Comparing Digital and Visual Evaluations for Accuracy and Precision in Estimating Tall Fescue Brown Patch Severity

Virginia R. Sykes,* Brandon J. Horvath, Scott E. Warnke, Shawn D. Askew, Anton B. Baudoin, and James M. Goatley

ABSTRACT
Brown patch (*Rhizoctonia solani* Kuhn), a destructive disease of tall fescue (*Festuca arundinacea* Schreb.), is typically evaluated visually. The subjectivity of visual evaluations may be reduced using technology like digital image analysis (DIA). This study compared DIA and visual evaluations for accuracy and precision of brown patch ratings of glasshouse grown tall fescue plants. Across four experiments, 112 plants were inoculated with *R. solani*. Disease was rated visually and using DIA-WP (digital image analysis whole plant canopy). In two experiments, disease evaluations were replicated using three images and three visual evaluations per pot. Absolute error was calculated as the difference between actual disease severity ([calculated using an individual leaf DIA method previously quantified as highly predictive of actual brown patch disease severity on tall fescue (*r*² = 0.99)] and DIA-WP and visual evaluations, respectively. Standard deviations within repeated measures were also calculated. A mixed-model ANOVA was used to determine differences (*P* < 0.05) in mean absolute error and mean standard deviation by method, disease range, and method by disease range. Disease ranged from 0 to 100%. Mean absolute error did not differ between methods but did by disease range, exhibiting a bell-shaped curve from 0% to 100% disease severity. Mean standard deviation exhibited significant method by disease range interaction. Mean standard deviation did not differ across the disease range within DIA-WP evaluations but did across the disease range within visual evaluations. The more consistent precision of DIA across the disease range could reduce variability in brown patch evaluations of tall fescue.

BROWN PATCH, caused by the fungus *Rhizoctonia solani* Kuhn (Piper and Coe, 1919), can be a highly destructive disease on tall fescue (*Festuca arundinacea* Schreb.) grown in the transition zone. Visual evaluations are commonly used in turfgrass cultivar development, both in the initial stages, where breeders may use visual assessment to screen large numbers of individual turfgrass genotypes for disease resistance under controlled conditions (Beirn et al., 2015; Curley et al., 2005; Elliot, 1995; Fu et al., 2005), and in the latter stages, where cultivars are evaluated for brown patch resistance under field conditions by multiple evaluators across many locations. Visual evaluators typically report either the percentage of diseased tissue or use one of several variations of disease rating scales. Visual evaluation is a preferred disease evaluation method because it is quick, easy, and inexpensive. However, in visual disease evaluations there may be compromises to both precision, defined as “the closeness to each other of repeated measurements of the same quantity” and accuracy, defined as “the nearness of a measurement to the actual value of the variable being measured” (Horst et al., 1984; Nutter Jr. et al., 1993; Zar, 1999).

*Corresponding author (vsykes@utk.edu). Assigned to Associate Editor Michael Richardson.

Abbreviations: DIA, digital image analysis; DIA-IL, digital image analysis of individual leaves, DIA-WP, digital image analysis of whole plant canopy; PDA, potato dextrose agar; PI, plant introduction; V, visual disease evaluation.

Published in Crop Sci. 57:3303–3309 (2017).
doi: 10.2135/cropsci2016.08.0699

© Crop Science Society of America | 5585 Guilford Rd., Madison, WI 53711 USA All rights reserved.
Clear differentiation of disease severity between treatments through both accurate and precise disease evaluations is important in a number of different areas of turfgrass research. These include breeding studies that seek to identify disease resistant genotypes and applied research assessing the impact of inputs and management strategies on disease. A study on the assessment of visual evaluation techniques done by Horst et al. (1984) showed visual assessment to be inadequate at evaluating turfgrass quality and density. Ten evaluators took quality and density ratings of ten cultivars of Kentucky bluegrass (Poa pratensis L.) and tall fescue grown in Oregon in 1980 and Texas in 1981. Evaluator ratings and rankings differed significantly. These inconsistencies suggest that the criteria used for evaluation were inconsistent among evaluators. Studies performed by Nutter Jr. et al. (1993) on visual assessment of dollar spot (Sclerotinia homoeocarpa F.T. Benn) severity in creeping bentgrass (Agrostis stolonifera L.) showed similar results. Significant variation was reported for both intrarater repeatability and interrater reliability.

Digital image analysis (DIA) may reduce both inter- and intrarater variability through increased objectivity and standardization of practices. The accuracy and/or precision of DIA has been shown to be highly effective for a number of different crops with varied leaf shapes and types of pathogen infections (Barbedo, 2014; Bock et al., 2008; Lindow and Webb, 1983; Martin and Rybicki, 1998; Pourreza et al., 2015). However, poor correlations have also been found in DIA estimations of powdery mildew (Erysiphe cichoracearum) D.C. Ex Merat) on squash (Cucurbita maxima) Dutch) (Moya et al., 2005) and powdery mildew (Podosphaera clandestine Waller. Fr.) on sweet cherry [Prunus avium (L.) L.] (Olmstead and Lang, 2001). These mixed results indicate potential variation in DIA estimates due to differences in quantification methodology, image capture methods, and/or the type of disease signs and symptoms evaluated.

Although many studies have looked at disease evaluation using DIA, very few have examined the accuracy and precision of this method in evaluating leaves that are not perpendicular to the image capture device as is the case in a turfgrass canopy. In turfgrass research, the accuracy of DIA has been determined for evaluating turfgrass color and percent cover (Karcher and Richardson, 2003; Richardson et al., 2001). Digital image analysis was also used to successfully measure disease severity of the turfgrass disease dollar spot (Horvath and Vargas, 2005; Steketee et al., 2016). No reports are yet available comparing the precision and accuracy of DIA estimates to visual estimates of turfgrass disease severity using a quantified method as a standard of accuracy. Using visual evaluations as a standard of accuracy is common; however, this may have limited efficacy due to the subjectivity of visual evaluations (Zink and Kartunova, 1998).

The objective of this study was to compare DIA and visual evaluations for accuracy and precision in estimating brown patch disease severity on tall fescue.

**MATERIALS AND METHODS**

### Plant Maintenance

Material for Exp. 1 consisted of six seeds of ‘Kentucky 31’ tall fescue planted in 3.8-cm diam. cone-tainers (SC7 stubby cells, Steuwe & Sons, OR) filled with Pro-Mix potting media containing biofungicide (Bacillus subtilis MBI 600) (Code 0532, Premier Horticulture, Quebec, Canada). Plants were maintained in a glasshouse, watered at a rate of 2.55 cm h\(^{-1}\) for 3 min daily using an overhead irrigation system, and fertilized with three applications per month of a 20:20:20 mix at an N rate of 48.8 kg ha\(^{-1}\). Plants were cut twice a week to a height of 7.6 cm.

Material for Exp. 2 consisted of 20 seeds each from 13 plant introductions (PIs) obtained from the USDA germplasm database. Seeds were planted singly in each cone-tainer and maintained in the same manner described for Exp. 1.

### Inoculation

Fourteen isolates of *R. solani* were collected from creeping bentgrass putting greens at the Virginia Tech Turfgrass Research Center in Blacksburg, Virginia. These were assessed for virulence on tall fescue. Three of the most virulent isolates (data not shown) were used to create inoculum. Inoculum was created using filter paper cut to 2 cm by 0.5 cm, autoclaved for 1 h prior to use, and placed radially around a 4-mm diam. plug of *R. solani* on potato dextrose agar (PDA) (Fig. 1). Plates were maintained for 2 wk to allow sufficient colonization of the filter paper. Plants were then inoculated by placing three infected filter paper pieces, one piece from each isolate, within the plant canopy of each cone-tainer. Plants were inoculated approximately 6 wk after planting.

### Disease Treatment

Disease treatment was set up similarly to other breeding studies screening turfgrass genotypes for disease resistance under controlled conditions (Beirn et al., 2015; Curley et al., 2005; Elliot, 1995; Fu et al., 2005). Plants were placed in a sealed chamber to increase air temperature and provide high humidity for disease development. Plants were cut to a height of 7.6 cm prior to inoculation and again prior to disease evaluation. Two methods were used to obtain plants exhibiting a range of disease severity. In Exp. 1, plants were subject to varying days under disease pressure. Plants were placed in a randomized complete block design and six plants were removed and evaluated at 0, 2, 4, 6, 8, and 10 d after inoculation, respectively. After evaluation, plants were not replaced in the disease
again after the 24-hr drying period. Leaves exhibiting natural senescence due to age were identified and removed by hand prior to imaging. Leaves exhibiting senescence due to severe disease pressure, where the dark edges of coalescing lesions could be seen on the leaf, were not removed. When capturing images, cones were lined up with an index mark placed on the cone-tainer stand and cone prior to inoculation to provide consistent leaf orientation between the initial and final pictures. The stand was set at an angle of approximately 45˚ for the initial and final pictures. The stand was set at an angle of approximately 45˚ and a Canon PowerShot G6 digital camera (Canon, Inc., Tokyo, Japan) was placed on a tripod with the orientation adjusted to take a top down image (Fig. 2). The aperture on the camera was set to f/8 and the shutter speed was manually adjusted for each image until the light meter on the camera read zero. Two Calumet (Calumet Photographic, Inc., Bensenville, IL) light stands, each containing four 40-W compact semispiral, daylight balanced (5500 K) fluorescent bulbs (model 0L2000, Calumet Photographic, Inc. Bensenville, IL) were placed on either side of the stand to provide even illumination. Three images were obtained for each cone with each cone placed in the stand, photographed, and removed from the stand between each image.

Pictures were analyzed using APS Assess image analysis software (APS Press, St. Paul, MN). Hue thresholds of 31 (low) and 191 (high) were used to distinguish the plant from the background. The background was then substituted for a solid blue background. An intensity threshold of 66 (low) and 255 (high) was used to select the soil, which was then replaced with a solid blue background. Hue thresholds of 31 (low) and 191 (high) were used to determine the total plant area. Lesion area was determined with hue thresholds of 31 (low) and 101 (high). A ratio of lesion area to total plant area was used by APS Assess to determine percent disease severity. Total percent disease severity was calculated as the difference in disease severity between the pre-inoculation picture and post-inoculation picture.

3. Individual leaf DIA (DIA-IL): All leaves from each plant were cut at the soil surface, placed within a clear plastic cover sheet, and scanned at 300 dpi to produce JPEG images using a CanoScan LiDE 50 scanner (Canon, Inc., Tokyo, Japan). Each set of leaves was scanned three times. The cover sheet was removed and the leaves within it were rearranged between each scan. APS Assess image analysis software was used to evaluate percent disease in each image. Hue thresholds of 31 (low) and 191 (high) were used to distinguish the plant from the background. The background was then substituted for a solid blue background. An intensity threshold of 66 (low) and 255 (high) was used to select the soil, which was then replaced with a solid blue background. Hue thresholds of 31 (low) and 191 (high) were used to determine the total plant area. Lesion area was determined with hue thresholds of 31 (low) and 101 (high). A ratio of lesion area to total plant area was used by APS Assess to determine percent disease severity. Total percent disease severity was calculated as the difference in disease severity between the pre-inoculation picture and post-inoculation picture.

Disease Evaluation

Upon removal from the chamber, plants were subirrigated on greenhouse benches for a 24-hr period to allow the leaf canopy to dry prior to disease evaluation. Percent disease was estimated for each plant using visual evaluation (V) and DIA of the whole plant canopy (DIA-WP). A third method, DIA of individual leaves (DIA-IL), was used as a standard of accuracy by which to judge the accuracy of DIA-WP and visual estimations. The DIA-IL method was shown to be highly predictive of actual brown patch severity in previous studies ($r^2 = 0.99$) (Sykes, 2009). Similar individual leaf methods using APS assess software have also been quantified as highly accurate ($r = 0.97$) in other plant–disease combinations (Bock et al., 2008). The V, DIA-WP, and DIA-IL methods are described below.

1. Visual analysis: Three evaluators performed visual ratings of disease severity using a percentage scale. All three evaluators were trained in the evaluation of brown patch on tall fescue, though years of experience varied. Evaluators had approximately 1, 3, and 14 yr of experience.

2. DIA of whole plant canopy (DIA-WP): Pictures were taken of each plant prior to inoculation and...
solid blue background. Hue thresholds of 31 (low) and 191 (high) were used to determine the total plant area and hue thresholds of 31 (low) and 87 (high) were used to determine the lesion area. The ratio of lesion area to total plant area was used by APS Assess to determine percent disease severity.

In Exp. 2, only one picture, scan, or visual evaluation was performed for each plant. Visual estimation of plant disease was performed by a rater with 3 yr of experience. All other methods followed those described for Exp. 1.

**Statistical Analysis**

The absolute error was calculated for each disease severity estimate as the difference between the mean disease severity estimated using the DIA-IL method (standard of accuracy) and the disease severity estimated using either DIA-WP or visual estimation. Standard deviations were calculated within the three images, scans, or visual estimations performed on each plant. Standard deviation was only calculated for Exp. 1.1 and 1.2.

To compare the accuracy (absolute error) and precision (standard deviation) of each method, analyses of variance were performed in SAS v. 9.3 (SAS Institute, Inc., Cary, NC) using the GLIMMIX procedure with the following mixed model in which either absolute error or standard deviation were considered the dependent variable; estimation method and disease range were considered fixed effects and experiment was considered a random effect.

\[
Y_{ijkl} = \mu + M_i + D_j + MD_{ij} + E_k + MDE_{ijk} + \epsilon_{l(ijk)}
\]

where \(Y_{ijkl}\) is the observed value of absolute error or standard deviation for the \(l\)th replication within the \(i\)th method by the \(j\)th disease range by the \(k\)th experiment; \(\mu\) is the population mean; \(M_i\) is the effect of the \(i\)th method (\(i = \text{DIA-WP or Visual}\)); \(D_j\) is the effect of the \(j\)th disease range (\(j = 0\) to 20, 20.1 to 40, 40.1 to 60, 60.1 to 80, or 80.1 to 100); \(MD_{ij}\) is the interaction effect of the \(i\)th method with the \(j\)th disease range; \(E_k\) is the effect of
the $k$th experiment ($k = 1.1, 1.2, 2.1, \text{or} 2.2$); $MDE_{ijk}$ is interaction effect of the $i$th method with the $j$th disease range and $k$th experiment; $e_{ijk}$ is experimental error or residual.

Means were separated using Tukey’s HSD. Statistical significance was determined using a $P$-value of less than 0.05 for all tests.

RESULTS AND DISCUSSION

The mean and distribution of disease severity differed among experiments (Fig. 3). Blocking by experiment helped control this variation and allowed for comparison of methods across the full range of potential disease values (0% to 100%). Across the four experiments, the total number of observations per disease group were $n = 226$ (0 to 20% disease), $n = 126$ (20.1 to 40% disease), $n = 66$ (40.1 to 60% disease), $n = 40$ (60.1 to 80% disease), and $n = 46$ (80.1 to 100% disease). There was no interaction between experiment and method for either absolute error or standard deviation analyses.

Mean absolute error values between DIA-WP and visual estimations did not differ and the interaction effect between evaluation method and disease range was not significant. Because methods did not differ, data were combined across methods to assess differences in mean absolute error by disease range. Mean absolute error did differ by disease range, exhibiting a bell-shaped curve across the range of disease severity (Fig. 4). The highest mean absolute error occurred at the 40.1 to 60% range while the lowest mean absolute errors occurred at the 0 to 20% and 80.1 to 100% ranges.

In the model examining precision, standard deviation differed by method and disease range and a significant interaction between method and disease range was observed. This indicates the differences in standard deviation by disease range differed by method. Therefore, each method was considered separately. Within the DIA-WP method, standard deviation did not differ by disease range (Fig. 5). However, within the visual estimation method, the standard deviation did differ by disease range, with a significant increase in standard deviation being observed.
between the 0% to 20%, 20.1% to 40%, and 40.1% to 60% ranges. Standard deviation for DIA-WP was significantly lower than that of visual estimations within the 40.1% to 60% disease range. Within the remaining disease ranges, these two methods did not differ.

Although DIA-WP was more consistently precise across the range of disease severity, it did not show improved accuracy compared with visual evaluations. Potential improvements in image capture were identified which might provide improved accuracy in future DIA evaluations. In all DIA-WP evaluations, a purple cloth was placed around the cone to mask the stand and provide a uniform background for the image. Although the same cloth was used for each image capture, the color appears different in certain images. Light stands were used to provide even illumination; however, this did not entirely negate the influence of light from outside sources. Better control of outside light sources could help to further reduce variability and potentially improve accuracy.

Hue variance may also have been caused by the camera settings used when capturing images. The aperture was set to f/8 and the shutter speed adjusted to balance the light meter to zero for each image. This resulted in slight overexposure of images with greater disease severity. This minor overexposure did not affect the ability of DIA to detect disease when parameters were defined on a single image basis. However, when evaluating disease using a DIA macro, where the parameters are set prior to batch evaluation of images, this could impact the ability of DIA to consistently define diseased areas in each image due to differences in exposure. Further reducing variability by using a closed light source and balancing the camera settings to a neutral image prior to image capture could aid in increasing the accuracy of DIA methods.

The DIA method exhibited improved precision, and potential improvements to accuracy were identified. However, other factors may compromise overall efficacy in adoption of this technology. Our disease inoculations were performed in a glasshouse setting with highly controlled conditions that limited the introduction of abiotic or biotic stresses other than the pathogen of interest. Images were also captured in a highly controlled setting with uniform plant placement and lighting. This type of methodology would work very well for disease resistance screenings in the initial stages of cultivar development, where hundreds of plant introductions could be rapidly assessed for resistance to specific pathogens. This type of controlled environment is common in the initial stages of breeding studies, as it allows for screening of larger numbers of genotypes and reduces the potential for escapes being identified as resistant. In later stages of cultivar development under field conditions, these methods may not be as effective. Modification of image capture methods by using a portable light box or drone technology to assess field plots could allow these analysis methods to be used for field scale disease evaluation. Continued research would need to assess whether the image analysis methods described would translate to larger areas of turfgrass cover under less controlled conditions where additional abiotic and biotic stressors could potentially influence efficacy.

Additional concerns revolve around the time and expense of this technology. One of the distinct advantages of visual evaluation is that it requires no expensive equipment and analysis is very rapid. Although the cost of equipment required for digital image analysis has dropped significantly in the past decade, the time required to capture and process an image still exceeds the time required to visually assess a plant for a single trait. With increased levels of automation and the use of macros to assess multiple parameters such as color, percent cover, and disease simultaneously, the time requirement of digital image analysis might become more comparable to visual evaluation.

**CONCLUSION**

The DIA-WP method evaluated in this study was significantly more precise than visual evaluation within the 40.1% to 60% disease range. Additionally, while no differences in precision were observed among the disease ranges within the DIA-WP method, precision differed between the 0 to 20%, 20.1 to 40%, and 40.1 to 60% range within visual evaluations. The DIA-WP method demonstrated improved precision; however, it did not show improved accuracy compared with visual evaluation. An increase in absolute error under either evaluation method as disease values approached the 40.1 to 60% disease range indicates a strong need for continued efforts to improve the accuracy of disease evaluation methods. Reduction in image variability introduced by differences in light quality and camera settings may result in further improvements to both the accuracy and precision of DIA in evaluating brown patch disease severity on tall fescue. These results are important for turfgrass evaluation since leaf orientation differs from many of the previous research studies examining the use of DIA to evaluate plant disease. This technology offers improved phenotyping capability that is critical to the early stages of cultivar improvement, when large numbers of individual genotypes are screened for disease resistance under controlled conditions. Continued research focused on improving accuracy and translating this methodology to field-based turfgrass evaluations would further expand the utility of this technology in turfgrass research.

**References**

Barbedo, J.G. 2014. An automatic method to detect and measure leaf disease symptoms using digital image processing. Plant Dis. 98:1709–1716. doi:10.1094/PDIS-03-14-0290-R-E
Beirn, L.A., W.A. Meyer, B.B. Clarke, and J.A. Crouch. 2015. A greenhouse-based inoculation protocol for fungi causing crown rust and stem rust diseases of Kentucky bluegrass turf. HortScience 50:1509–1513.

Bock, C.H., P.E. Parker, A.Z. Cook, and T.R. Gottwald. 2008. Visual rating and the use of image analysis for assessing different symptoms of citrus canker on grapefruit leaves. Plant Dis. 92:530–541. doi:10.1094/PDIS-92-4-0530

Curley, J., S.C. Sim, S. Warnke, S. Leong, R. Barker, and G. Jung. 2005. QTL mapping of resistance to gray leaf spot in ryegrass. Theor. Appl. Genet. 111:1107–1117. doi:10.1007/s00122-005-0036-x

Elliott, M.L. 1995. Disease response of Bermudagrasses to Gaeumannomyces graminis var. graminis. Plant Dis. 79:699–702. doi:10.1094/PD-79-0699

Fu, D., N.A. Tisserate, Y. Xiao, D. Settle, S. Muthukrishnan, and G.H. Liang. 2005. Overexpression of rice TLPD34 enhances dollar-spot resistance in transgenic bentgrass. Plant Sci. 168:671–680. doi:10.1016/j.plantsci.2004.09.032

Horst, G.L., M.C. Engelkle, and W. Meyers. 1984. Assessment of visual evaluation techniques. Agron. J. 76:619–622. doi:10.2134/agronj1984.00021962007600040027x

Horvath, B.J., and J.M. Vargas. 2005. Analysis of dollar spot disease severity using digital image analysis. Int. Turfgrass Soc. Res. J. 10:196–201.

Karcher, D.E., and M.D. Richardson. 2003. Quantifying turfgrass color using digital image analysis. Crop Sci. 43:943–951. doi:10.2135/cropsci2003.9430

Lindow, S.E., and R.R. Webb. 1983. Quantification of foliar plant disease symptoms by microcomputer-digitized video image analysis. Phytopathology 73:520–524. doi:10.1094/Phyto-73-520

Martin, D.P., and E.P. Rybicki. 1998. Microcomputer-based quantification of maize streak virus symptoms in Zea mays. Phytopathology 88:422–427. doi:10.1094/PHYTO.1998.88.5.422

Moya, E.A., I.R. Barrales, and G.E. Apablaza. 2005. Assessment of the disease severity of squash powdery mildew through visual analysis, digital image analysis and validation of these methodologies. Crop Prot. 24:785–789. doi:10.1016/j.cropro.2005.01.003

Nutter, F.W., Jr., M.L. Gleason, J.H. Jenco, and N.C. Christians. 1993. Assessing the accuracy, intra-rater repeatability, and inter-rater reliability of disease assessment systems. Phytopathology 83:806–812. doi:10.1094/Phyto-83-806

Olmstead, J.W., and G.A. Lang. 2001. Assessment of severity of powdery mildew infection of sweet cherry leaves by digital image analysis. HortScience 36(1):107–111.

Piper, C.V., and H.S. Coe. 1919. Rhizoctonia in lawns and pastures. Phytopathology 9(2):89–92.

Pourreza, A., W.S. Lee, E. Etxeberria, and A. Banarjee. 2015. An evaluation of a vision-based sensor performance in Huanglongbing disease identification. Biosystems Eng. 130:13–22. doi:10.1016/j.biosystemseng.2014.11.013

Richardson, M.D., D.E. Karcher, and L.C. Purcell. 2001. Quantifying turfgrass cover using digital image analysis. Crop Sci. 41:1884–1888. doi:10.2135/cropsci2001.1884

Steketee, C.J., A.D. Martinez-Espinoza, K.R. Harris-Shultz, G.M. Henry, and P.L. Raymer. 2016. Effects of genotype and isolate on expression of dollar spot in seashore paspalum. HortScience 51(1):67–73.

Sykes, V.R. 2009. Use of digital image analysis to identify Rhizoctonia solani and Rhizoctonia zeae resistance in Festuca arundinacea plant introductions. Virginia Tech.

Zar, J.H. 1999. Biostatistical analysis. 4th ed. Prentice Hall, Inc., Upper Saddle River, NJ.

Zink, A.G., and E. Kartunova. 1998. Wood failure in plywood shear samples measured with image analysis. For. Prod. J. 48:69–74.