Illusions in Visual Assessment of Stagonospora Leaf Spot of Orchardgrass

R. T. Sherwood, C. C. Berg, M. R. Hoover, and K. E. Zeiders

Research plant pathologist, research agronomist, research associate, and research plant pathologist, respectively, U.S. Regional Pasture Research Laboratory, USDA-ARS, University Park, PA 16802 (also adjunct professor of plant pathology, adjunct assistant professor of plant breeding, support scientist, and adjunct instructor, respectively, Pennsylvania State University, University Park).

Contribution 8203 from the U.S. Regional Pasture Research Laboratory, USDA-ARS, University Park, PA 16802.
Accepted for publication 6 July 1982.

ABSTRACT

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.

Two groups of five experienced scorers estimated percent leaf area spotted by matching orchardgrass leaves infected by *Stagonospora arenaria* against published area diagrams. Actual area of spotting was determined by weighing paper replicas of photographs. Nine of the scorers usually overestimated spotted areas. Overestimation was greatest when infected area was smallest, often being two to three times the actual area, and it decreased as the infected area increased. When two leaves had equal total spotted areas, the leaf having substantially more (but smaller) spots was usually scored higher. Regression analysis showed that overestimation was inversely proportional to the natural logarithm of the disease area for all scorers and also directly proportional to the number of spots for five scorers. There were significant group, scorer within group, leaf, and scorer × leaf effects. If visual area assessments overestimate disease more seriously at low disease incidence, their use in equations for predicting disease increase or in equations for yield reduction will result in underestimation of the true rate or amount of loss. The coefficients of variation ($\bar{x} = 21.5\%$) indicated that visual estimates were not highly precise.

Additional key words: *Dactylis glomerata*, epidemiology, disease loss, modeling, forage crops disease, Gompertz transformation, disease assessment, Vanderplank equation.

Accurate measurement of disease severity is essential for quantitative prediction of disease progression or yield reduction and for elucidating inheritance of resistance. The impracticality of counting and measuring individual lesions for assessing leaf spot in large plant populations motivated scientists to rely upon other more rapid visual methods. The simplest procedures involve scoring severity on a graded scale of arbitrary values, eg, 1–9. Because disease scoring apparently follows the Weber-Fechner law, which states that the response of an organism to a stimulus is a linear function of the log of the stimulus, graded readings are often converted to percentages on a logarithmic curve (5,6). A somewhat more sophisticated approach uses reference diagrams that depict lesions covering known percentages of area. The sets of diagrams are approximately logarithmic. Intervals between successive disease proportions depicted become progressively wider as the disease proportion increases. Diseased area is estimated by matching infected plant parts with the diagrams (7,12,16), and intermediate values are interpolated.

A recent treatise on plant pathology contends the eye is an objective "photocell" for measuring disease intensity provided that treatment identity is unknown (6). It claims close agreement among observers for several diseases. However, psychological studies provide evidence that the eye often grades inaccurately (1,3). Krantz (9,10) showed that visual ratings made without reference diagrams may overestimate actual area of spotting. For the most part, critical studies are lacking to test the accuracy of visual methods against independent, reliable, mechanical measurements.

In our work with the purple leaf spot of orchardgrass (*Dactylis glomerata* L.), caused by *Stagonospora arenaria* Sacc., we frequently encountered significant discrepancies among scores assigned by different operators using standards. The disease is characterized by scattered, circular or elongate, dark spots of variable size, frequency, and distribution.

The purpose of the experiment reported here was to investigate our lack of agreement in scoring and to determine whether there are systematic errors in the area diagram matching method for estimating purple leaf spot.

MATERIALS AND METHODS

Plants were maintained in the greenhouse and inoculated with *S. arenaria* as previously described (13,17). About 14 days after inoculation, 8-cm lengths of infected leaves were removed and placed in test tubes with Carnoy's solution (ethanol:acetic acid, 3:1). The solution removed chlorophyll but did not alter the leaf spots. This permitted repeated assessment and measurement without dimensional changes.

Measurement of actual area. Groups of five leaves were photographed on 3-mm color transparency film. The transparencies were projected on paper at an enlargement of ×20 actual leaf length. The outlines of each leaf and spot were drawn. The edge of the spot was drawn at the interface between apparently healthy tissue and the lightly brown pigmented outer margin of the lesion. The percentage of leaf area covered by spots was calculated from the weights of the paper replicas. Repeated determinations from individual leaves gave close agreement. For example, five determinations of one leaf gave 3.136 ± 0.061%.

Visual estimation of area. The percent area covered by spots was visually estimated using the method of James (7). Leaves from 40 plants were scored by two groups of five scorers using key diagrams 1.2, 1.3, and 1.4 (7). These depict 1, 5, 15, 25, and 50% coverage of cereals by rust or powdery mildew. The diagram spot sizes and shapes closely resembled purple leaf spot of orchardgrass. Group I scorers (scorers 1–5) had extensive experience using this method on cereals. Group II scorers (scorers 6–10), from a different institution, were experienced scoring forage crop leaf diseases by various systems. The groups scored the leaves on separate dates, but in the same room under identical conditions. Double-blind procedure was used to present 40 leaves, one at a time, in unmarked petri dishes to each scorer in three randomized replications. There were no identifying marks on leaves and no communication among scorers.

Statistical analyses. Data were analyzed by two-way and three-way analyses of variance, courtesy of R. R. Hill, Jr., U.S. Regional

---

Pasture Research Laboratory, University Park, PA. Differences of means were tested by a Duncan's modified (Bayesian) least significant difference test program (DLSD) of The Pennsylvania State University Computation Center. Correlation coefficients were calculated and regression analyses were performed by using SAS procedures (4).

RESULTS

Estimation of spotted area. The 40 leaves had a range of 0.46–19.25% of their area actually covered by spots (Fig. 1). Group I visual estimates exceeded the actual value for each leaf. Group II estimates for 34 leaves were intermediate between the actual area and Group I estimates, and for six leaves were less than the actual area. Analysis of variance showed that leaf, scorer, and scorer × leaf effects were highly significant within each group ($P > 0.01$). Correlations between scorers and actual area were 0.922–0.967 and among scorers were 0.849–0.985.

When two leaves had similar actual total spot areas, but substantially different numbers of spots, usually the visual estimate for the leaf with the greater number of spots exceeded that for the leaf with fewer (but larger) spots (Fig. 1). Accordingly, we tested various regression models to determine the influence of actual spot area and spot frequency on visual estimates. The independent variables tested were actual area, A; $\sqrt{(A)}$; $\log_e 100\ A$; number of spots, N; $\sqrt{(N)}$; and $\log_e N$. (Values for area were multiplied by 100 to avoid negative log at low percentages of infection.) The dependent variables were visually estimated diseased area, Y; and $\sqrt{(Y)}$. Coefficients of determination (R^2) were calculated for all combinations of variables for each scorer (4).

The best single-factor model for all scorers was $Y = a + bA$ in which a = intercept and b = slope of A. The R^2 values ranged from 0.849 to 0.935 among scorers. The most satisfactory two-factor model was $Y = a + bA + cN$ in which c = slope of N. For this two-factor model, R^2 values ranged from 0.879 to 0.957 among scorers and were slightly to substantially higher than R^2 for the one-factor model for all except scorer 10. The slope due to N was significantly different from 0 for all scorers except 6 and 10 (Table 1). Scorer and scorer × slope effects were significant, but groups were not significantly different; this is illustrated in Fig. 2. Square root and natural logarithmic transformations did not improve the models. Examination of the residuals (4) confirmed the essential linearity of the models.

Ratio of visual estimate to actual area. Visual rating usually overestimated, but sometimes underestimated, area (Fig. 1). We tested whether the error in estimation was random throughout the range of areas studied. The ratio (estimated area):(actual area) was calculated for each score for each leaf. If there were no errors, the ratio of (estimated area):(actual area) would be one at all levels of disease. However, as shown in Fig. 3, the ratio was usually greater than one and tended to be larger at low levels of infection than at high levels of infection. Therefore, we postulated that the ratio (estimated area):(actual area) was influenced by A or N or their square root or log transformations. Regression models were tested as above.

The most satisfactory single-factor model was $Y = a + b \log_e 100A$ ($R^2 = 0.090$ to 0.595). The slope was significant ($P > 0.05$) for scorers 2 and 9, and highly significant ($P > 0.01$) for the other scorers. The two-factor model $Y = a + b \log_e 100A + cN$ gave substantially improved estimates for all scorers, with $R^2 = 0.255$–0.753 (Table 2). In this model the slope due to N was significant for scorers 1, 2, 4, 7, and 9 and slope due to $\log_e 100A$ was highly significant for all scorers. Group I had significantly greater overestimation than Group II. Scorers within groups varied. Thus, for all scorers, the tendency to overestimate area was inversely proportional to the natural log of the area; for some scorers overestimation was also directly proportional to number of spots.

Precision of visual estimates. To test the precision (ie,

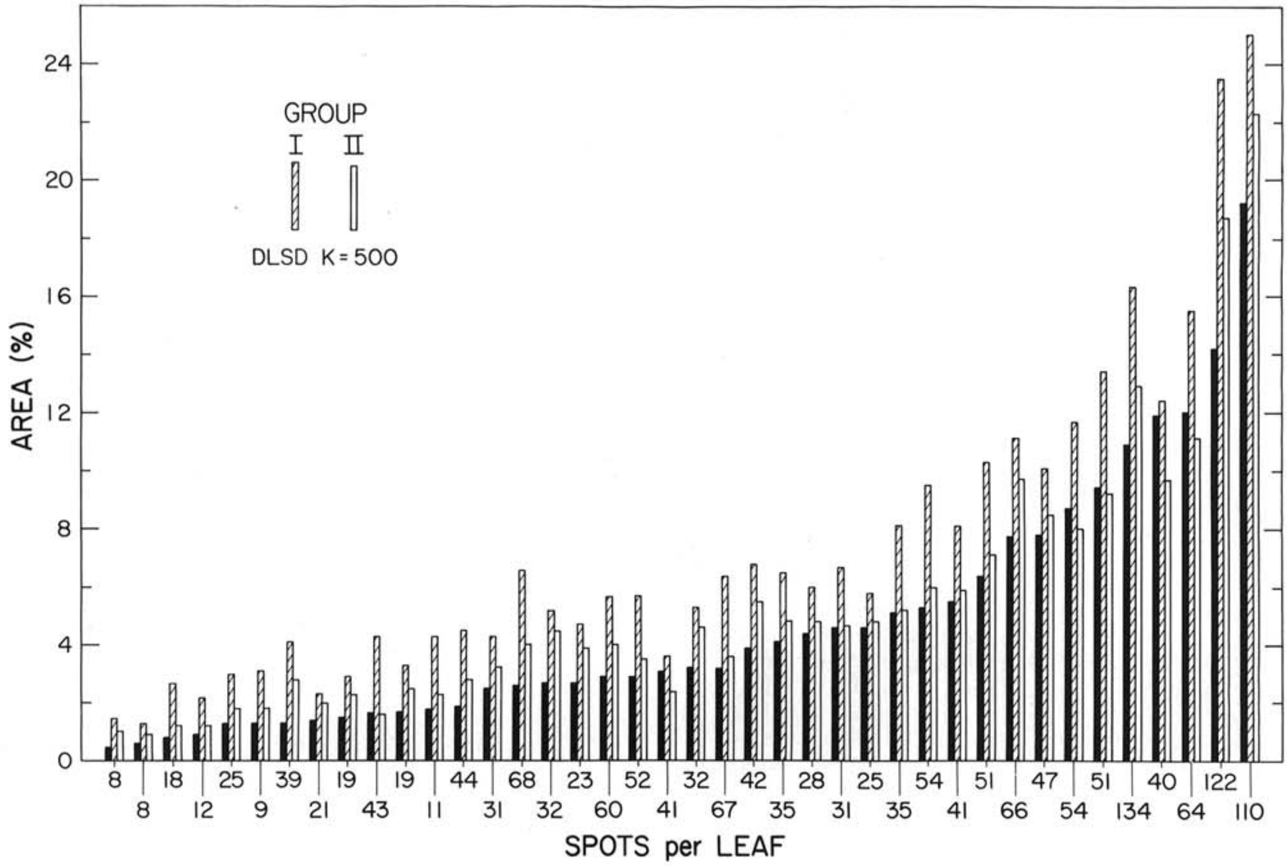

Fig. 1. Percentage of leaf area covered by purple leaf spot on 40 orchardgrass leaves, and number of spots per leaf. Each group of three bars represents one leaf, and leaves are ranked by actual area. Solid bars = actual area. Cross-hatched bars = estimated area, mean of five scorers in Group I. Open bars = estimated area, mean of five scorers in Group II. Duncan's least significant difference (DLSD) K = 500; Group I = 2.3% and Group II = 2.2%.

reproducibility of measurement) of visual estimation, we calculated the coefficient of variation ($C = 100\ \sigma n - 1/\bar{x}$) for each set of three determinations on each leaf by each scorer. Analysis of variance of C did not show significant differences between groups of scorers or among scorers within groups (range, 15.7–27.8%). There were significant differences among leaves, but the differences were not associated with lesion number or size. The coefficients of variation were not significantly correlated with actual area of spotting. The mean coefficient of variation for the experiment was 21.5%.

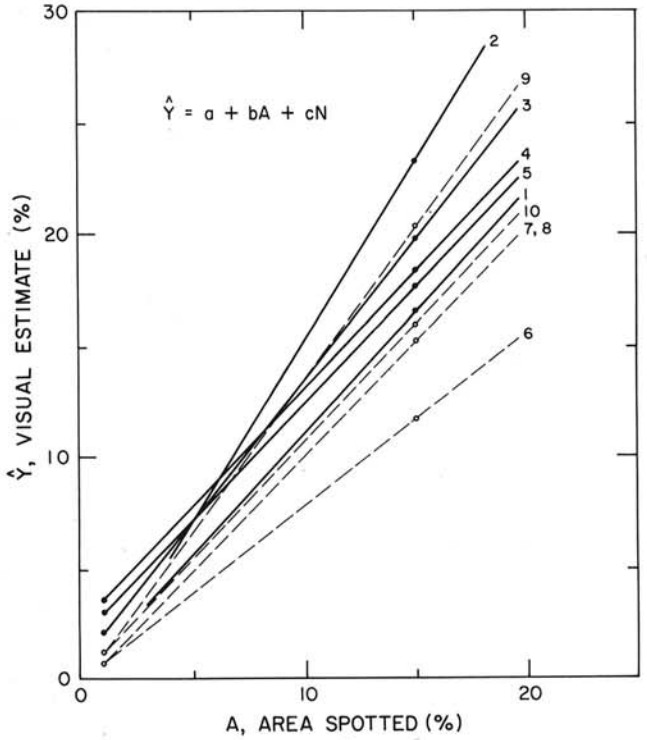

Fig. 2. Regression of visual estimates ($\hat{Y}$) of percent area spotted vs actual area (A) spotted for each of 10 scorers. Scorers 1–5 (solid lines) from Group I, 6–10 (broken lines) from Group II. See Table 1 for values of a, b, and c and regression analysis. Points drawn assuming that number of spots (N) = 15 at 1% A, and $N = 75$ at 15% A, in accord with trends in Fig. 1.

DISCUSSION

The eye is readily deceived when judging geometric figures (1,3). Shape, orientation, shading, surrounding figures, and personal traits enter into perception. Numerous illusions have been defined, but their psychological and psychophysical bases are poorly understood.

Ten experienced scorers, using standard area diagrams, showed general agreement in ranking leaves for percent area occupied by spots ($r = 0.849–0.985$). However, most scorers overestimated the actual amount of spotting. Visual overestimation was greatest at lowest levels of infection, often being two to three times the true value. This bias was inversely proportional to the natural logarithm of the actual area. For five scorers the bias was also proportional to the number of spots.

The results indicate that two kinds of illusion influence visual

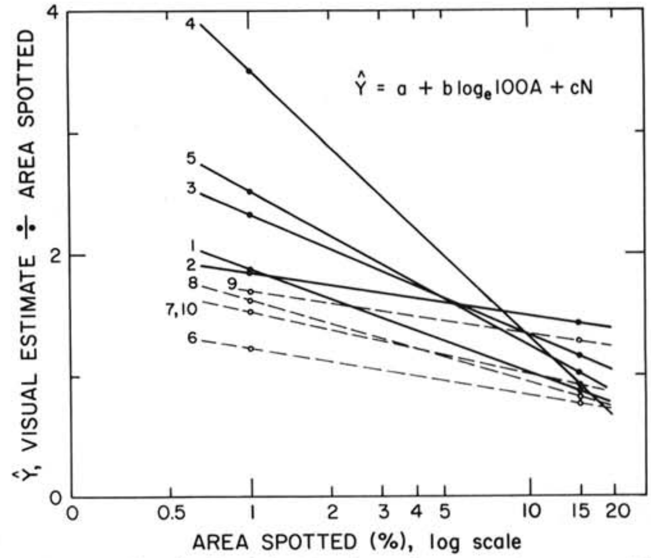

Fig. 3. Regression of ratio of visually estimated area:actual area spotted ($\hat{Y}$) vs actual area (A) spotted for each of 10 scorers. Scorers 1–5 (solid lines) from Group I, 6–10 (broken lines) from Group II. See Table 2 for values of a, b, and c and for regression analysis. Points drawn assuming that number of spots $N = 15$ at 1% A, and $N = 75$ at 15% A, in accord with trends in Fig. 1.

TABLE 1. Regression analysis of visual estimates of percent spotted leaf area for two groups of five scorers for the model $Y = a + bA + cN$[a]

Group	Scorer	a	b	c	R^2
I	1	−0.090	0.931**	0.0348**	0.939
	2	−1.446**	1.279**	0.0736**	0.957
	3	0.821*	1.154**	0.0256*	0.942
	4	1.529**	0.690**	0.0879**	0.944
	5	1.821**	0.948**	0.0227*	0.940
II	6	0.370	0.700**	0.0127	0.940
	7	−0.988*	0.728**	0.0710**	0.934
	8	−0.202	0.898**	0.0296*	0.911
	9	−0.571	1.131**	0.0450**	0.953
	10	0.518	1.029**	0.0006	0.879
Sources of variation			d.f		Mean square
Group (G)[b]			1		14.90
Scorers within group (S)			8		13.04**
Area by paper weight (A)			1		2,579.28**
Number of spots (N)			1		214.03**
Interactions					
$A \times G$			1		7.62**
$N \times G$			1		9.65**
$A \times S$			8		12.36**
$N \times S$			8		10.87**
Error			370		1.82

[a] Y = predicted visual estimate of area spotted; a = intercept; b and c = slopes; A = actual percent area spotted; N = number of spots per sample. Based on data from 40 leaves. See regression lines in Fig. 2.

[b] Scorers within groups were used to test significance of groups. Asterisks * and ** indicate statistical significance at $P = 0.05$ and 0.01, respectively.

judgment of this material. These operate even when reference diagrams are used. One illusion is that the spots occupy more area than they actually do. This illusion is more pronounced at lowest levels of spotting. Pathologists early realized that when disease occupies less than 50% of total area the eye focuses on the diseased tissue (6,16).

The second illusion involves perception of the components of total area, eg, size and number of spots. Within pairs of leaves having similar total areas in spots, the leaf having a substantially greater number of spots was often perceived as having a greater total area of spots. The eye apparently discriminates among frequencies more readily than among sizes. A given area of spotting can be doubled either by doubling the number of spots or by expanding the diameter of each spot by 1.414. A doubling of spot count makes more impression than a change of 1.414 × diameter.

Agreement among scorers was not as impressive as some scientists have indicated (6). In both experiments with standards, there were significant differences among scorers within groups and significant scorer × leaf interactions. Scorers from one institution consistently had higher scores and larger errors in estimation. The two scoring panels were conducted under identical conditions. We believe the differences in groups reflect differences in training. Scorer II-6, with scores lower than all others, previously had been trained to consciously score lower than perception indicated. The important points revealed by this study are that at low levels of disease typical of those in the field, all scorers, regardless of training, overestimated the area of infection and that overestimation decreased with increasing infection.

Smith et al (14) estimated tomato leaf areas infected by *Cladosporium fulvum* without reference to disease keys. In agreement with our findings, their estimates exceeded the actual areas of infection. If the correction factors illustrated in their Fig. 1B are used to calculate the proportion of overestimation at each level of infection, the percent of overestimation was greatest at lowest levels and declined with increasing infection, as in our experiment. Kranz (9) obtained a similar result in experiments rating paper models of spotted apple leaves.

A situation in which visual scores overestimate intensity, and overestimation decreases with increased intensity and decreased number of spots, leads to some interesting practical consequences.

TABLE 2. Regression analysis of the ratio (estimated area):(actual area spotted) for two groups of five scorers for the model $Y = a + b \log_e 100\,A + cN$[a]

Group	Scorer	a	b	c	R^2
I	1	4.153**	−0.519**	0.0066*	0.582
	2	3.536**	−0.402**	0.0122**	0.361
	3	4.810**	−0.551**	0.0050	0.492
	4	9.870**	−1.452**	0.0221**	0.753
	5	5.596**	−0.683**	0.0061	0.603
II	6	2.116**	−0.202**	0.0015	0.255
	7	3.465**	−0.451**	0.0100**	0.433
	8	3.411**	−0.396**	0.0038	0.561
	9	2.966**	−0.306**	0.0070**	0.315
	10	2.738**	−0.261**	0.0014	0.333
Sources of variation			d.f.		Mean square
Group (G)[b]			1		10.545*
Scorers within group (S)			8		1.979**
Log$_e$100A (A)			1		39.796**
Number of spots (N)			1		8.306**
Interactions					
$A \times G$			1		5.778**
$N \times G$			1		1.106**
$A \times S$			8		1.364**
$N \times S$			8		0.471**
Error			370		0.128

[a] Y = predicted ratio (estimated area):(actual area); a = intercept; b and c = slopes; A = actual percent area spotted; N = number of spots per sample. Based on data from 40 leaves. See regression lines in Fig. 3.

[b] Scorers within groups were used to test significance of groups. Asterisks, * and **, indicate statistical significance at $P = 0.05$ and 0.01, respectively.

TABLE 3. Effect of diminishing overestimates of disease proportion on parameters used in predicting rate of disease increase; calculated from data for scorer 1 in Table 2 and Fig. 3

	Disease proportion (x_i) at indicated time (t_i)				Gompits (Y) at indicated time				Calculated expression for disease increase equation[b]	
	Actual		Estimated[a]		Actual		Estimated			
Interval measured	t_1	t_2	t_1	t_2	t_1	t_2	t_1	t_2	Actual	Estimated
Change from 1% area at t_1 to 4% area at t_2	0.01	0.04	0.0186	0.0548	−1.527	−1.169	−1.382	−1.066	1.417	1.118
$t_2 - t_1$					0.358		0.316			
Estimated (as % of actual)					88.3%				78.9%	
Change from 4% area at t_1 to 16% area at t_2	0.04	0.16	0.0548	0.1360	−1.169	−1.606	−1.066	−0.691	1.520	0.999
$t_2 - t_1$					0.563		0.375			
Estimated (as % of actual)					66.6%[b]				65.7%	

[a] From scorer 1 regression line; Fig. 3 multiplied by percent area.

[b] From references 15 and 16, $\ln(x_2/(1-x_2)) - \ln(x_1/(1-x_1))$.

When selection in a breeding scheme is based on perceived area of infection, there may be a tendency to discard plants having numerous small lesions and save those with fewer but somewhat larger lesions of the same total area. Paradoxically, the plants with smaller spots could be better sources of rate-limiting resistance. Estimates of disease intensity are integral to determining the relation between disease and crop loss. Use of exaggerated visual estimates as the independent variable in critical-point models for predicting loss (11) would undervalue the impact of disease. Shifting overestimates used in multiple-point or response-surface models would also provide erroneous solutions.

The equation (15,16) $r = [1/(t_2-t_1)]\ [\ln(x_2/(1-x_2)) - \ln(x_1/(1-x_1))]$ in which x is the proportion of tissue affected, is commonly used to calculate the apparent infection rate. If the degree of overestimation increases at lower values of disease, estimates of r will be decreased because of grossly overestimated x_1 (disease proportion at t_1), even though x_2 is also overestimated, but not as much. This can be illustrated by comparing the calculations of $(\ln(x_2/(1-x_2)) - \ln(x_1/(1-x_1)))$ in the above equation for actual vs estimated purple leaf spot. Table 3 shows values for this expression calculated for scorer 1 using data of Table 2 and assumptions of Fig. 3. From Table 3, it is seen that for a fourfold increase of actual disease from 1% at t_1 to 4% at t_2, the value of the expression from visual estimates = 1.118 and from actual measurement of area is 1.417. Estimators would assign an r value of only 78.9% of the actual rate. If the disease is 4% at t_1, and 16% at t_2 (a fourfold increase), the discrepancy in r values is greater; estimated r is 65.7% of actual. For these examples it was assumed that increase in disease came from increase both in lesion size and lesion number. This is based on the observation (*unpublished*) that lesion size shows a positive correlation with lesion frequency in this disease. If, for the sake of argument, we assume that between t_1 and t_2 there is a fourfold increase in lesion size, and no change in lesion number, the discrepancy between actual and estimated r becomes markedly greater (*unpublished*).

Berger (2) recently showed that a Gompertz transformation gave more accurate estimates of epidemic rate than logistic equations for several leaf spotting diseases. The data presented in Table 3 indicate decreasing overestimations with increasing disease could lead to underestimation in equations by using the Gompertz transformation. In the 1–4% increase example, the underestimation of rate by Gompertz transformation would be less serious than in the logistic model (88.3 vs 78.9%). In the 4–16% example, errors would be similar (66.6 vs 65.7%). Such misjudgments are unacceptable in quantitative studies. The wide differences among scorers working together and using standards demonstrates that visual scoring may not be a satisfactory quantitative tool.

The most heavily infected leaf in this study had about 20% area covered. Most purple leaf spot infections in the field cover less area, although some leaves may have approximately 20–35% coverage. This seems to be the maximum possible coverage by individual spots (data not shown). The study was conducted within the range of reactions normally encountered. Kranz (10) noted that maximum disease severity of two other leaf spots was usually less than 13.8% and rarely exceeded 37% on individual leaves.

Kranz (10) considered that the high levels of disease severity published may partly reflect the psychological error of expectation. Koch and Hau (8) demonstrated that scorers tend to prefer certain values (eg, 1, 5, and 10%), so that scores are "knotted" at these values. Our experiments did not test the errors of expectation or clustering, although it is possible that in addition to the illusions detected by our tests, these factors also influenced the scores.

The rather sizeable variability of determination ($C = 21.5\%$) indicates visual estimates were not precise. Unlike accuracy, precision was not shown to be influenced by area of spotting, frequency of spotting, or scorer.

One should obtain background information from accurate quantitative measurements appropriate to a specific disease before applying grading, Horsfall-Barrett conversions, diagram matching, or any other visual assessment method. It is clear that for purple leaf spot, visual rating may be useful for ranking of plants varying widely in disease severity, but it is unreliable for quantifying disease progress and yield loss.

Substantial improvement of accuracy will require modification of existing methods or development of new technology. Training alone will not eliminate illusion. Although all scorers showed area-dependent overestimation, the scorers varied in degree of error and in response to spot number. Therefore, no correction factors for regression equations could be derived that would apply universally. The area diagram matching method might be improved by using additional diagrams emphasizing low levels of infection and different spot sizes in the manner of Petersen et al (12). We are presently evaluating computerized analysis of video images as a potential system for measuring purple leaf spot.

LITERATURE CITED

1. Baird, J. C., and Elliott, N. 1978. Fundamentals of Scaling and Psychophysics. John Wiley & Sons, New York. 287 pp.
2. Berger, R. D. 1981. Comparison of the Gompertz and logistic equations to describe plant disease progress. Phytopathology 71:716-719.
3. Coren, S., and Girgus, J. S. 1978. Seeing is Deceiving: The Psychology of Visual Illusions. Wiley, New York. 255 pp.
4. Helwig, J. T., and Council, K. A. (eds.). 1979. SAS User's Guide. Statistical Analysis Institute, Inc., Raleigh, NC. 494 pp.
5. Horsfall, J. G., and Barratt, R. W. 1945. An improved grading system for measuring plant disease. (Abstr.) Phytopathology 35:655.
6. Horsfall, J. G., and Cowling, E. B. 1978. Pathometry: The measurement of plant disease. Pages 119–136 in: Plant Disease: An Advanced Treatise. Vol. II. J. G. Horsfall and E. B. Cowling, eds. Academic Press, New York.
7. James, W. C. 1971. An illustrated series of assessment keys for plant disease: Their preparation and usage. Can. Plant Dis. Surv. 51:39-65.
8. Koch, H., and Hau, B. 1980. Ein psychologischer Aspekt biem Schätzen von Pflanzenkrankheiten. Z. Pflanzenkrankh. Pflanzenschutz 87:587-593.
9. Kranz, J. 1970. Schätzklassen für Krankheitsbefall. Phytopathol. Z. 69:131-139.
10. Kranz, J. 1977. A study on maximum severity in plant diseases. Pages 169–173 in: Trav. ded. à G. Viennot-Bourgin.
11. Main, C. E. 1977. Crop destruction—the raison d'être of plant pathology. Pages 55–78 in: Plant Disease: An Advanced Treatise. Vol. I. J. G. Horsfall and E. B. Cowling, eds. Academic Press, New York.
12. Peterson, R. F., Campbell, A. B., and Hannah, A. E. 1948. A diagrammatic scale of estimating rust intensity on leaves and stems of cereals. Can. J. Res. C26:496-500.
13. Sherwood, R. T. 1982. Pathological anatomy of *Dactylis glomerata* infected by *Stagonospora arenaria*. Phytopathology 72:146-150.
14. Smith, P. M., Last, F. T., Kempton, R. J., and Gisborne, J. H. 1969. Tomato leaf mould: Its assessment and effects on yield. Ann. Appl. Biol. 63:19-26.
15. Vanderplank, J. E. 1963. Plant Diseases: Epidemics and Control. Academic Press, New York. 349 pp.
16. Zadoks, J. C., and Schein, R. D. 1979. Epidemiology and Plant Disease Management. Oxford University Press, New York. 427 pp.
17. Zeiders, K. E., Sherwood, R. T., and Berg, C. C. 1974. Reaction of orchardgrass cultivars to purple leaf spot caused by *Stagonospora arenaria*. Crop Sci. 14:205-208.