that it has relatively smaller standard errors of the y-estimate, this being a measure of the degree of error associated with a predicted y-value (10,16,19). Using these criteria, percent reflectance of sunlight recorded in the 600- and 800-nm wavelength bands provided estimates with higher levels of precision compared to the visual disease assessment method.

Accuracy can be defined as the closeness of an estimate (disease assessment) to the truest value (13). O'Brien and van Bruggen (14) defined accuracy to be the disease scores provided by the originator of the assessment scale they were using. We defined accuracy as the percent disease-severity value for a selected quadrat as determined by the acetate image-analysis technique and

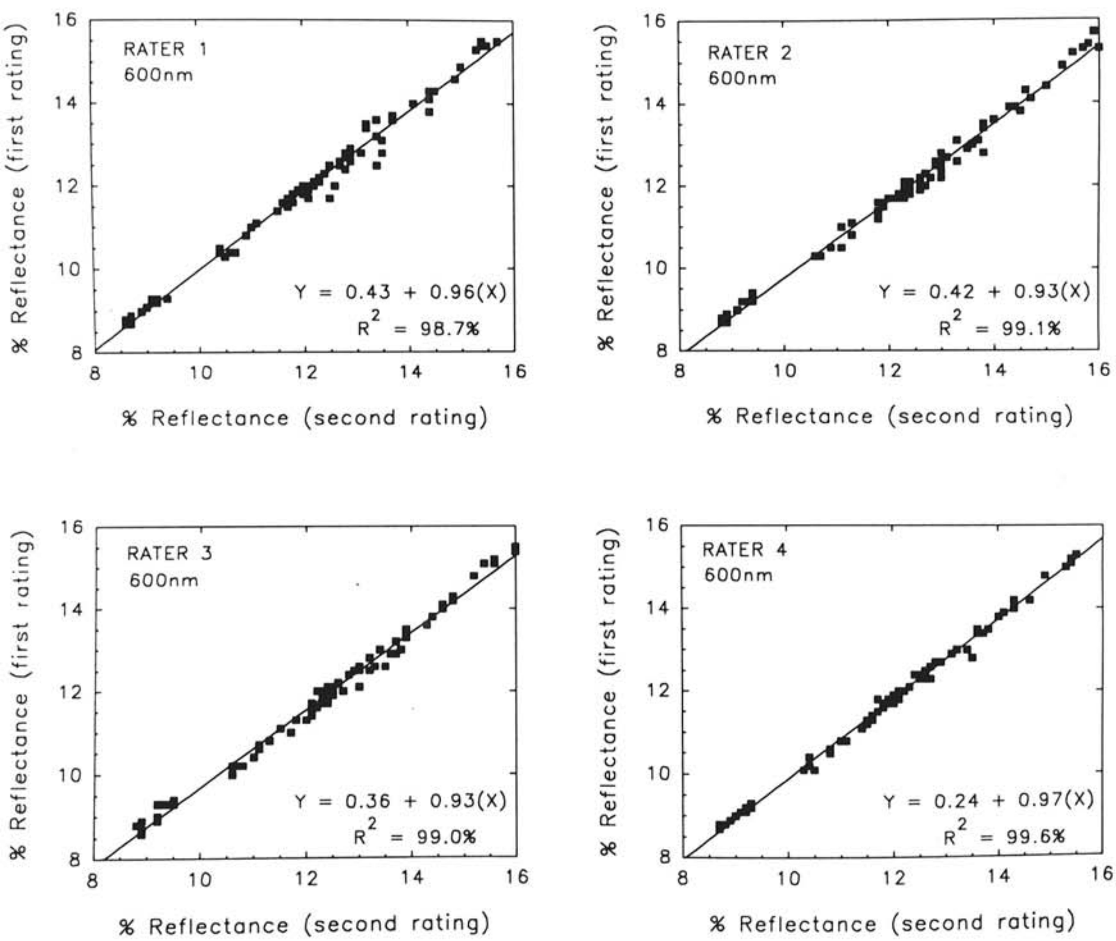

Fig. 4. Intra-rater repeatability of four raters measuring percent reflectance of sunlight (600 nm) from 80 bentgrass quadrats infected with different levels of dollar spot.

TABLE 1. Effect of disease assessment method on inter-rater reliability measured as the intercept, slope, and coefficient of determination (R^2) from linear regression equations relating disease assessments among raters

Comparison	Visual disease severity			Radiometer (800 nm)			Radiometer (600 nm)		
(*X*/*Y*)	Intercept	Slope	R^2 (%)	Intercept	Slope	R^2 (%)	Intercept	Slope	R^2 (%)
Rater 1/Rater 2	6.95* [a]	0.82* [b]	78.6	−4.08*	1.04	98.3	−0.06	1.01	99.1
Rater 1/Rater 3	2.33	0.74*	77.0	−5.30*	1.09*	98.4	−0.11	1.02	99.1
Rater 1/Rater 4	2.98	0.74*	89.2	−2.95*	1.06*	99.2	−0.20	1.03	99.6
Rater 2/Rater 3	−0.42	0.76*	70.4	−0.48	1.03	97.8	−0.02	1.01	99.4
Rater 2/Rater 4	0.60	0.77*	80.0	1.82	1.00	98.3	−0.07	1.02	99.4
Rater 3/Rater 4	6.01*	0.88*	86.4	2.80*	0.96	98.4	−0.03	1.01	99.5

[a] Y-intercepts followed by an * are significantly different from zero ($P \leq 0.01$).
[b] Slope parameters followed by an * are significantly different from a slope value of 1.0 ($P \leq 0.01$).

by determining the linear relationship between these values and the values obtained by visual or radiometric assessments. We found that radiometric measurements had a better relationship to the acetate image-analysis method than did the visual estimates of percent disease severity.

Correlation coefficients provide a measure only of the intensity of association between two variables (16). No cause-and-effect relationship can be inferred with correlation analysis. Correlation coefficients are independent of the units of measure (disease severity, percent reflectance, etc.) and, thus, are absolute or dimensionless quantities ranging from −1 to +1 (16). For many purposes, it is desirable to know both the degree to which two disease assessment methods (or raters) are associated and how they are related to one another in actual units of measure. For example, correlation coefficients among raters using the visual assessment method ranged from 0.87 to 0.94 ($P \leq 0.01$). Although these correlations are considered high, there were still systematic and constant sources of error undetected by this statistical method. Regression analysis provides additional information concerning the way in which one rater perceives a range of disease-intensity stimuli compared to other raters. For example, when visually assessing dollar spot severity, rater 1 consistently overestimated disease severity in relation to the other raters, as evidenced by the regression coefficients shown in Table 1. For each 1% increase in severity by rater 1, estimates by raters 2, 3, and 4 increased by 0.82, 0.74, and 0.74%, respectively. This systematic bias, which is not detected by correlation analysis, may be important in other pathosystems in which disease-management decisions are based on disease thresholds. Moreover, if assessment methods are used in which constant and systematic sources of variation are a potential hazard, then individual raters should be assigned to evaluate whole replications so these sources of error can be accounted for in the analysis. Although multispectral radiometer systems can be quite expensive, the use of radiometric methods to assess

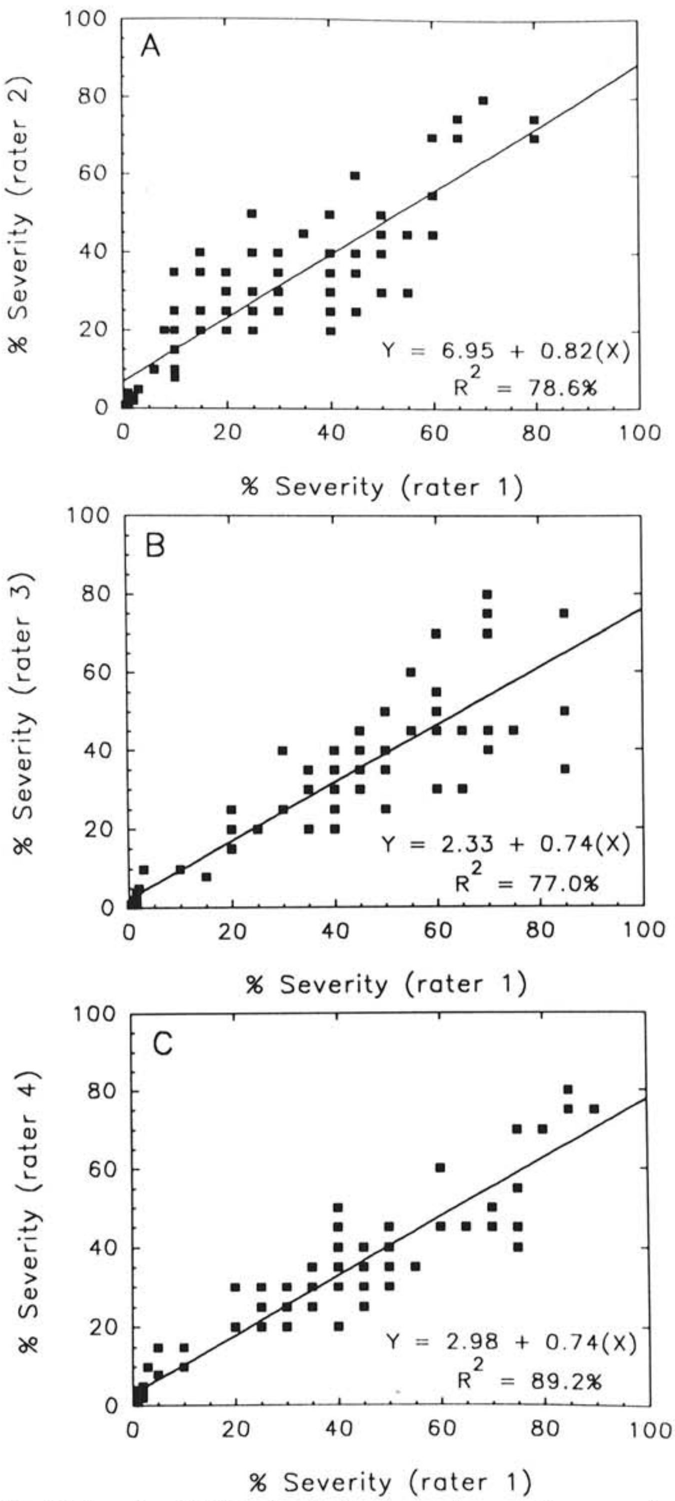

Fig. 5. Inter-rater reliabilities for visual estimates of percent disease severity performed by rater 1 compared to **A,** rater 2, **B,** rater 3, and **C,** rater 4.

TABLE 2. Effect of dollar spot assessment method on the standard error of the y-estimate and coefficient of variation among raters

Assessment method and comparison	SE of y-estimate (%)	Coefficient of variation (%)
Visual		
Rater 1 vs. Rater 2	9.39	30.47
Rater 1 vs. Rater 3	9.74	31.61
Rater 1 vs. Rater 4	6.67	21.65
Rater 2 vs. Rater 3	11.95	40.81
Rater 2 vs. Rater 4	9.87	33.71
Rater 3 vs. Rater 4	8.87	22.95
Reflectance (800 nm)		
Rater 1 vs. Rater 2	0.48	0.87
Rater 1 vs. Rater 3	0.47	0.84
Rater 1 vs. Rater 4	0.11	0.59
Rater 2 vs. Rater 3	0.52	0.91
Rater 2 vs. Rater 4	0.45	0.79
Rater 3 vs. Rater 4	0.42	0.74
Reflectance (600 nm)		
Rater 1 vs. Rater 2	0.17	1.40
Rater 1 vs. Rater 3	0.17	1.41
Rater 1 vs. Rater 4	0.12	0.98
Rater 2 vs. Rater 3	0.14	1.17
Rater 2 vs. Rater 4	0.14	1.18
Rater 3 vs. Rater 4	0.12	1.01

TABLE 3. Regression analyses to determine the accuracy of visual and percent reflectance (600 and 800 nm) disease assessments in relation to the actual level of disease severity as determined with an acetate image-analysis method

Method and rater	Intercept	Slope	R^2 (%)
Visual			
Rater 1	4.7	1.4	82.6
Rater 2	19.6	1.0	78.7
Rater 3	1.3	1.1	95.6
Rater 4	12.9	0.9	96.8
Reflectance (800 nm)			
Rater 1	609.0	−10.2	83.6
Rater 2	631.2	−10.3	86.1
Rater 3	727.3	−12.2	83.9
Rater 4	595.0	−10.1	79.0
Reflectance (600 nm)			
Rater 1	−134.8	15.0	98.3
Rater 2	−142.3	15.8	97.1
Rater 3	−135.3	15.3	96.4
Rater 4	−135.5	15.7	95.8

disease severity eliminates much of the bias present among and within raters.

O'Brien and van Bruggen (14) suggested that image-analysis systems, such as the one developed by Lindow and Webb (8), could be used to eliminate human bias in visual assessments in which leaf surfaces are fairly uniform and there is a clear distinction between diseased and nondiseased areas. The square-meter quadrats of bentgrass infected with dollar spot were analogous to large green, "uniform" leaf surfaces with spots and, thus, satisfy the requirement for surface uniformity. However, image analysis based on gray scales could not precisely distinguish between diseased and nondiseased turfgrass areas. The inability of image analysis of color slides to measure dollar spot severity precisely may be due to several factors, including 1) variance in lighting of the turfgrass quadrats during the period (1300–1600) when photographs were taken, 2) inconsistencies in film development, and 3) the subjective decision making required on the part of the machine operator as to the grey-scale values corresponding to diseased areas and those corresponding to healthy areas. The presence of chlorotic and necrotic leaves in a mosaic throughout the bentgrass green, even in nondiseased plot areas, may interfere with this subjective part of the image-analysis process. The use of acetate drawings, however, shifts the decision-making step from the machine operator to the persons making the acetates, and, therefore, dollar spot-diseased areas were differentiated from the background mosaic and were delineated more precisely with this method.

Researchers evaluating disease-control tactics aimed at reducing dollar spot severity presently assess disease intensity by reporting the frequency of dollar spots per plot (3,5) or by visually estimating percent dollar spot severity (2). Our study showed that radiometric assessments can be used to provide accurate and precise estimates of dollar spot severity in a shorter period of time compared to the visual percent disease-severity method. These advantages may outweigh the initial cost of the equipment and software (10,11). Because the CropScan radiometer measures incident as well as reflected radiation, it can provide accurate and precise measurements of percent reflectance during cloud cover as well as in full sunlight. However, reliable measurements can only be obtained within 2 h of solar noon because of the affect of sun angle on percent reflectance values.

LITERATURE CITED

1. Berger, R. D. 1980. Measuring disease intensity. Pages 28-31 in: Crop Loss Assessment; Proceedings of E. C. Stakman Commemorative Symposium. Misc. Publ. 7 Agric. Exp. Stn. Univ. Minn., St. Paul.
2. Campbell, C. L., and Madden, L. V. 1990. Introduction to Plant Disease Epidemiology. John Wiley & Sons, New York.
3. Dernoeden, P. H., and Schmitt, E. 1992. Dollar spot and algae control in bentgrass, 1991. Fungic. Nematicide Tests 47:284.
4. Gaunt, R. E. 1987. Measurement of disease and pathogens. Pages 6-18 in: Crop Loss Assessment and Pest Management. P. S. Teng, ed. The American Phytopathological Society, St. Paul, MN.
5. Gleason, M. L., and Christians, N. C. 1992. Evaluation of fungicides for control of dollar spot in Emerald bentgrass, 1991. Fungic. Nematicide Tests 47:285.
6. Horsfall, J. G., and Cowling, E. B. 1978. Pathometry: The measurement of plant disease. Pages 120-136 in: Plant Disease: An Advanced Treatise. Vol. 2. J. G. Horsfall and E. B. Cowling, eds. Academic Press, Inc., New York.
7. Kranz, J. 1988. Measuring plant disease. Pages 35-50 in: Experimental Techniques in Plant Disease Epidemiology. J. Kranz and J. Rotem, eds. Springer-Verlag, New York.
8. Lindow, S. E., and Webb, R. R. 1983. Quantification of foliar plant disease symptoms by microcomputer-digitized video image analysis. Phytopathology 73:520-524.
9. Nelson, E. B., and Craft, C. M. 1992. Suppression of dollar spot on creeping bentgrass and annual bluegrass turf with compost-amended topdressings. Plant Dis. 76:954-958.
10. Nutter, F. W., Jr. 1989. Detection and measurement of plant disease gradients in peanut with a multispectral radiometer. Phytopathology 79:958-963.
11. Nutter, F. W., Jr. 1990. Remote sensing and image analysis for crop loss assessment. Pages 93-105 in: Crop Loss Assessment in Rice. Int. Rice Res. Inst., Manila, Philippines.
12. Nutter, F. W., Jr., Gleason, M. L., Jenco, J. H., and Christians, N. C. 1991. Effect of visual, remote sensing, and image analysis assessment methods on intra-rater and inter-rater reliability estimates in the dollar spot-bentgrass pathosystem. (Abstr.) Phytopathology 81:1182.
13. Nutter, F. W., Jr., Teng, P. S., and Shokes, F. M. 1991. Disease assessment terms and concepts. Plant Dis. 75:1187-1188.
14. O'Brien, R. D., and van Bruggen, A. H. C. 1992. Accuracy, precision, and correlation to yield loss of disease severity scales for corky root of lettuce. Phytopathology 82:91-96.
15. Sanders, P., and Cole, H., Jr. 1986. In vivo fungicide screening on field-grown turfgrasses. Pages 244-247 in: Methods for Evaluating Pesticides for Control of Plant Pathogens. K. D. Hickey, ed. The American Phytopathology Society, St. Paul, MN.
16. Schmidt, M. J. 1975. Understanding and Using Statistics: Basic Concepts. D. C. Heath and Co., Lexington, MA.
17. 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.
18. Shokes, F. M., Berger, R. D., Smith, D. H., and Rasp, J. M. 1987. Reliability of disease assessment procedures: A case study with late leafspot of peanut. Oleagineux 42:245-251.
19. Smiley, R. S., ed. 1987. Compendium of Turfgrass Diseases. The American Phytopathology Society, St. Paul, MN.
20. Steel, R. G. D., and Torrie, J. H. 1980. Principles and Procedures of Statistics. McGraw Hill, New York. 633 pp.
21. Watson, G., Morton, V., and Williams, R. 1990. Standardization of disease assessment and product performance reporting: An industry perspective. Plant Dis. 74:401-402.