Eur J Plant Pathol (2015) 142:603–614
DOI 10.1007/s10658-015-0638-3

A set of standard area diagrams to assess severity of frogeye leaf spot on soybean

Daniel Debona · Kelly Juliane Telles Nascimento · Danielle Rezende · Jonas Alberto Rios · Arthur Martins Almeida Bernardeli · Leandro Castro Silva · Fabrício Ávila Rodrigues

Accepted: 6 March 2015 / Published online: 14 March 2015

Abstract A set of standard area diagrams (SADs) was developed and validated to aid visual assessment of severity of frogeye leaf spot (FLS) caused by *Cercospora sojina*. The SAD has eight color images of diseased leaflets with severity values that ranged from 0.1 to 39.9 %. The SAD was validated by a group of 20 raters [10 experienced (ER) and 10 inexperienced (IR)], who assessed the same set of 50 images twice, the first without SADs and the second using SADs as an aid. The SADs significantly improved accuracy [coefficients of bias (C_b) were 0.64 and 0.99 for IR and 0.98 and 0.99 for ER, without and with SADs, respectively], precision [correlation coefficients (r) were 0.89 and 0.95 for IR and 0.94 and 0.97 for ER, without and with SADs, respectively] and overall agreement [Lin's concordance correlation coefficients (ρ_c) were 0.57 and 0.94 for IR and 0.92 and 0.97 for ER without and with SADs, respectively]. The estimates of severity of FLS were more reliable when using SADs. Both the inter-rater reliability (coefficient of determination, R^2) and intra-class coefficient (ρ) were significantly increased by using SADs. Therefore, it is believed that the SADs proposed in the present study will be a useful tool to aid accurate, precise and reliable estimates of severity of FLS in experiments (e.g., fungicide screening, assessment of partial resistance of soybean genotypes to FLS) and to aid in decision-making purposes.

Keywords *Cercospora sojina* · *Glycine max* · Disease assessment · Disease severity · Epidemiology · Phytopathometry

D. Debona · K. J. T. Nascimento · D. Rezende · J. A. Rios · A. M. A. Bernardeli · L. C. Silva · F. Á. Rodrigues (✉)
Departamento de Fitopatologia, Laboratório da Interação Planta-Patógeno, Universidade Federal de Viçosa, Viçosa, Minas Gerais State 36570-900, Brazil
e-mail: fabricio@ufv.br

Introduction

Frogeye leaf spot (FLS), caused by the fungus *Cercospora sojina* Hara, was first reported on soybean in 1915 in Japan (Mian et al. 2009). Subsequently, it has become an important disease in warm, humid tropical and subtropical regions (Akem and Dashiell 1994). Although *C. sojina* is able to infect all above ground organs, FLS is primarily a leaf disease, with symptoms initially forming as dark, water-soaked spots (with or without a lighter center) and developing into brown spots surrounded by a narrow, purple-brown margin (Nascimento et al. 2014). The lesions can coalesce and cause severe leaf blight or defoliation (Kim et al. 2013). FLS epidemics can cause yield losses greater than 60 % (Dashiell and Akem 1994). *Cercospora sojina* survives in infested soybean residues and infected seeds (Mian et al. 2009). Because FLS is a polycyclic disease, inoculum can increase continuously throughout the growing season, thereby causing severe epidemics when control measures are not properly adopted (Kim et al. 2013). Sowing disease-free seeds, seeds treated with fungicides, crop rotation and fungicide sprays applied at the R2-R5 growth stages are used to control FLS (Mian et al. 2009). However, resistance based on three single genes (*Rcs1-3*) is the most important strategy that has been adopted for disease control (Mian et al. 2009). In spite of the past effectiveness of fungicides in managing FLS, *C. sojina* isolates with reduced sensitivity to quinone outside inhibitor fungicides were recently detected in the U.S. (Zhang et al. 2012). The recent characterization of sexual reproduction in *C. sojina* (Kim et al. 2013) causes additional concerns regarding FLS management because recombination amplifies the genetic variability of the fungus, which may render resistance ineffective. The extent of sexual reproduction in *C. sojina* has not yet been investigated in Brazil, but the fungus is known to be genetically diverse (Yorinori and Klingelfuss 1999). Therefore, studies to better understand the biology of *C. sojina*, to identify soybean genotypes with partial resistance to FLS and to screen fungicides are urgently needed. However, a standard method to quantify FLS has not been developed, hindering efforts to address these important issues.

The assessment of disease severity in experiments to compare control measures and for characterization of pathogen aggressiveness is dependent on the availability of a standard method for disease quantification (Duarte et al. 2013). Poor disease quantification can result in erroneous conclusions and recommendations that might have catastrophic consequences for producers (Madden et al. 2007). The phytopathometric variable most commonly used for quantifying FLS on soybean is severity, using the percentage of symptomatic leaf area. However, estimation of disease severity by raters is difficult because it is subjective and estimates may vary according to the inherent ability of the rater (Campbell and Madden 1990). This subjectivity introduces undesirable error; thus the use of standardized criteria to measure disease is imperative (Lenz et al. 2010).

Ideally, a disease assessment scheme should be accurate, precise and reliable (Nutter and Schultz 1995; Madden et al. 2007). Accuracy measures the degree of closeness or bias between the estimates and the actual severity and involves the difference between the means of observed and expected values, and scale differences (Nutter and Schultz 1995; Madden et al. 2007), precision is a measure of the variability of the estimates (Madden et al. 2007; Yadav et al. 2013), and the concept of agreement combines both accuracy (measure of bias) and precision in the case when actual true values are being considered (Madden et al. 2007). Reliability is related to the ability of the rater to produce similar estimates from the same specimen under different conditions (Nutter et al. 1991).

A standardized method that has proven very useful for estimating disease severity is the standard area diagram (SAD), which consists of pictorial diagrams that depict the true proportion of severity on individual sampling units (quadrants, whole plants, leaves, fruit, tubers, etc.) and represents a range of disease severities (Bergamin Filho and Amorim 1996; González-Domínguez et al. 2014). Although it has long been recognized that the use of SAD provides more accurate and precise estimates of diseases severity (James 1971; Horsfall and Cowling 1978; Kranz 1988), it was only recently that improvements afforded by using SADs relative to estimates made without SADs were statistically supported (Corrêa et al. 2009; Duarte et al. 2013; Rios et al. 2013; Yadav et al. 2013; González-Domínguez et al. 2014). Furthermore, it has been demonstrated that inexperienced raters tend to improve more in accuracy and reliability when using SADs compared to experienced raters (Michereff et al. 2000; Nita et al. 2003; Godoy et al. 2006; Bock et al. 2009, 2013; Pedroso et al. 2011; Sachs et al. 2011; Klosowski et al. 2013; Yadav et al. 2013; González-Domínguez et al. 2014).

Severity of FLS has been assessed using various methods including an ordinal rating scale (Ross 1968), and disease category scales (Akem and Dashiell 1994), such as the Horsfall-Barrat (H-B) scale, which is a logarithmically-based scale with unequal intervals of severity related to the eye's supposed ability to discern differences (the so-called Weber-Fechner law) (Horsfall and Barrat 1945). Problems can arise when assessing severity of FLS with these methods: the first method is less informative and tends to be subjective, whereas the assumptions behind the H-B scale were never supported experimentally and have been severely criticized (Nutter and Esker 2006; Bock et al. 2010). No SADs have been developed and validated to assess severity of FLS. Two objectives were addressed in this study: i) development of SADs as an assessment aid for estimating severity of FLS, and ii) evaluation of the effects of SADs and rater experience on accuracy, precision and reliability of the estimates of severity of FLS.

Materials and methods

Diseased soybean leaves

Leaves of the soybean cultivars Bossier and Conquista, susceptible and resistant to FLS, respectively

(Nascimento et al. 2014), were obtained from plants previously inoculated with a pathogenic isolate of *C. sojina* (UFV/DFP-*Cs* 01). The inoculation procedure followed that described by Nascimento et al. (2014). Plants were maintained in a greenhouse at a temperature of 30±5 °C, relative humidity of 65±5 % and natural light radiation of 900±15 μmol photons m^{-2} s^{-1}, which was quantified with a LI-250A light meter (LI-COR, Inc., Lincoln, NE, USA).

Development and evaluation of the SAD

A total of 150 leaflets with a range of severities of FLS were collected at 16 days after inoculation and were individually scanned to obtain images with a resolution of 300 dpi. These images were processed, and the actual severity of FLS was measured using QUANT software (Vale et al. 2003). The minimum and maximum true severity of FLS obtained after image processing was 0.1 and 39.9 %, and following an approximately linear increase, six additional, intermediate diagrams were chosen resulting in a set of eight images. These soybean leaflet images with known values of severity of FLS were used as templates for the diagrams developed for the SADs. The typical symptoms of necrosis and purple-brown borders were included in the leaflet diagrams in the SADs.

A group of 20 raters (10 inexperienced and 10 experienced) assessed the 50 images. Raters were classified as experienced if they had received previous formal training and practice in assessment of severity of FLS and were familiar with FLS symptoms. Rater training was performed in three steps 2 weeks before assessments. First, a group of 50 digital images of FLS-diseased leaflets that was not used for SADs validation was projected to raters and the raters were provided with an explanation as to how to recognize FLS symptoms (to consider only lesions with light brown to light gray centers and showing a darker red to purple-brown border). Second, the same group of images was projected and the actual severity of FLS was immediately shown with the image. Raters were cautioned to take care with estimates of severity in images which had numerous small lesions (due to risks of overestimation, Madden et al. 2007) and they were informed that leaflets become chlorotic at severities near 40 %. Third, the raters estimated severity of FLS for each training image, immediately after which the actual severity was shown on the image. Images were projected one at a time at random for 15 s each. Images used in rater training and SADs validation were chosen to embrace a wide range of FLS severity and included leaflets with several small lesions. Inexperienced raters had no formal training or familiarity with FLS symptoms. For each group, the 50 digital images of the FLS-diseased leaflets were projected at random one at a time for 15 s. In the first assessment, the raters did not use SADs as an aid to estimate severity of FLS. Two weeks later, the same raters made a second assessment using SADs as an aid to estimate severity of FLS using the same set of images. To evaluate each leaflet, the rater compared the image with those in the SADs to obtain a new estimate of severity of FLS. Before performing both evaluations, raters were told to consider only the symptoms of necrosis and purple-brown borders typical of FLS lesions and to base their visual estimate on proportion of diseased area related to total leaflet area for each image as a percentage.

Data analyses

Agreement between estimated and actual severity of FLS for each rater was determined according to Lin's concordance correlation coefficient (LCCC) (ρ_c) (Lin 1989). The analysis was performed separately using the data obtained with and without the use of the SADs for both inexperienced and experienced raters. This analysis is the most appropriate for gauging agreement because it combines measures of accuracy and precision to assess the fit of pairs of observations to the line of concordance (45°). The LCCC is comprised of two statistics where $\rho_c = C_b \, . \, r$, and where C_b is a measure of bias or accuracy of the best-fit, and r equals the correlation coefficient between the estimated severity (Y) and the actual severity (X), and in this case, measures the precision of the best-fit line. C_b is derived from the following: $C_b = 2/[(\upsilon + 1/\upsilon + \mu^2)]$, where $\upsilon = \sigma_y / \sigma_x$, and σ_y and σ_x are the standard deviations of Y and X, respectively, and where $\mu = (\mu_y - \mu_x) / \sqrt{(\sigma_y \, . \, \sigma_x)}$ and μ_y and μ_x are the mean values of Y and X, respectively. The term υ (scale shift or systematic bias) describes the dependency of over or underestimation on the magnitude of X (it can be described as the difference in the slopes of the fitted line and the line of concordance). Slopes of 45° would have a υ of 1. The term μ (location shift or constant bias) is the tendency, on average, to over or underestimate severity. Equal heights would have a μ of 0. A perfectly accurate measurement occurs when estimates equal actual values, and all points fall on the concordance line e.g., r=1, C_b=1 υ=1, μ=0 and thus, ρ_c=1 (Bock et al. 2010; Nita et al. 2003).

For all parameters analyzed (υ, μ, C_b, ρ_c and r), the differences between the means (i.e., with SADs minus without SADs) were calculated and an equivalence test was used to test their significance (Yi et al. 2008; Bardsley and Ngugi 2013; Yadav et al. 2013). The equivalence test was used to calculate 95 % confidence intervals (CIs) for each statistic (the difference between the means) by bootstrapping using the percentile method (with an equivalence test, the null hypothesis is the converse of H_0, i.e., the null hypothesis is non-equivalence). All analyses were based on 2,000 balanced bootstrap samples using 95 % CIs that were calculated as the difference between the means of the groups. If the CIs spanned zero, there was no significant difference (P= 0.05). Precision was also determined using analysis of the absolute error (estimated minus actual severity of FLS).

The inter-rater reliability of the estimates was determined in two ways. First, the coefficient of determination (R^2) from linear regression analyses of relationships between severity estimates for all pairs of raters was used (Nutter and Schultz 1995). Second, the inter-rater reliability was assessed based on the intra-class correlation coefficient (ICC, ρ) (Shrout and Fleiss 1979), which, unlike most other correlation measures, operates on data structured as groups rather than data structured as paired observations, which provide only relative measures. For the ICC analysis, the model was assumed to be two-way, with absolute agreement and single measures (Shrout and Fleiss 1979). The effect of SADs use on the inter-rater reliability for each assessment time was measured based on the confidence interval of the ICC estimated by the model.

All statistical analyses were calculated in R (R Core Team 2013). The epi.ccc function of the epiR package (Stevenson 2012) was used to obtain Lin's CC statistics. The built-in boot.sample R function was used for the equivalence test. The ICC was calculated with the icc function of the irr R package (Gamer et al. 2012).

Results

Soybean leaflets developed FLS symptoms that began as dark, water-soaked circular to somewhat irregular spots with light centers (Fig. 1). As the lesions expanded, the centers turned light brown to light gray in color and developed a darker red to purple-brown border as they aged. When severe (~40 %), the leaflets showed intense chlorosis. The SADs developed in the present study comprised eight color frogeye spot-diseased leaflet images, each of a known severity value ranging from 0.1 to 39.9 % (Fig. 2).

Lin's concordance analysis revealed that the raters' estimates of severity of FLS approached the actual severity when they used SADs, irrespective of their experience. There was a linear relationship between estimated and actual severity of FLS for all raters. Use of SADs resulted in significant improvement for all bias and accuracy statistics of Lin's concordance correlation coefficient (υ, u, C_b, ρ_c and r) for inexperienced raters and for C_b, ρ_c and r for experienced raters (Table 1).

Use of SADs significantly reduced scale shift (υ) for inexperienced raters (Table 1), which ranged from 0.78 to 2.73 (mean of 1.93) and from 0.80 to 1.05 (mean of 0.93) without and with SADs, respectively. The proportion of raters who had υ values between 0.90 and 1.10 increased from 10 to 70 % with use of SADs, indicating that the slope of the best fitting line was closer to the concordance line when SADs were used (Fig. 3a).

Location shift (μ) for inexperienced raters was significantly reduced by using SADs (Table 1). Inexperienced raters overestimated severity of FLS without SADs [μ ranged from 0.02 to 1.47 (mean of 0.85)]. In contrast, μ values ranged from −0.16 to 0.16 (mean of −0.01) when inexperienced raters used SADs.

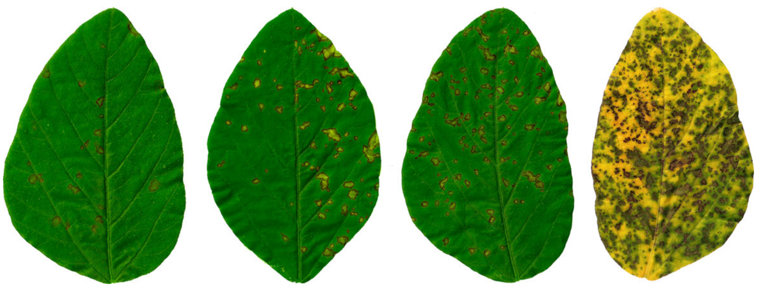

Fig. 1 Symptoms of frogeye leaf spot (FLS) on soybean leaflets inoculated with *Cercospora sojina*. Characteristic lesions have a necrotic center surrounded by a purple-brown margin. At high FLS severity (~40 %), the leaflets become chlorotic and senesce prematurely

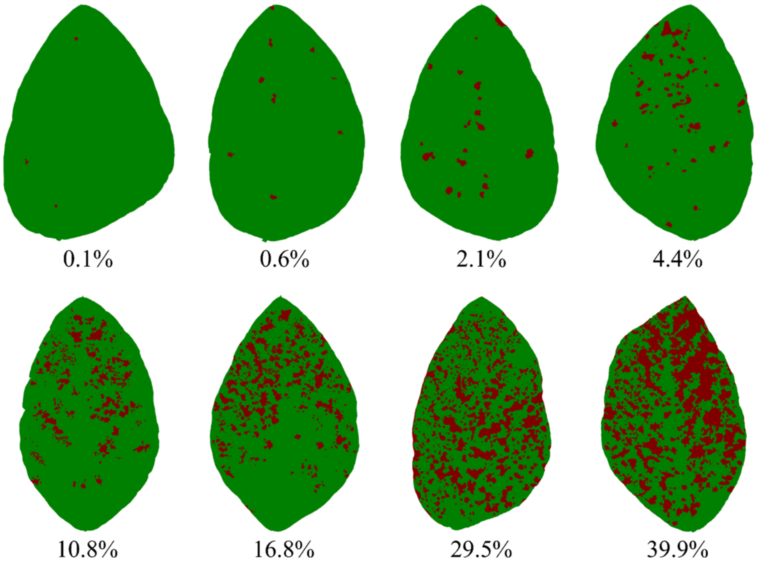

Fig. 2 Set of standard area diagrams (SADs) developed as an aid for assessment of frogeye leaf spot (FLS) on soybean leaflets. Values represent the percentage (%) of the leaflet area showing FLS symptoms

The proportion of inexperienced raters with μ values between −0.10 and 0.10 was 30 % without use of SADs, but 90 % when SADs were used, indicating a reduction in overestimation when SADs are used as an aid to assess severity of FLS (Fig. 3b).

Accuracy (C_b) was significantly increased when either experienced or inexperienced raters used SADs (Table 1). Accuracy without SADs ranged from 0.38 to 0.97 (mean of 0.64) for inexperienced raters and from 0.95 to 0.99 (mean of 0.98) for experienced

Table 1 Effect of the use of a set of standard area diagrams (SADs) as an assessment aid on bias, precision and agreement of estimates of severity of frogeye leaf spot (FLS) on soybean leaflets made by raters with or without experience in FLS assessment

Experience	LCC statistic	Means		95 % CI[a] of the difference between means
		No SADs	With SADs	
Inexperienced	Scale shift (υ)[b]	1.93	0.93	**−1.340, −0.629**
	Location shift (μ)[c]	0.85	−0.01	**−1.075, −0.634**
	Bias correction factor (C_b)[d]	0.64	0.99	**0.228, 0.465**
	Correlation coefficient (r)[e]	0.89	0.95	**0.028, 0.081**
	Concordance coefficient (ρ_c)[f]	0.57	0.94	**0.254, 0.474**
Experienced	Scale shift (υ)	0.93	0.92	−0.088, 0.053
	Location shift (μ)	0.05	−0.01	−0.154, 0.027
	Bias correction factor (C_b)	0.98	0.99	**0.005, 0.026**
	Correlation coefficient (r)	0.94	0.97	**0.017, 0.048**
	Concordance coefficient (ρ_c)	0.92	0.97	**0.032, 0.063**

[a] Bootstrap calculated difference between means and confidence intervals (CIs). If the CIs embrace zero, difference is not significant at the 5 % level. Bold numbers represent significance of the difference

[b] Scale or slope shift (systematic bias) relative to the perfect relationship (1=perfect relation between x and y)

[c] Location or height shift (constant bias) relative to the perfect relationship (0=perfect relation between x and y)

[d] Bias correction factor that measures how far the best-fit line deviates from a line at 45°. No deviation from the 45° line occurs when C_b=1. C_b is calculated from υ and μ and is a measure of accuracy

[e] Correlation coefficient (r) that measures precision

[f] Lin's concordance correlation coefficient (ρ_c) combines both precision (r) and accuracy (C_b) ($\rho_c = r{:}C_b$) to measure agreement with the actual value (Lin 1989)

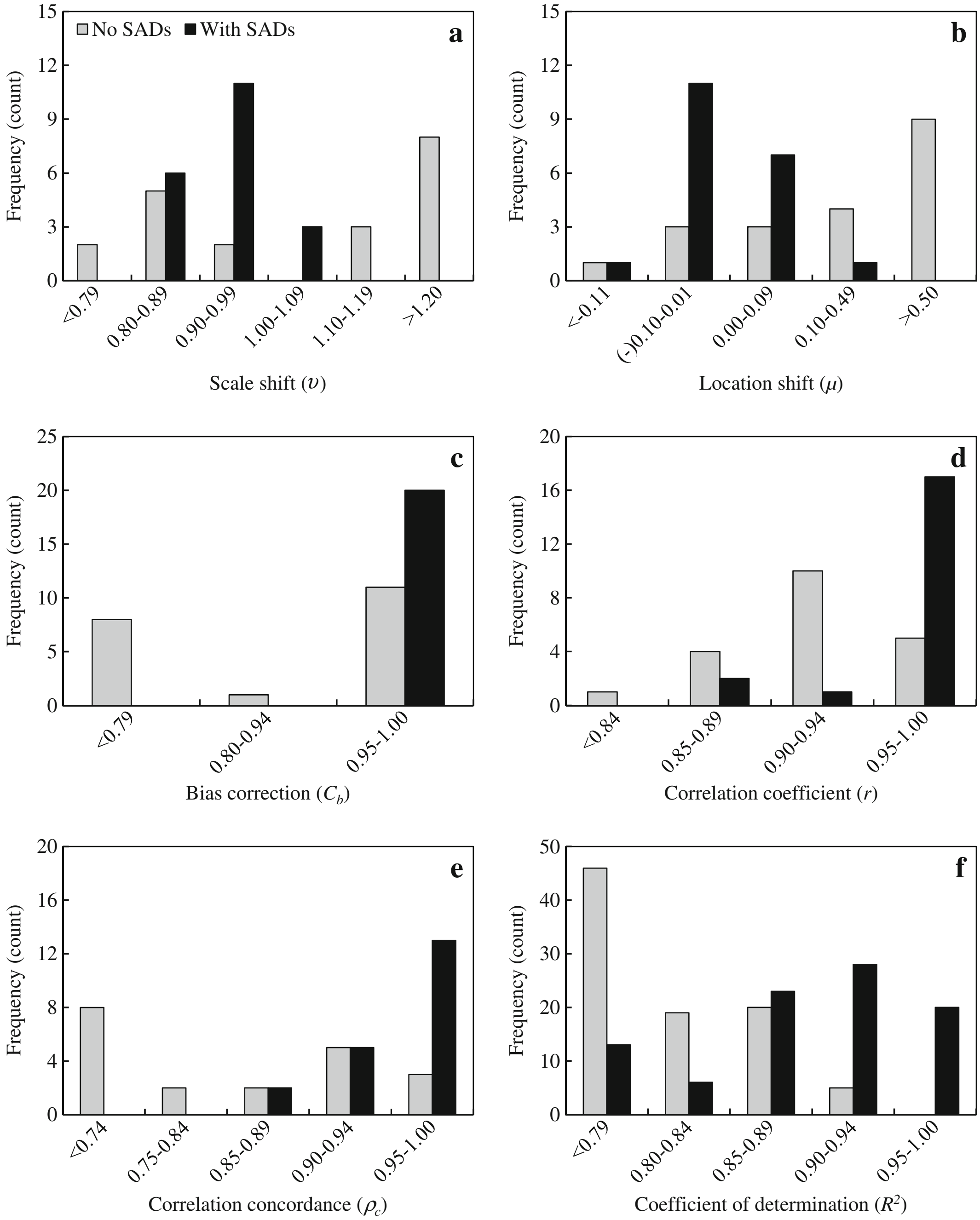

Fig. 3 Frequency of bias, precision and agreement and inter-rater reliability without and with use of a set of standard area diagrams (SADs) as an assessment aid by 20 raters who assessed 50 images of frogeye leaf spot-diseased soybean leaflets. (**a**) Scale shift (or systematic bias, υ), (**b**) location shift (or constant bias, μ), (**c**) bias correction factor (C_b), (**d**) correlation coefficient (r), (**e**) Lin's concordance correlation coefficient (ρ_c), and (**f**) coefficient of determination (R^2)

raters. In contrast, accuracy when using SADs for inexperienced raters ranged from 0.96 to 0.99 (mean of 0.99) and for experienced raters from 0.97 to 0.99 (mean of 0.99). Use of SADs also increased the proportion of raters with accuracy >0.95 from 55 to 100 % (Fig. 3c).

Without the SADs aid, precision (r) ranged from 0.83 to 0.93 (mean of 0.89) and from 0.90 to 0.97 (mean of

0.94) for inexperienced and experienced raters, respectively (Table 1). When assessments were made using SADs, precision improved for both inexperienced [0.89 to 0.98 (mean of 0.95)] and experienced raters [0.95 to 0.99 (mean of 0.97)]. The proportion of raters with precision (r) >0.95 was 25 and 85 % without and with use of the SADs aid, respectively, which indicates an increase in precision using the SADs (Fig. 3d).

Agreement (ρ_c), which combines measures of both accuracy and precision, was significantly improved for all raters, irrespective of experience, when SADs were used (Table 1). In the absence of SADs, ρ_c for inexperienced raters ranged from 0.32 to 0.81 (mean of 0.57) and for experienced raters from 0.88 to 0.97 (mean of 0.92). In contrast, when raters assessed using SADs inexperienced raters had a ρ_c ranging from 0.87 to 0.97 (mean of 0.94) and experienced raters from 0.94 to 0.99 (mean of 0.97). The improvement in agreement was reflected in the increase in the proportion of raters with a ρ_c value >0.90 (40 and 90 % without and with the SADs aid, respectively) (Fig. 3e).

The increase in precision was reflected in the reduction in absolute errors when raters used SADs, and was particularly evident for inexperienced raters (Fig. 4). When estimates were made without SADs, 80 % of the raters had error >25 %. When SADs were used as an aid to estimate severity, the errors were consistently <25 %.

Without SADs, experienced raters had less bias and better accuracy, precision and agreement (υ, u, C_b, r and ρ_c, respectively) compared to inexperienced raters (Table 2). However, when the experienced raters used SADs, only precision (r) and agreement (ρ_c) were significantly improved compared with inexperienced raters. Although estimates of severity were improved for all raters regardless of their experience, the biggest reduction in bias and gains in accuracy, precision and agreement were obtained for inexperienced raters (Fig. 5).

The equivalence test revealed that inter-rater reliability measured by both the coefficient of determination (R^2) and the intra-cluster coefficient (ρ) was significantly improved when the raters used SADs (Table 3). The mean coefficient determination (R^2) among inexperienced raters increased from 0.76 to 0.84 when using SADs and for experienced raters from 0.81 to 0.93.

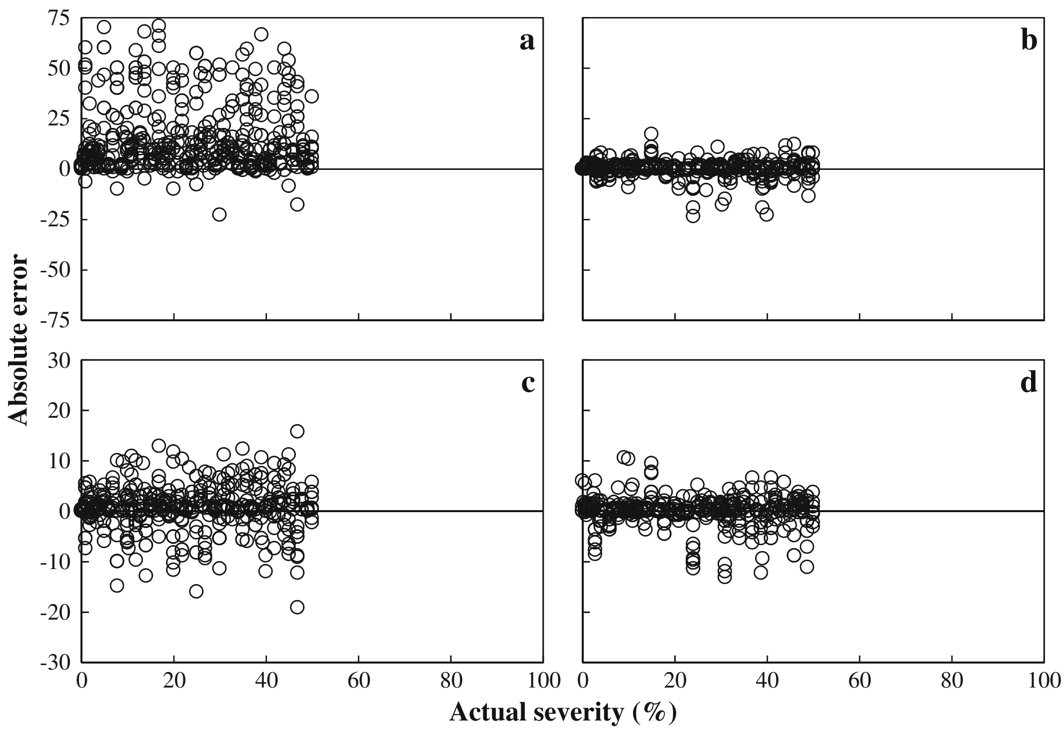

Fig. 4 Absolute error (estimated severity minus actual severity) of estimates of severity of frogeye leaf spot on 50 diseased soybean leaflets for each of ten inexperienced (**a**, **b**) and ten experienced raters (**c**, **d**), without (**a**, **c**) and with (**b**, **d**) the use of a set of standard area diagrams (SADs). Low absolute errors indicate that the estimated severity was similar to the actual severity

Table 2 Effect of rater experience on bias, precision and overall agreement of estimates of severity of frogeye leaf spot made by raters either unaided or aided by a set of standard area diagrams (SADs)

Assessment	LCC statistic	Means		95 % CI[a] of the difference between means
		Inexperienced	Experienced	
No SADs	Scale shift (υ)[b]	1.93	0.93	**−1.388, −0.589**
	Location shift (μ)[c]	0.85	0.05	**−1.049, −0.507**
	Bias correction factor (C_b)[d]	0.64	0.98	**0.220, 0.449**
	Correlation coefficient (r)[e]	0.89	0.94	**0.023, 0.078**
	Concordance coefficient (ρ_c)[f]	0.57	0.92	**0.238, 0.455**
With SADs	Scale shift (υ)	0.93	0.92	−0.071, 0.041
	Location shift (μ)	−0.01	−0.01	−0.068, 0.071
	Bias correction factor (C_b)	0.99	0.99	−0.004, 0.009
	Correlation coefficient (r)	0.95	0.97	**0.006, 0.050**
	Concordance coefficient (ρ_c)	0.94	0.97	**0.006, 0.055**

[a] Bootstrap calculated difference between means and confidence intervals (CIs). If the CIs embrace zero, difference is not significant at the 5 % level. Bold numbers represent significance of the difference

[b] Scale or slope shift (systematic bias) relative to the perfect relationship (1=perfect relation between x and y)

[c] Location or height shift (constant bias) relative to the perfect relationship (0=perfect relation between x and y)

[d] Bias correction factor that measures how far the best-fit line deviates from a line at 45°. No deviation from the 45° line occurs when C_b=1. C_b is calculated from υ and μ and is a measure of accuracy

[e] Correlation coefficient (r) that measures precision

[f] Lin's concordance correlation coefficient (ρ_c) combines both precision (r) and accuracy (C_b) ($\rho_c = r.C_b$) to measure agreement with the actual value (Lin 1989)

Without use of SADs, 72.2 % of the rater pairwise comparisons had a coefficient of determination <0.85, but when SADs were used, only 2.1 % of the pairwise comparisons had a coefficient of determination <0.85 (Fig. 3f). The intra-class correlation coefficient (ρ) for inexperienced raters increased from 0.78 to 0.91 and from 0.88 to 0.95 for experienced raters, indicating that use of the SADs aid resulted in greater inter-rater reliability irrespective of experience.

Discussion

Several studies have previously demonstrated that use of SADs improves the accuracy of raters to estimate the severities of leaf diseases including Asian rust on soybean (Godoy et al. 2006), blast (Rios et al. 2013) and brown spot (Domiciano et al. 2014) on wheat, brown spot on rice (Lenz et al. 2010), early blight on potato (Duarte et al. 2013) and rust on coffee (Capucho et al. 2011). Despite the increase in importance of FLS on soybean, no adequate method has been developed to help ensure accurate estimation of severity. Accurate methods of assessment are required to minimize the risk of type II errors in hypothesis testing (Bock et al. 2010) and thus are very important when the data is to be analyzed to compare control strategies and also for ensuring accurate estimates of severity for other studies on the biology of *C. sojina*. The SADs developed and validated in this study to assess severity of FLS was demonstrated to provide great improvements in accuracy, precision and reliability of estimates of severity of FLS, and when used, should ensure severity estimates are closer to the actual values.

The SADs proposed as an aid to assess severity of FLS in this study utilized eight color images of leaflets with severity values ranging from 0.1 to 39.9 %. This severity range is considered quite sufficient to embrace the range of severities of FLS observed in the field. Furthermore, severities <0.1 % can be difficult to identify and have little epidemiological value; and leaflet senescence was found to occur at severities near 40 %. The eight severities represented in the SADs developed in the present study are

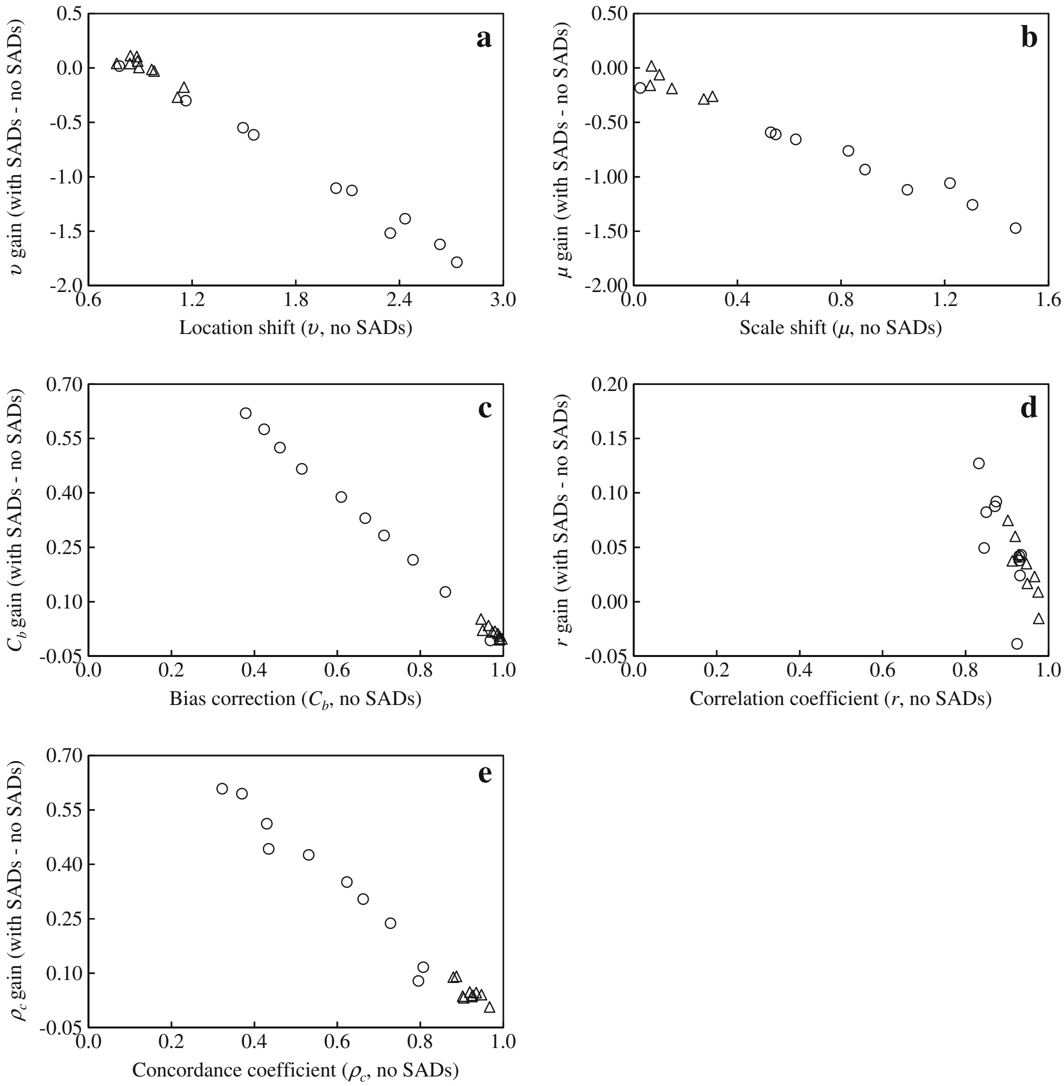

Fig. 5 The relationship between gain [difference between the estimate with and without use of a set of standard area diagrams (SADs)] for measures of (**a**) location shift (or systematic bias, υ), (**b**) scale shift (or constant bias, μ), (**c**) bias correction factor (C_b), (**d**) correlation coefficient (r), and (**e**) Lin's concordance correlation coefficient (ρ_c) of visual severity estimates made by 10 inexperienced (*circles*) and 10 experienced (*triangles*) raters for a set of 50 images of frogeye leaf spot-diseased soybean leaflets

considered adequate to give accurate and rapid estimates of severity of FLS. Similar numbers of represented severities have been employed in SADs developed and validated for other host-pathogen interactions, which are in practical use (Corrêa et al. 2009; Yadav et al. 2013; González-Domínguez et al. 2014). The eight severities follow an approximately linear increase, in contrast to some SADs that use a logarithmic arrangement based on the so-called Weber-Fechner law, which assumes a logarithmic relationship between estimates and actual severity (Horsfall and Barrat 1945; Godoy et al. 2006; Lenz et al. 2010). However, this law was never substantiated, and recent studies have questioned its usefulness for providing good estimates of disease severity (Nutter and Esker 2006; Bock et al. 2010). Therefore, consistent with the SADs that were recently developed (Domiciano et al. 2014; Duarte et al. 2013; Rios et al. 2013; Yadav et al. 2013; González-Domínguez et al. 2014), a linear relationship between estimates and actual severity of FLS was adopted as the rationale for the SADs described here.

Table 3 Inter-rater reliability of assessments by 10 inexperienced and 1experienced raters of frogeye spot on 50 soybean leaflets both without and with use of a set of standard area diagrams (SADs) as an assessment aid. Inter-rater reliability was measured using the intra-class correlation coefficient (ρ) and coefficient of determination (R^2)

Statistics	Inexperienced raters (n=10)		Experienced raters (n=10)	
	No SADs	With SADs	No SADs	With SADs
Intra-class correlation coefficient, ρ (95 % CI)[a]	0.784 (0.709 – 0.853)	0.912 (0.876 – 0.943)	0.882 (0.835 – 0.922)	0.959 (0.941 – 0.974)
Mean inter-rater coefficient of determination (R^2)[b] (95 % CI)	0.759 (0.733 – 0.783)	0.842 (0.823 – 0.861)	0.808 (0.785 – 0.829)	0.926 (0.918 – 0.935)

[a] 2,000 bootstrap samples were used to obtain the confidence intervals (CIs). If the CIs embrace zero, the difference was not significant (P=0.05)

[b] Mean coefficients of determination estimated from pairwise comparisons of assessments by visual raters

Overestimation of disease severity by raters is well-documented (Lenz et al. 2010; Capucho et al. 2011; Godoy et al. 2006; Rios et al. 2013; Domiciano et al. 2014). In the present study, raters (primarily the inexperienced) overestimated severity of FLS when they did not use SADs. However, when raters used SADs as an aid in severity estimation (irrespective of their experience) not only was accuracy greatly improved, but some (60 % of the raters) actually underestimated severity of FLS (negative μ values). Underestimates were most pronounced at severities of FLS above 20 %, which is consistent with some previous reports (Duarte et al. 2013; González-Domínguez et al. 2014). Furthermore, absolute errors of the estimates were greatly reduced by using SADs, which was particularly evident for inexperienced raters. These results are consistent with the effects of the use of SADs or computer-based assessment training in improving accuracy of disease estimates (Nutter and Schultz 1995).

For most raters scale shift (υ) and location shift (μ) were closer to 0 and 1, respectively, when using SADs, indicating that the slope and height of the fitted line approached the line of concordance. However, the gain in these parameters for experienced raters who used SADs was negligible, as has been noted before (González-Domínguez et al. 2014). However, the other measures of precision, accuracy and agreement (r; C_b and ρ_c) were all improved by using SADs for all raters (irrespective of experience), indicating that overall the estimates of severity of FLS were more accurate and precise when raters used SADs as an aid to assess severity of FLS. These results demonstrate the value of the SADs we developed as an aid for estimating severity of FLS, even for experienced raters.

Many studies showed that rater experience, familiarity with symptoms and training in disease assessment influence the reliability and accuracy of disease estimates (Bardsley and Ngugi 2013). In general, experienced raters produced estimates that were more accurate and precise than inexperienced raters, which is consistent with previously reported results (Nita et al. 2003; Bock et al. 2009; Yadav et al. 2013; González-Domínguez et al. 2014). According to Yadav et al. (2013), an inexperienced rater may inherently have a high degree of accuracy and precision, but most often the estimates of inexperienced raters deviate more from the actual severity compared to experienced rater estimates. Thus inexperienced raters are expected to benefit most from using SADs, which also improve inter-rater reliability (Yadav et al. 2013).

The proposed SAD should be a valuable tool where estimates of severity of FLS are required, including in experiments comparing disease management treatments, for epidemiological studies or for monitoring disease for decision-making purposes.

Acknowledgments Prof. F. A. Rodrigues thanks the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for his fellowship. Mr. D. Debona and Mr. J. A. Rios were supported by CNPq and Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG), respectively. This study was supported by grants to Prof. F. A. Rodrigues from CAPES, CNPq and FAPEMIG.

Conflicts of Interest Authors declare that there are no conflicts of interest regarding this study.

References

Akem, C. N., & Dashiell, K. E. (1994). Effect of planting date on severity of frogeye leaf spot and grain yield of soybeans. *Crop Protection, 13*, 607–610.

Bardsley, S. J., & Ngugi, H. K. (2013). Reliability and accuracy of visual methods used to quantify severity of foliar bacterial spot symptoms on peach and nectarine. *Plant Pathology, 62*, 460–474.

Bergamin Filho, A., & Amorim, L. (1996). *Doenças de plantas tropicais: epidemiologia e controle econômico*. São Paulo: Editora Agronômica Ceres.

Bock, C. H., Gottwald, T. R., Parker, P. E., Cook, A. Z., Ferrandino, F., Parnell, S., & van den Bosch, F. (2009). The Horsfall-Barratt scale and severity estimates of citrus canker. *European Journal of Plant Pathology, 125*, 23–38.

Bock, C. H., Poole, G., Parker, P. E., & Gottwald, T. R. (2010). Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *Critical Reviews in Plant Sciences, 29*, 59–107.

Bock, C. H., Wood, B. W., van den Bosch, F., Parnell, S., & Gottwald, T. R. (2013). The effect of Horsfall-Barratt category size on the accuracy and reliability of estimates of pecan scab severity. *Plant Disease, 97*, 797–806.

Campbell, C. L., & Madden, L. V. (1990). *Introduction to plant disease epidemiology* (p. 532). New York: Wiley.

Capucho, A. S., Zambolim, L., Duarte, H. S. S., & Vaz, G. R. O. (2011). Development and validation of a standard area diagram set to estimate severity of leaf rust in *Coffea arabica* and *C. canephora*. *Plant Pathology, 60*, 1144–1150.

Corrêa, F. M., Bueno Filho, J. S. S., & Carmo, M. G. F. (2009). Comparison of three diagrammatic keys for the quantification of late blight in tomato leaves. *Plant Pathology, 58*, 1128–1133.

Dashiell, K. E., & Akem, C. N. (1994). Yield losses in soybeans from frogeye leaf spot caused by *Cercospora sojina*. *Crop Protection, 10*, 465–468.

Domiciano, G. P., Duarte, H. S. S., Moreira, E. N., & Rodrigues, F. A. (2014). Development and validation of a set of standard area diagrams to aid in estimation of spot blotch severity on wheat leaves. *Plant Pathology, 63*, 922–928.

Duarte, H. S. S., Zambolim, L., Capucho, A. S., Nogueira Júnior, A. F., Rosado, A. W. C., Paul, P. A., Cardoso, C. R., & Mizubuti, E. S. G. (2013). Development and validation of a set of standard area diagrams to estimate severity of potato early blight. *European Journal of Plant Pathology, 137*, 249–257.

Gamer, M., Lemon, J., & Fellows, I. (2012). *irr: Various Coefficients of Interrater Reliability and Agreement*. R Package Version 0.84.

Godoy, C. V., Koga, L. J., & Canteri, M. G. (2006). Diagramatic scale for assessment of soybean rust severity. *Fitopatologia Brasileira, 31*, 63–68.

González-Domínguez, E., Martins, R. B., Del Ponte, E. M., Michereff, S. J., García-Jiménez, J., & Armengol, J. (2014). Development and validation of a standard area diagram set to aid assessment of severity of loquat scab on fruit. *European Journal of Plant Pathology, 139*, 413–422.

Horsfall, J. G., & Barrat, R. W. (1945). An improved grading system for measuring plant disease. *Phytopathology, 35*, 655.

Horsfall, J. G., & Cowling, E. B. (1978). Pathometry: the measurement of plant disease. In J. G. Horsfall & E. B. Cowling (Eds.), *Plant disease: An advanced treatise* (Vol. II, pp. 20–136). New York: Academic.

James, W. C. (1971). An illustrated series of assessment keys for plant diseases, their preparation and usage. *Canadian Plant Disease Survey, 51*, 39–65.

Kim, H., Newell, A. D., Cota-Sieckmeyer, R. G., Rupe, J. C., Fakhoury, A. M., & Bluhm, B. H. (2013). Mating-type distribution and genetic diversity of *Cercospora sojina* populations on soybean from Arkansas: evidence for potential sexual reproduction. *Phytopathology, 103*, 1045–1051.

Klosowski, A. C., Ruaro, L., Bespalhok Filho, J. C., & May De Mio, L. L. (2013). Proposal and validation of diagrammatic scale for assessment of orange rust of sugarcane. *Tropical Plant Pathology, 38*, 166–171 (in Portuguese).

Kranz, J. (1988). Measuring plant disease. In J. Kranz & J. Rotem (Eds.), *Experimental techniques in plant disease epidemiology* (pp. 35–50). New York: Springer.

Lenz, G., Balardin, R. S., Corte, G. D., Marques, L. N., & Debona, D. (2010). Diagrammatic scale for assessment of rice brown spot severity. *Ciência Rural, 40*, 752–758 (in Portuguese).

Lin, L. I. K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics, 45*, 255–268.

Madden, L. V., Hughes, G., & van den Bosch, F. (2007). *The study of plant disease epidemics* (p. 421). St. Paul: APS Press.

Mian, R., Bond, J., Joobeur, T., Mengistu, A., Wiebold, W., Shannon, G., & Wrather, A. (2009). Identification of soybean genotypes resistant to *Cercospora sojina* by field screening and molecular markers. *Plant Disease, 93*, 408–411.

Michereff, S. J., Maffia, L. A., & Noronha, M. A. (2000). Diagrammatic scale for assessment of the severity of leaf blight of taro. *Fitopatologia Brasileira, 25*, 612–629 (in Portuguese).

Nascimento, K. J. T., Debona, D., França, S. K. S., Gonçalves, M. G. M., DaMatta, F. M., & Rodrigues, F. A. (2014). Soybean resistance to *Cercopora sojina* infection is not enhanced by silicon. *Phytopathology, 104*, 1183–1191.

Nita, M., Ellis, M. A., & Madden, L. V. (2003). Reliability and accuracy of visual estimation of phomopsis leaf blight of strawberry. *Phytopathology, 93*, 995–1005.

Nutter, F. W., & Esker, P. D. (2006). The role of psychophysics in phytopathology. *European Journal of Plant Pathology, 114*, 199–213.

Nutter, F. W., & Schultz, P. M. (1995). Improving the accuracy and precision of disease assessments: selection of methods and use of computer-aided training programs. *Canadian Journal of Plant Pathology, 17*, 174–184.

Nutter, F. W., Teng, P. S., & Shokes, F. M. (1991). Disease assessment term and concepts. *Plant Disease, 75*, 1187–1888.

Pedroso, C., Lage, D. A. C., Henz, G. P., & Café-Filho, A. C. (2011). Development and validation of a diagrammatic scale for estimation of anthracnose on sweet pepper fruits for epidemiological studies. *Journal of Plant Pathology, 93*, 219–225.

R Core Team 2013. *R: A language & environment for statistical computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org/.

Rios, J. A., Debona, D., Duarte, H. S. S., & Rodrigues, F. A. (2013). Development and validation of a standard area diagram set to assess blast severity on wheat leaves. *European Journal of Plant Pathology, 136*, 603–611.

Ross, J. P. (1968). Additional physiological races of *Cercospora sojina* on soybeans in North Carolina. *Phytopathology, 58*, 708–709.

Sachs, P. J. D., Neves, C. C. S. V. J., Canteri, M. G., & Sachs, L. G. (2011). Diagrammatic scale for assesment of the Phaeosphaeria leaf spot severity in maize. *Summa Phytopathologica, 37*, 202–204.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.

Stevenson, M. (2012). *Epir: An R Package for the Analysis of Epidemiological Data*. R package version 0.9-43.

Vale, F. X. R., Fernandes Filho, E. I., & Liberato, J. R. (2003). QUANT: a software plant disease severity assessment. In R. Close, M. Braithwaite, & I. Havery (Eds.), *Proceedings of the 8th International Congress of Plant Pathology* (pp. 105). New Zealand.

Yadav, N. V., Vos, S. M., Bock, C. H., & Wood, B. W. (2013). Development and validation of standard area diagrams to aid assessment of pecan scab symptoms on fruit. *Plant Pathology, 62*, 325–335.

Yi, Q., Wang, P. P., & He, Y. (2008). Reliability analysis for continuous measurements: equivalence test for agreement. *Statistics in Medicine, 27*, 2816–2825.

Yorinori, J. T., & Klingelfuss, L. H. (1999). Novel races of *Cercospora sojina* on soybean. *Fitopatologia Brasileira, 24*, 509–512 (in Portuguese).

Zhang, G., Pedersen, D. K., Phillips, D. V., & Bradley, C. A. (2012). Sensitivity of *Cercospora sojina* isolates to quinone outside inhibitor fungicides. *Crop Protection, 40*, 63–68.