



How much do standard area diagrams improve accuracy of visual estimates of the percentage area diseased? A systematic review and meta-analysis

Emerson M. Del Ponte¹ · Luis Ignacio Cazón¹ · Kaique S. Alves¹ · Sarah J. Pethybridge² · Clive H. Bock³

Received: 1 July 2021 / Accepted: 26 October 2021 / Published online: 11 January 2022
© The Author(s), under exclusive license to Sociedade Brasileira de Fitopatologia 2022

Abstract

Plant disease severity, when expressed as percentage area diseased, is commonly estimated visually without or with the aid of standard area diagram sets (SADs). It is generally thought that the use of SADs leads to more accurate and thus more precise estimates, but the degree of improvement has not been characterized in a systematic manner. We built on a previous review and screened 153 SAD studies published from 1990 to 2021. A systematic review resulted in a selection of 72 studies that reported three linear regression statistics for individual raters, which are indicative of the two components of bias (intercept = constant bias; slope = systematic bias) and precision (Pearson's correlation coefficient, r), to perform a meta-analysis of these accuracy components. The meta-analytic model determined an overall gain of 0.07 (r increased from 0.88 to 0.95) in precision from using SADs. Overall, there was a reduction of 2.65 points in the intercept (from 3.41 to 0.76) indicating a reduction in the constant bias. Slope was less influenced and was reduced slightly (from 1.09 to 0.966), indicating a marginal reduction in systematic bias when using SADs. A multiple correspondence analysis suggested an association of less accurate, unaided estimates with diseases that produce numerous lesions and for which maximum severity of 50% is seldom reached. In contrast, estimates of severity for diseases that cause only a few lesions and those diseases where the lesions coalesce and occupy more than 50% of the organ surface had greater accuracy, which was most pronounced for specimen types other than leaves. By quantitatively exploring how characteristics of the pathosystem and how SADs affect precision and constant and systematic biases, we affirm the value of SADs for reducing bias and imprecision of visual assessments. We have also identified situations where SADs have greater or lesser utility as an assessment aid.

Keywords Diagrammatic scales · Phytopathometry · Visual assessment

Introduction

The choice of the most appropriate method for quantifying plant disease depends on various factors including the objective of the work, availability of equipment and feasibility, spatial scale (organ, plant, plot or field), as well as time and

cost constraints (Madden et al. 2007). Measures of disease intensity include prevalence, incidence and severity. Plant disease severity, or the degree to which a specimen (plant or plant part) is diseased (Bock et al. 2021a), can be expressed by various means. A commonly used metric of disease severity is the percentage area of the considered plant tissue affected by the presence of disease symptoms on the host or signs of the pathogen (Nutter et al. 1991). Disease severity is particularly valuable to (1) quantify and understand plant disease epidemiology, (2) quantify yield losses, (3) evaluate control methods, and (4) screen plant genotypes for host resistance (Bock et al. 2016).

Disease severity based on the percentage of plant tissue area visibly diseased is most commonly estimated via visual assessment using the human eye as a remote sensor (Bock et al. 2010, 2016). Despite the recent advances in technology and availability of tools based on remote sensing and image

✉ Emerson M. Del Ponte
delponte@ufv.br

¹ Departamento de Fitopatologia, Universidade Federal de Viçosa, Viçosa, MG 36570-900, Brazil

² Plant Pathology & Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell AgriTech, Cornell University, Geneva, NY 14456, USA

³ United States Department of Agriculture—Agriculture Research Service—Southeastern Fruit and Tree Nut Research Station, Byron, GA 31008, USA

analysis (e.g., image segmentation by color thresholding), for measuring disease severity, visual assessment prevails, especially in field research (Bock et al. 2020, 2021b). Visually, a rater perceives and assesses the percentage of diseased area (via estimation) or, in some cases, depending on the type of scale used, applies a class grade based on an ordinal scale that encompasses a distinct pattern or a percentage interval (Bock et al. 2016; Chiang et al. 2014).

Estimation of percentage areas is deemed more challenging than classification of severity into ordinal classes, considering the large number of options to choose from on the percentage scale, compared to the finite and small number of classes in ordinal scales (Large 1966; James 1974). Moreover, estimation is subjective and known to vary considerably among raters (Nutter 1993; Bock et al. 2009). Specific tools have been proposed to train raters prior to and to aid them during the assessment process so that the visual estimates are as accurate as possible. Accuracy of a severity estimate is the closeness to the actual value; therefore a qualitative concept that can only be assessed in relation to what are accepted as the “actual,” “true,” or “gold standard” values (Bock et al. 2016). Consistently accurate estimates of severity will be defined as “precise” (statistically less variable), but reliable or precise estimates are not necessarily accurate (Madden et al. 2007). Hence, precision is a second important component of overall accuracy. The definition and importance of accuracy in visual estimates have been recently reviewed and discussed (Bock et al. 2016).

Standard area diagram sets (SADs), or a similar pictorial or graphic representation of selected disease severities on a plant organ, are among the aids used to calibrate a raters’ eyes for performing a visual assessment (Del Ponte et al. 2017). SADs were first developed over a century ago but have received considerable attention during the last quarter-century during the modern era of plant pathology research (Del Ponte et al. 2017; Bock et al. 2020). Indeed, from a handful of black and white hand drawings of diseased specimens developed at the end of the nineteenth century (Cobb 1892), recent research has proposed more elaborate SADs that have been validated—i.e., demonstrated to increase accuracy and reliability of visual estimates compared with unaided estimates (Del Ponte et al. 2017). In the last 20 years, plant pathologists have taken advantage of advances in image processing and analysis tools and from knowledge gained from the psychophysical and measurement sciences to develop SAD sets that are realistic (e.g. true color photographs), with appropriate validation and illustrated severities to maximize estimation accuracy (Domiciano et al. 2014; Schwanck and Del Ponte 2014; Araújo et al. 2019; Franceschi et al. 2020).

To date, research in the field has focused largely on the proposition and validation of new SADs for diseases that lack the tool and much less on improving upon existing ones

(Domiciano et al. 2014; Schwanck and Del Ponte 2014; Araújo et al. 2019) or exploring the effects of SAD-specific factors (e.g., diagram appearance and number, scale structure, instructions) or other aspects that may affect rater performance including symptomatic patterns (Domiciano et al. 2014; Schwanck and Del Ponte 2014; Bock et al. 2015; de Melo et al. 2020). A list of future research needs and best operating practices for developing and using SADs has been prepared to guide researchers in the design and evaluation of more effective systems that will ultimately lead to improved accuracy (Del Ponte et al., 2017; Bock et al. 2020). Since our previous review published on the topic (Del Ponte et al. 2017), 50 further studies have been published. A qualitative evaluation suggests the use of SADs results in less bias and consequently greater precision. But the degree of improvements in the accuracy and precision of estimates, the circumstances of rating, and characteristics of symptoms (size, number, distribution, and maximum severity) under which SADs have proven most valuable has not been characterized in a systematic manner. However, some early research indicated effects of lesion size, number, and distribution (Amanat 1976; Kranz et al. 1977; Sherwood 1983; Forbes and Jeger 1987; Hock et al. 1992).

A common statistical method to investigate the overall accuracy of visual estimates of severity is Lin’s concordance correlation coefficient (LCCC) (Lin 1989), which is the product of precision (as Pearson’s correlation coefficient, r) and generalized bias (the product of scale-shift [systematic bias] and location-shift [constant bias] measures) (Nita et al. 2003; Madden et al. 2007). Prior to LCCC’s first use in SAD research (Spolti et al. 2011; Yadav et al. 2013), linear regression models were the statistical standard used to relate rater estimates to the actual severity values (see reviews by Del Ponte et al. 2017; Bock et al. 2020). Indeed, linear regression continues to be used for evaluating SADs (Santos et al. 2017; Camara et al. 2018; Trojan and Pria 2018; Arias et al. 2020; Robaina et al. 2020; Kublik et al. 2020). The two linear regression coefficients (intercept and slope), together with the coefficient of determination (R^2), are usually reported for each rater in the SAD validation study, while those researchers reporting LCCC statistics have most often reported means of all raters combined (Del Ponte et al. 2017; Bock et al. 2020; Brás et al. 2020; Nascimento et al. 2020; Rivera et al. 2020; Castellar et al. 2021; Montero et al. 2021).

When using linear regression, the constant (intercept, a) and the first regression coefficient (slope, b) represent two measures of bias, constant bias and systematic bias, respectively. The closer a is to zero and b is to one, the smaller the bias, and thus the closer the visual estimates of percentage area are to the actual values (Teng 1981). The two coefficients considered separately do not inform us how close the estimates are to the actual severity values. The overall

variability in estimates in relation to the actual values is provided by the coefficient of determination (R^2). By taking the root-square, we obtain Pearson's correlation coefficient, r , which is a measure of precision and a component of the LCCC (Madden et al. 2007). In contrast to LCCC, which provides a global measure of concordance (or overall accuracy), precision and the two measures of bias in linear regression are evaluated independently.

The current study is a sequel to our previous systematic review of those SADs that had been published in peer-reviewed journals up to 2017 (Del Ponte et al. 2017). We have since updated the database to include all additional published articles since 2017. Firstly, we considered all peer-reviewed SAD studies to systematically select studies and conduct a meta-analysis to obtain mean estimates of each of the three linear regression coefficients indicative of the two components of bias and precision for the visual estimates that were published in the original articles. Secondly, we explored factors related to disease symptom characteristics and SADs that could explain the variability in precision and bias of the unaided estimates as well as the gains in precision and reduction of bias resulting from use of SAD aids.

Material and methods

Data sources and variables used

In addition to the variables listed in our previous systematic review of published SADs (Del Ponte et al. 2017), we have included additional numeric variables representing the bias and precision of the visual estimates. Data on the regression statistics (intercept, slope and coefficient of determination) associated with the estimates of severity were obtained from each of the peer-reviewed studies in which SADs had been validated in a systematic manner, and for which the data existed for each rater in the study. Using the same bibliographic search criteria described in our previous review on the topic (Del Ponte et al. 2017), we were able to locate, catalogue, and extract information from an additional 48 articles published since 2017, which resulted 153 (105 + 48) articles that validated SADs published between 1990 and 2021, regardless of the disease or host plant species. Finally, we classified each SAD in the 153 studies according to a nominal scale of symptomatic patterns created specifically for this study (details provided in a later section).

Criteria for study and rater selection

We systematically selected the studies for analysis following objective criteria (Fig. 1). We first queried our database to select studies where summary statistics (intercept, slope and coefficient of determination) of severity estimates were

available by individual rater and obtained both the data for the unaided and aided estimates using SADs. Nine studies were not included due to a lack of validation, and 19 additional studies were not included due to lack of unaided assessments. We excluded a further 32 studies for which the linear regression statistics were not available by individual rater. A further filter screening rater number was applied to ensure a minimum of six raters participated in the SAD validation. Based on the final screen, a further 15 studies were excluded. Thus, 78 studies remained, in which there were 923 raters who assessed severity on a range of diseases without and with SADs.

A preliminary, exploratory analysis of all regression statistics showed outliers, defined as raters with exceptionally high or low intercepts and slopes, relative to zero and one, respectively, or with very low Pearson's correlation coefficients, r (precision). We excluded raters, regardless of study, who exhibited an $r < 0.3$; an intercept value > 30 or < -30 , and a slope > 5 . A total of 75 raters were thus removed, reducing the total number of raters to 848 (8.1% were outliers). Finally, we eliminated a further six studies due to the number of raters being less than six. The final number of studies was 72 and included 823 raters (Fig. 1).

Lesion patterns

Each selected disease represented by a SAD was assigned to a nominal variable based on a descriptive pictorial key for classifying symptomatic patterns developed specifically for this study. The pictorial key comprising a 4-category scale was constructed based on characteristics of the lesions including number and size (single and large, few and medium sized, and numerous and medium to small in size), and the presence or absence of coalescence of the lesions as the disease progressed in the illustrated SADs. The pictorial key (Fig. 2) was used by five independent evaluators to classify each SAD accordingly. The five evaluators are experienced in disease assessment (the authors of this study). The most voted nominal variable applied to a SAD was the one assigned. Maximum severity represented in the SADs for each study was included in the analysis.

Exploratory analysis

First-order linear regression analysis was used in the original articles to make inferences regarding bias and precision of the visual estimates. We graphically explored the distribution of the intercepts and slopes and that of the correlation coefficient (calculated from the coefficient of determination) from the assessments without and with SADs. Subsequently, we averaged each regression statistic across raters and calculated the gain or loss by subtracting the mean value of the unaided estimates from the mean

Fig. 1 Flow diagram describing the steps and criteria for systematically reviewing studies on standard area diagrams (SADs) for inclusion in a meta-analysis of the bias and precision of visual estimates of plant disease severity using SADs

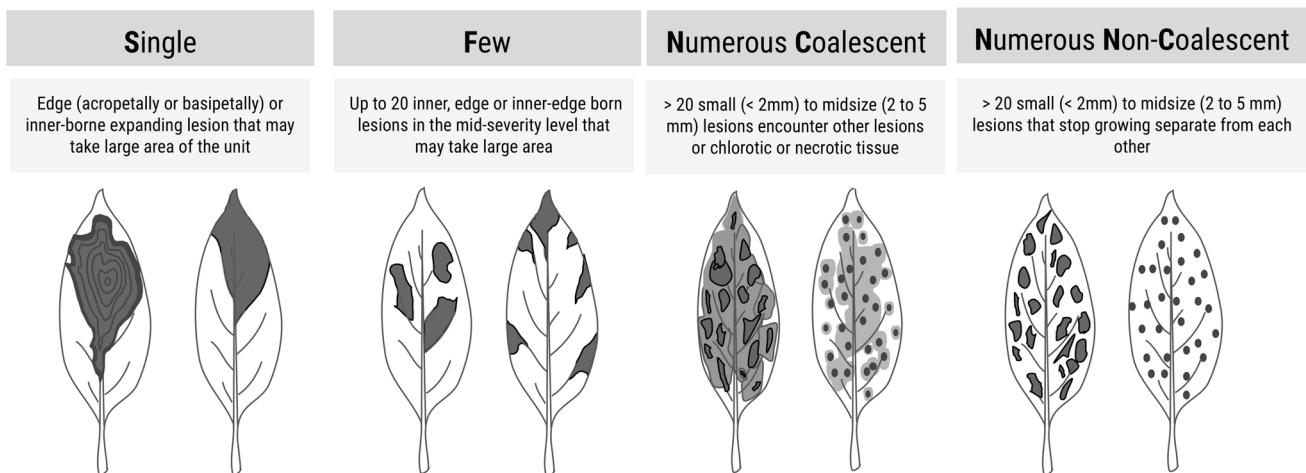
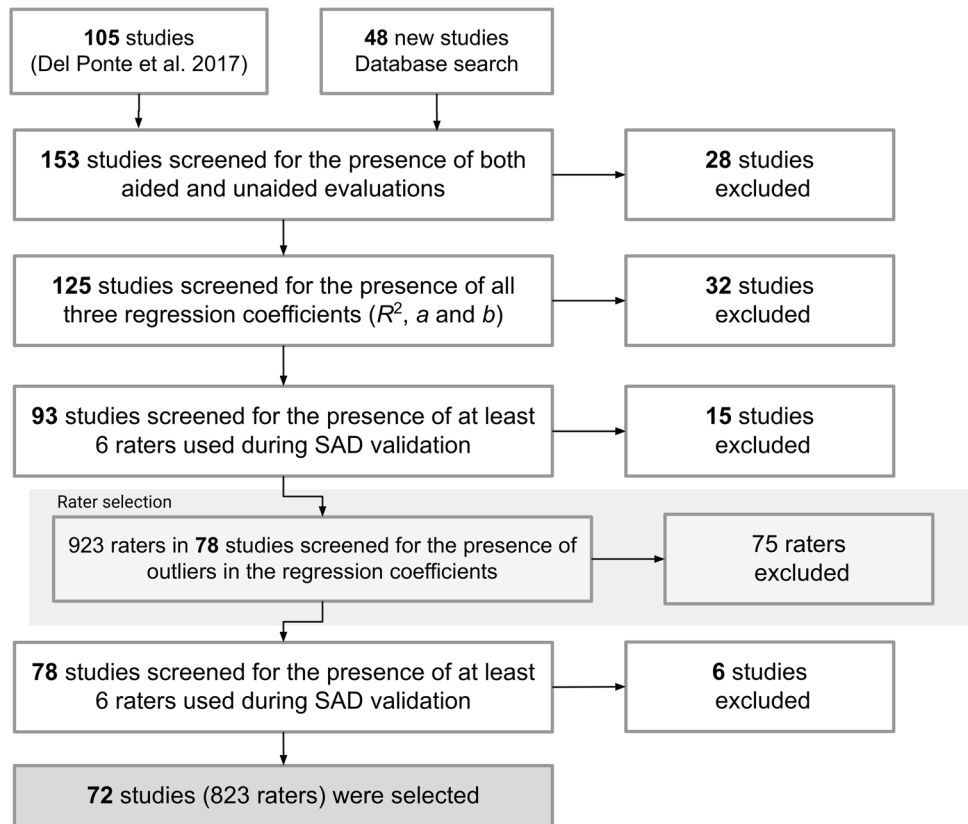


Fig. 2 Pictorial key used for the classification of symptoms based on the number, size, and coalescence of lesions

value for the SAD-aided estimates. For example, a positive or a negative difference in the Pearson's r represents a gain or loss in precision, respectively. The data for each statistic were summarized as means and standard deviations (across raters) both unconditioned (overall means) and conditioned to each study and were explored graphically.

Meta-analysis of regression coefficients

We fitted an arm-based meta-analysis, also called a two-way unconditional linear mixed model (Madden et al. 2016), directly to the means for each of the three statistics. The model can be written as:

$$Y_i \sim N\left(\mu, \sum + S_i\right), \quad (1)$$

where Y_i is the vector of the means of the regression statistic (Pearson's r , intercept, or slope) for the two treatments for the i th study, μ is a vector representing the mean of Y_i across all studies, Σ is a 2×2 between-study variance–covariance matrix, and S_i is the within-study variance–covariance matrix for the i th study. N indicates a multivariate normal distribution. The method (unaided or SAD-aided) was treated as a fixed effect and both the study and method as random effects, the latter to account for the dependency (as the same raters evaluated both without and with the aid). An unstructured Σ matrix was used, and the models were fitted to the data with a maximum-likelihood parameter. We used the inverse of the sampling variance as a weighting variable. This was calculated as the variance of each statistics' value across the raters divided by the number of raters for a specific study. Hence, studies with lower variance (indicating the method was more reproducible among raters) contributed more weight to the meta-analytic estimate. All models were fitted to the data using the *rma.mv* function of the *metafor* package (Viechtbauer 2010) of R (R Core Team 2021).

The model was expanded to include categorical moderator variables that could explain, at least in part, the heterogeneity of the means across studies (Madden et al. 2016). The categorical moderator variables included symptom number-size classification (F: few; Nc: numerous coalescent and Nnc: numerous non coalescent); maximum severity in the SAD ($\leq 50\%$ or $> 50\%$); and organ type (leaf or other). Linear contrasts were used to estimate the mean effect sizes and their standard errors and 95% CIs for each level of the categorical moderator variable (Madden et al. 2016). The gains in precision and accuracy were given by the difference between the estimates using the SAD and those obtained unaided.

Multiple correspondence analysis of the unaided estimates

In addition to the calculation of the meta-analytic estimate for each parameter, we created categorical variables for each regression parameter obtained from the unaided estimates. These were two nominal variables for the intercept as low (≤ 3.4) or high constant bias (> 3.4); positive (> 1) or negative (< 1) systematic bias; and three classes of precision: low (≤ 0.84), moderate (0.84 to 0.93), and high (> 0.93). A multiple correspondence analysis (MCA) was performed, given the nature of multiple variables was nominal, to detect and represent underlying structure in the database (Hjellbrekke 2018). The same categorical variables used for the meta-analysis were implemented in this analysis.

Results

What is the magnitude of the benefit to using SADs?

In general, the Pearson's r values, indicative of precision, were closer to 1 when using SADs (Figs. 3a, b), but the gains (increase in precision) from using the aid varied considerably among the studies, ranging from negative (a loss in precision) in one study to an increase of > 0.2 in five of the studies. The variation in the intercepts (constant bias) was greater when not using the aid compared with the aided estimates and was mostly positive, indicating an overall overestimation of severity (Figs. 3d, e). The intercept values were generally reduced (closer to zero) when using the SAD aid (Fig. 3f). The range of the slope (systematic bias) values was slightly reduced overall but was also increased (greater than 1) in some studies when using the SADs compared with the unaided estimates (Figs. 3g, h, i).

Results of the meta-analysis showed that the intercept was the most affected parameter when using the SADs (Fig. 4). Globally, there was a 2.65 reduction in the intercept, from 3.41 (95% confidence interval (CI) 2.78–4.04) to 0.76 (95% CI 0.45–1.03). However, the 95% CI of the estimate did not span zero, suggesting an overall tendency of raters to overestimate severity even when using a SAD (Fig. 4e). The slope was the least affected statistic, slightly reduced from 1.09 (95% CI 1.008–1.183) to 0.966 (95% CI 0.914–1.018). The 95% CI of the meta-analytic estimates of the slopes without the SAD did not embrace zero suggesting systematic bias was alleviated when using the SAD. Finally, the meta-analytic estimate of precision was significantly increased when using the SAD compared with the unaided estimates. Precision increased, on average, by 0.071, from 0.871 (95% CI 0.849–0.893) to 0.943 (95% CI 0.933–0.953) (Fig. 4c).

Do symptomatic patterns affect bias and precision of unaided estimates?

The three symptom characteristics of the diseases represented in the SADs (size, number and maximum severity) influenced the precision and accuracy of the unaided estimates. Lesion number, coalescence, and maximum severity (Fig. 5) significantly affected the rater statistics. The greater the number of lesions that did not coalesce and were smaller in size, (Figs. 5A–C), combined with a low maximum severity (Fig. 5D,E,F), resulted in lower precision and greater bias of the estimates. The precision of the estimates for specimens with few lesions (< 20 lesions) was most often greater than the meta-analytic estimate but was more variable for the symptoms characterized by numerous coalescent lesions (Fig. 5A).

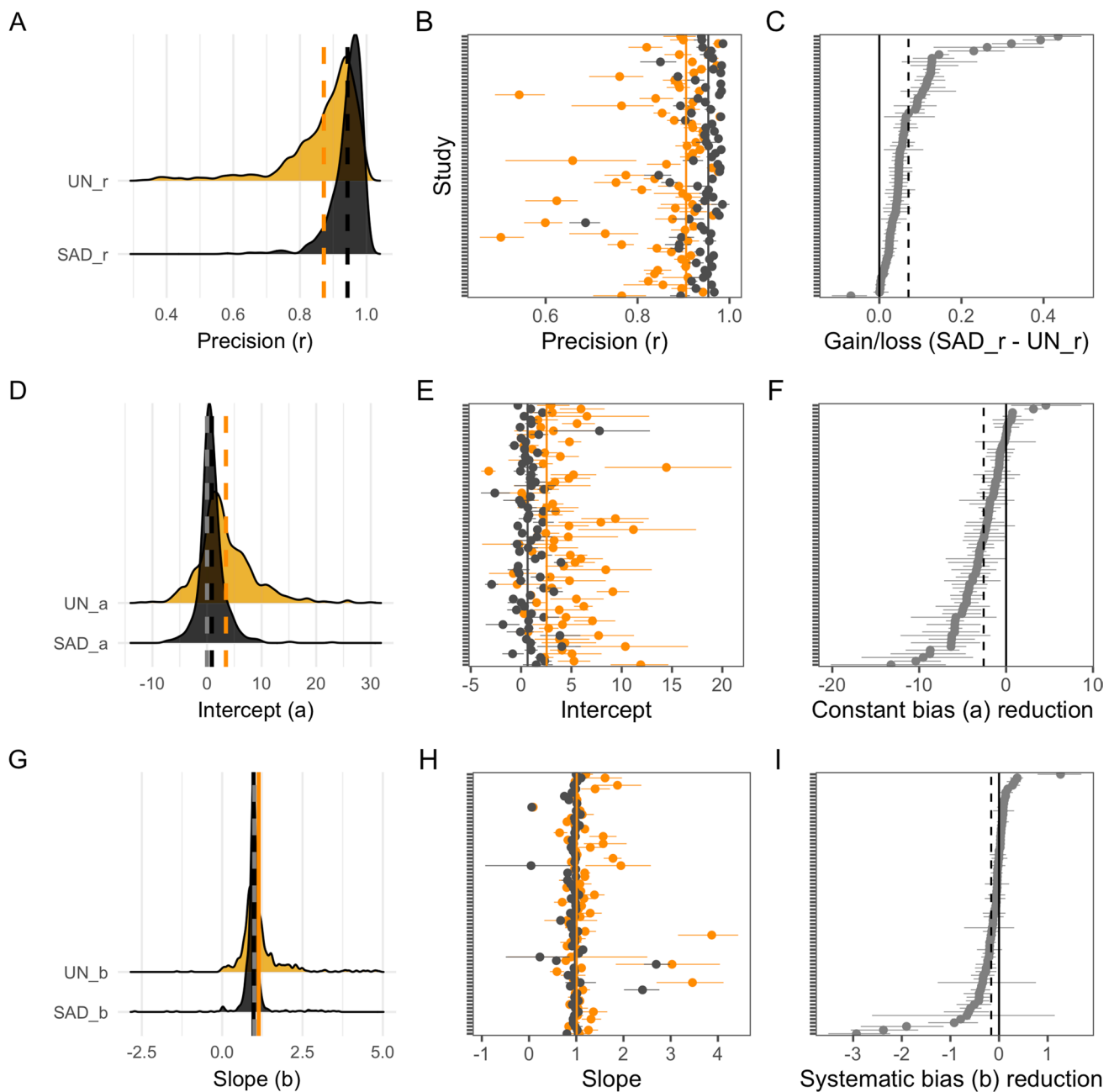


Fig. 3 Density (A, D, G) and dot (B, E, H) plots for the distribution of values of Pearson's r (A, B, C), indicative of precision of visual estimates of severity, and intercepts (a) (D, E, F) and slopes (b) (G, H, I) as indicative of bias, both unaided (UN) or aided by a standard area diagram (SAD) ($n=72$ studies). Plots C, F, and I display the

cumulative distribution of the gain (loss) in precision and reduction (increment) in constant (a, intercept) and systematic bias (b, slope). Each dot represents the means across raters for each published study; error bars indicate the standard deviations of the means

There was a tendency for the unaided estimates to be more precise on plant organs other than leaves (Other), which may be related to the characteristics of the lesion which were most often larger and fewer than on leaves (Leaf) in general (Fig. 5G, H, I; Fig. 6).

The MCA map provides a perspective for the relationships between disease characteristics and organ-related

factors and the magnitudes of Pearson's r and the slope and intercept statistics (Fig. 6). A Pearson's $r > 0.92$ clustered with other organs and a maximum severity $> 50\%$, suggesting an association between these variables (Fig. 6 lower left). Lower precision (< 0.84) was more strongly associated with numerous non-coalescent lesions and a maximum severity $\leq 50\%$ (Fig. 6 upper right). Intercepts > 3.4 were

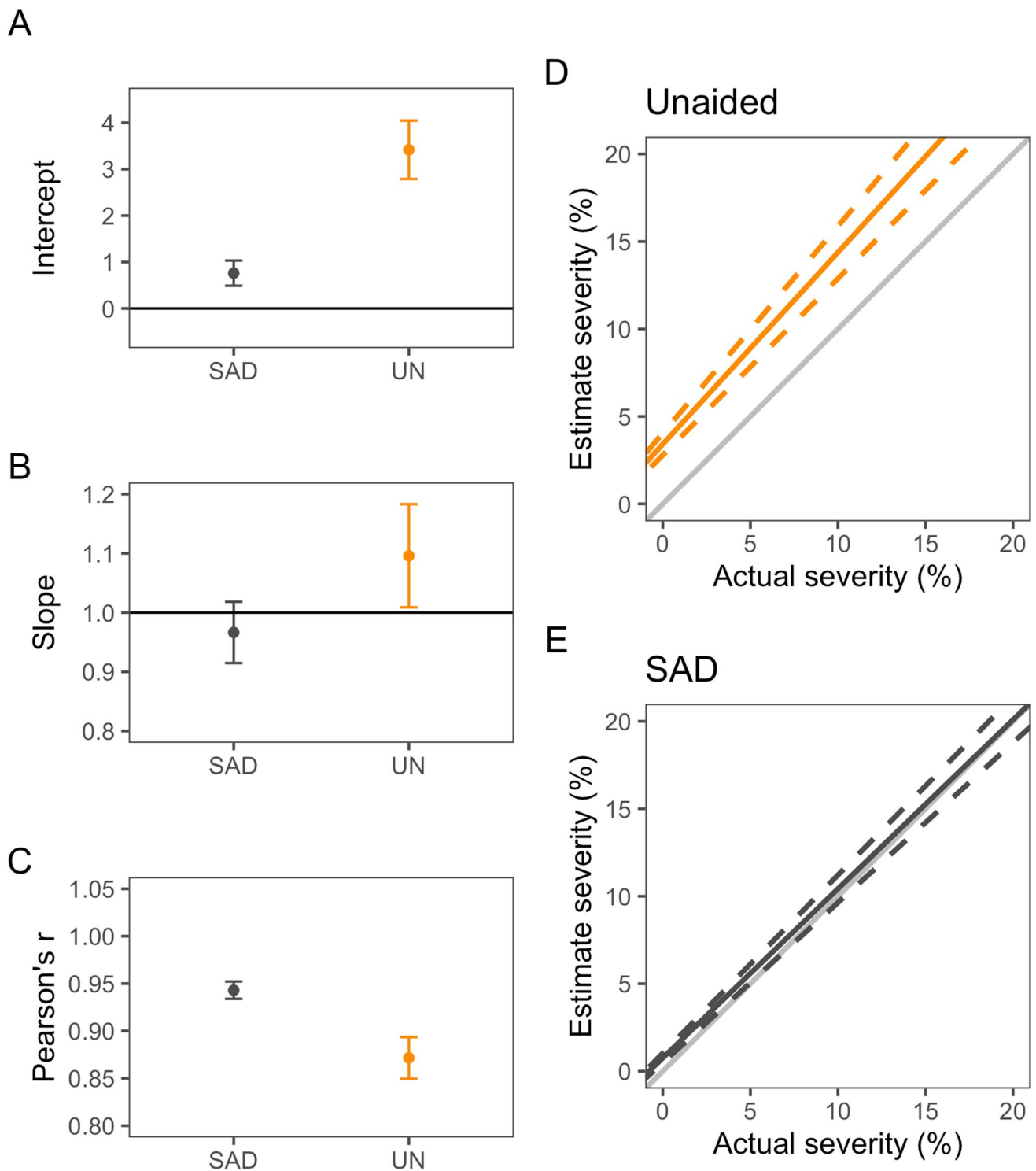


Fig. 4 Overall estimates using a meta-analytic model ($n=72$ studies) for the statistics of linear regression analysis for making inferences regarding accuracy (intercept and slope) (A, B) and precision (Pearson's r) (C) of visual estimates of severity either unaided (UN)

or aided by a standard area diagram (SAD). The predicted line (solid line) and respective 95% confidence interval (CI, dashed lines) using the estimates for the statistics are presented in D and E. Error bars indicate the 95% CI of the estimate

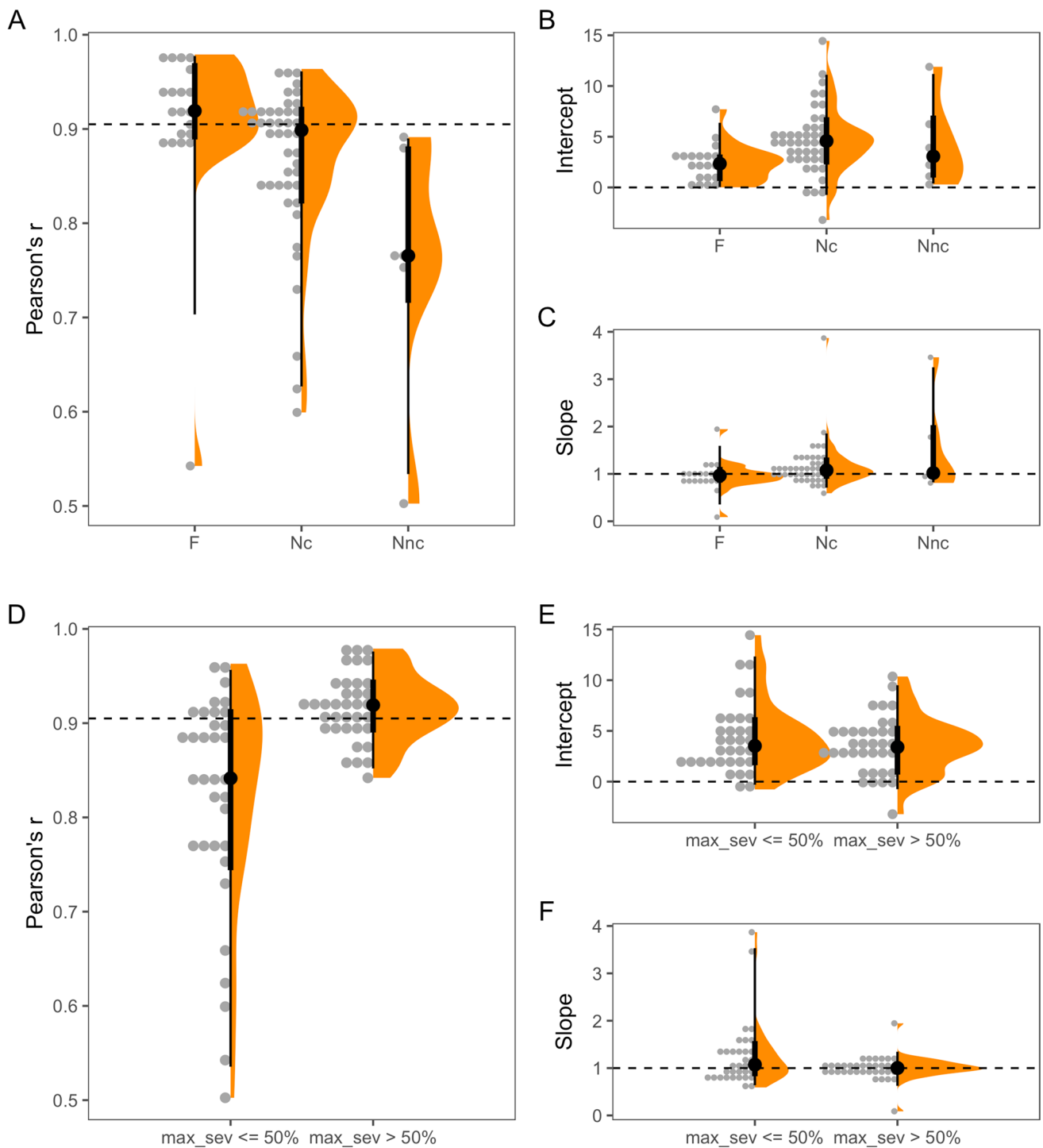


Fig. 5 Half-eye density and dots plots for linear regression statistics indicative of bias (slope and intercept) and precision (Pearson's r) for visual estimates of percentage area affected by disease (severity) obtained unaided (no standard area diagram set, SADs) and conditioned to a classification of symptoms according to size and coales-

cence (A, B, C), maximum severity represented in the disease SAD (D, E, F). Each grey dot represents a published research study on SADs. The black dot in the middle represents the mean and the solid vertical lines the intervals of a frequentist distribution

associated with numerous coalescent lesions (Fig. 6 lower right). Finally, lower bias was associated with fewer lesions (Fig. 6 upper left).

What factors influence bias and precision when using SADs?

We explored the effect of the two disease-specific factors that had the greatest effect on the unaided estimates (lesion number/symptomatic pattern and maximum severity) on estimation bias and precision using SADs. As expected, the gains in precision and reductions in bias were generally greater for diseases that are characterized by numerous lesions (> 20 small lesions), especially if they do not coalesce and rarely reach 50% severity (Fig. 7). It was not clear whether the true-color image SADs or those that had more illustrations of diseased leaves (or other organ type) increased precision and reduced bias of the estimates given the substantial variability across the studies within each defined group (Fig. 8). Finally, there was an inverse relationship between the unaided estimates with the gain/loss in precision and reduction in the constant and systematic biases (Fig. 9). The more biased and imprecise the unaided estimates, the greater the benefits to using SADs, especially for situations where the lesion classification was numerous compared to those classified as few. In only a

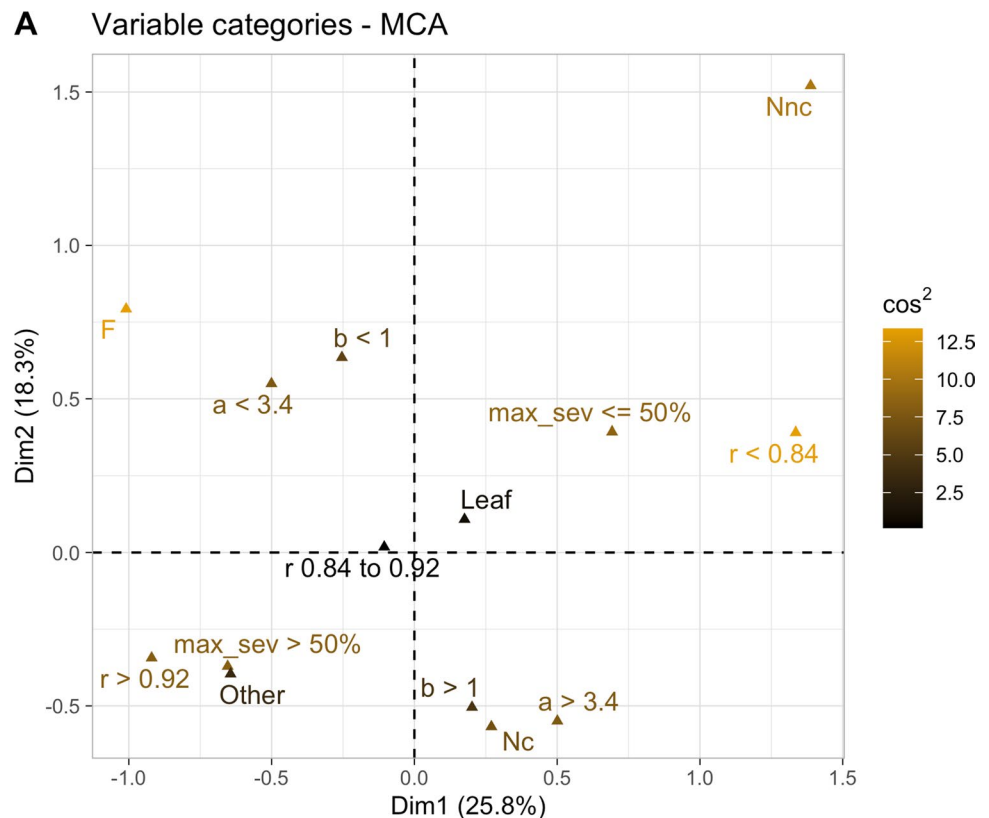
single study was the precision lower when using SADs compared to when estimates were made unaided (Fig. 9a).

Discussion

Our systematic review builds on our previous research (Del Ponte et al. 2017) and has identified 48 new studies in which SADs were developed and evaluated with the aim of increasing accuracy and precision of visual estimates of disease severity. On average, 10 studies have been published each year since 2017, and the total number of publications on SADs catalogued in our database (as of July 2021) is 153 (a searchable interface for SADs articles is publicly available at <https://sadbanks.netlify.app/>). The results confirm that the use of SADs most often results in improved accuracy and precision of visual estimates. Our analysis contributes to an improved understanding of SADs through the consideration of factors related to SAD design and structure, disease symptoms, and actual severity. Combined with other knowledge on SADs, the results should help increase the accuracy and precision of visual estimates (Franceschi et al. 2020; Belan et al. 2020; de Melo et al. 2020; Pereira et al. 2020).

Through meta-analysis, we quantified the overall benefits of SADs and also identified factors that could explain, at least in part, the variability of the summary statistics from

Fig. 6 Multiple correspondence analysis map for the association among classes of linear regression statistics (a =intercept: smaller or larger than 3.4; b =slope: small or larger than 1; and r =Pearson's r : smaller than 0.84, between 0.85 and 0.92 or greater than 0.92) obtained for the relationship between unaided estimates of severity and actual severity summarized across the 72 published studies; and the maximum severity represented in the standard area diagrams (SADs) that was divided into two classes (those small or equal to 50 or greater 50%). The SADs were developed for a specific plant disease and organ type (leaf or other)



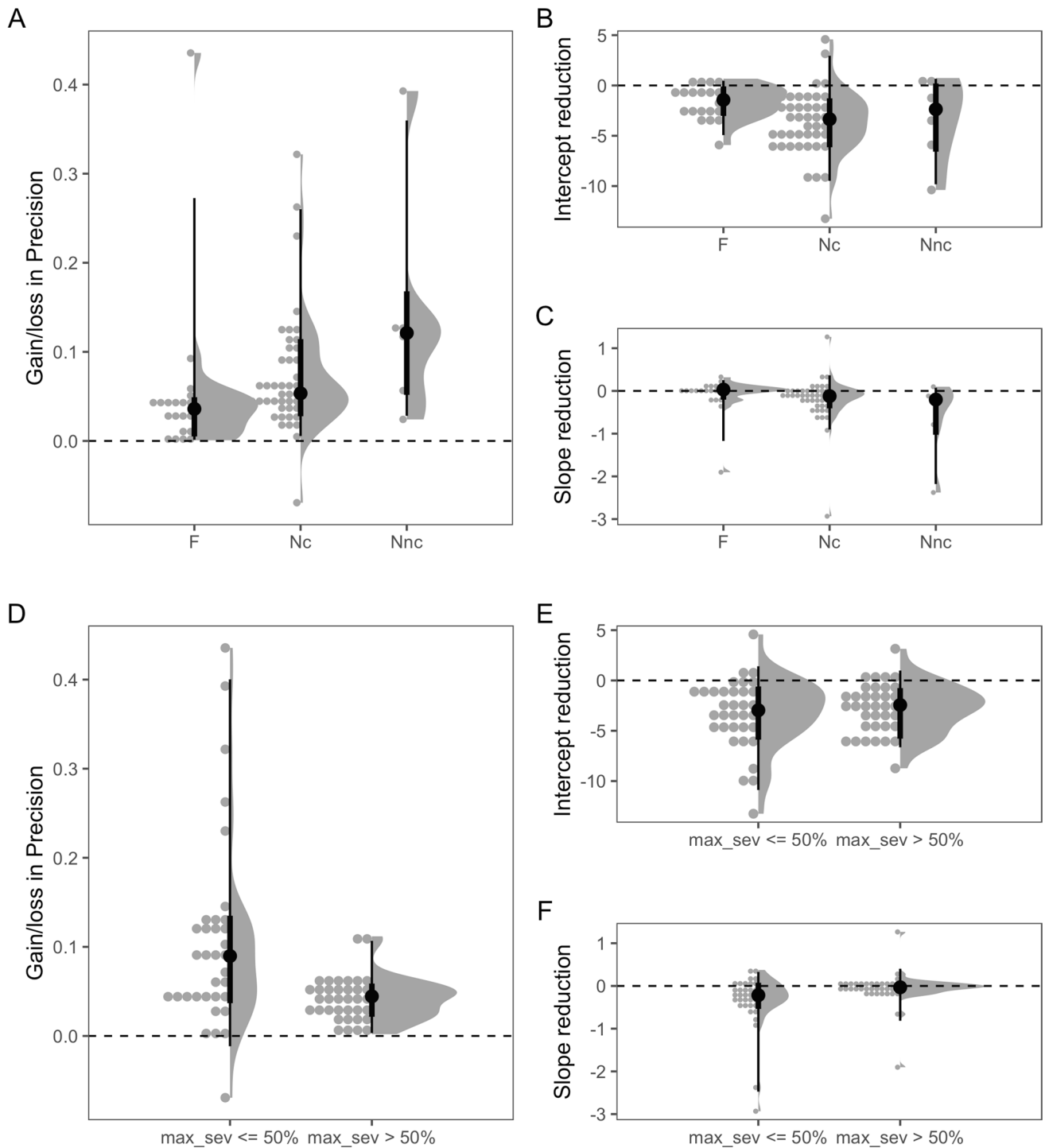


Fig. 7 Half-eye density and dots plots for the linear regression statistics indicative of bias (slope and intercept) and precision (Pearson's r) for visual estimates of percent area diseased (severity) made with the aid of standard area diagram set (SADs) and conditioned to a classification of symptoms according to size and coalescence (F=few; Nc=numerous coalescent and Nnc=numerous non coalescent) (A,

B, C) or maximum severity represented in the SAD (greater or equal to 50% or greater than 50%) (D, E, F). Each grey dot represents a published research study on SADs. The black dot in the middle represents the mean and the solid vertical lines the intervals of a frequentist distribution

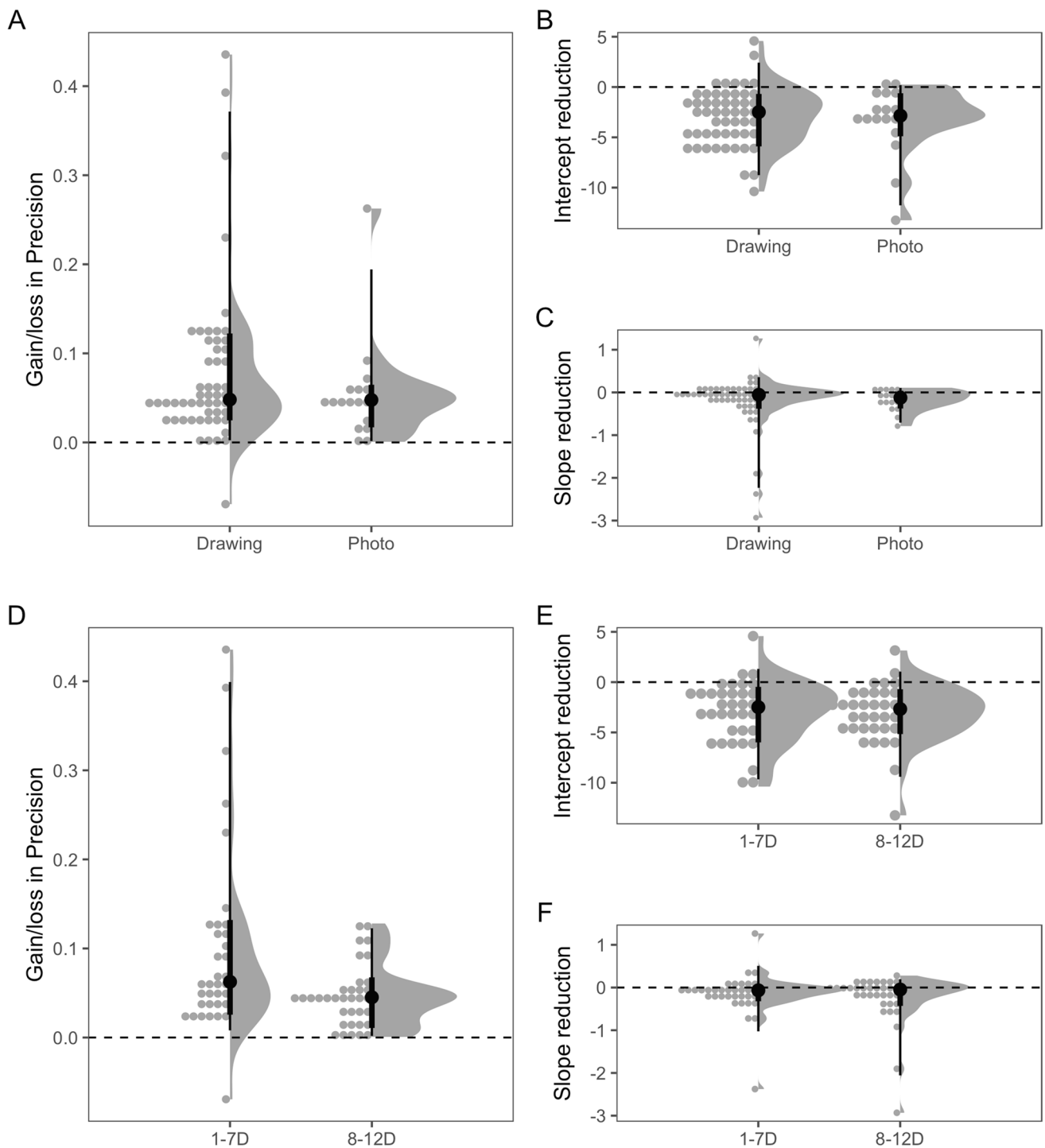


Fig. 8 Half-eye density and dot plots for linear regression statistics indicative of bias (slope and intercept) and precision (Pearson's r) for visual estimates of percent area diseased (severity) made with the aid of standard area diagram sets (SADs) and conditioned to two classes of image type (drawing or photo) (A, B, C) and two classes of num-

ber of illustrations in the SADs (one to seven diagrams or eight to 12 diagrams) (D, E, F). Each grey dot represents a published research study on SADs. The black dot in the middle represents the mean and the solid vertical lines the intervals of the distribution

linear regression (intercept, slope and the coefficient of determination) that are indicative of bias and precision. Typically, meta-analysis requires the application of criteria for

search, selection, inclusion, and further exclusion of studies based on the required criteria for the analysis. In our case, we focused on the summary statistics of the linear regression

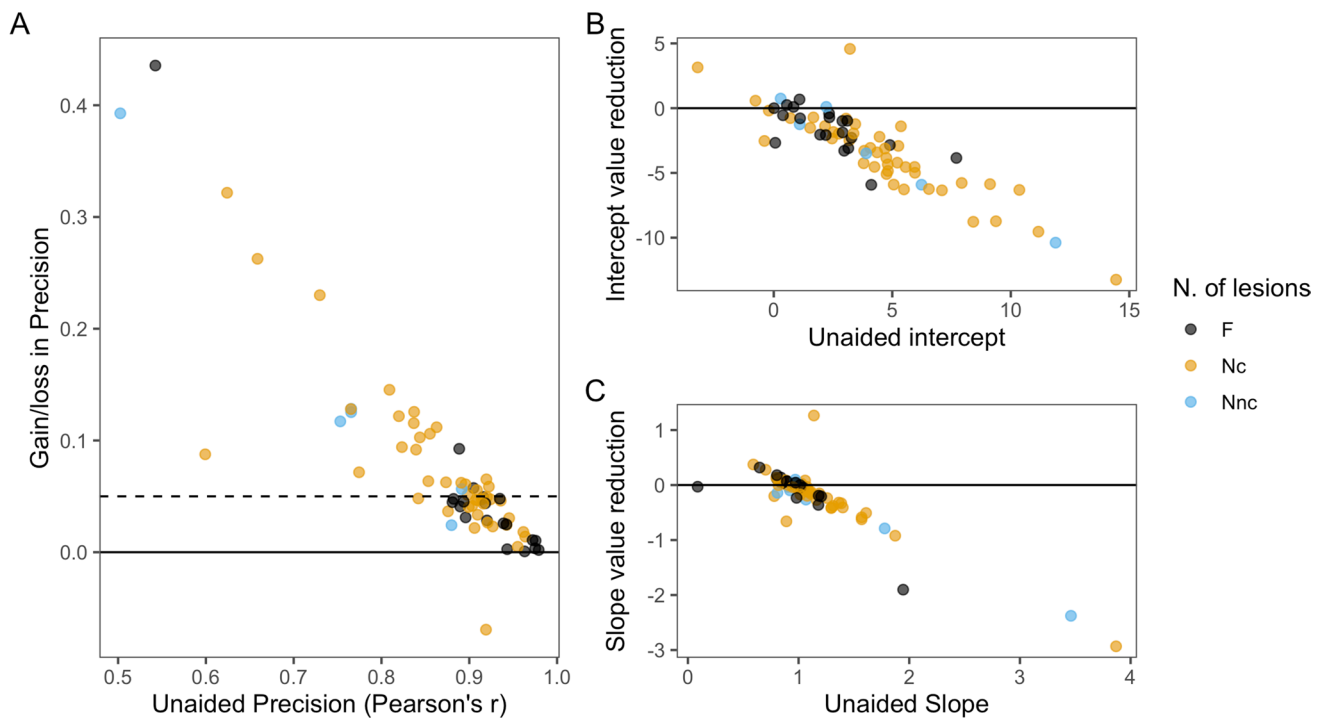


Fig. 9 The gain or loss in linear regression statistics indicative of precision (Pearson's r) (A), bias (slope and intercept) (B, C) for visual estimates of percent area diseased (severity) made first without, and then with the aid of standard area diagram set (SADs) ($n=72$ studies). The dashed line in A is indicative of the overall means of gain in

precision characteristics of lesions are indicated by color (see legend, where F=few, Nc=numerous and coalescent, and Nnc=numerous and non-coalescent). Each dot represents a published research study on SADs

coefficients and did not include studies that used LCCC (with the exception of our own data, for which linear regression statistics could be calculated). LCCC is actually considered a more suitable method for the purpose of gauging accuracy and precision (Madden et al. 2007). However, our objective was to select studies where all statistics of bias and precision were reported for individual raters (available for at least six raters), allowing us to calculate the sample variance. This is usually not available when authors use LCCC because the means across raters tends to be reported for statistical comparison between methods (Yadav et al. 2013; Rivera et al. 2020; Castellar et al. 2021). Another reason is that the values of the slopes and intercepts, which are measures of constant and systematic bias, respectively, cannot be directly compared with the location-shift and scale-shift statistics of the LCCC, respectively. Future meta-analytic studies could focus on the use of LCCC statistics to confirm our results. We note that more than 60% of the studies published following our 2017 review paper used LCCC (data not shown) to determine agreement with actual values. Further comprehensive analysis could be performed if the raw data (estimates at the rater levels, not only the summary statistics) was shared in the original SAD studies. This is one of the best operating practices when conducting SAD research (Del Ponte et al. 2017) but is still not practiced by many plant

pathologists reporting SAD research. We encourage authors of SAD studies to provide the original rater data for unaided and SAD aided estimates in support of establishing an open science culture in the field (Sparks et al. 2021) and because those data can be valuable in future analyses that further our knowledge and understanding. We support the use of Lin's CCC as a substitute for, or in addition to, linear regression.

Pearson's r , a component of LCCC that is indicative of precision, was calculated from the coefficient of determination of the linear regression. The overall increase in precision of the estimates was 0.071 (from 0.87 unaided to 0.943 when using a SAD), but considerable variability was observed across the studies. Similarly, constant bias (intercept), estimated at a value much larger than 1 for unaided estimates, was greatly reduced when using the SADs, suggesting that overestimation was reduced considerably at all severities when using the aid. The systematic bias (slope) was also reduced (closer to 1), suggesting that the overestimation proportional to the magnitude of the actual severity was greater when not using the SAD aid.

The gains when using the SADs are influenced by how well raters are able to accurately estimate disease severity unaided (a rater's baseline accuracy), a phenomenon that has been observed for individual raters (Yadav et al. 2013; Braido et al. 2014, 2015; Bock et al. 2015). However, we

found the improvement in accuracy varied according to symptom characteristics and maximum severity in the SAD set. To better understand the effect of symptoms, we developed a new key used to classify the SADs into four groups according to symptom number, size and degree of lesion coalescence. Using this approach, we found that SADs were of greatest value when the lesions are numerous, small and do not coalesce, and for those diseases (based on the maximum SAD value) that rarely exceed 50% severity. The characteristics were demonstrated to have strong association with each other based on the MCA visualization. Conversely, the smallest gains were in cases where unaided estimates were already accurate and were observed where there were few large lesions or where small lesions grew in size or coalesced, resulting in more than 50% of the specimen surface area being diseased, corroborating earlier reports of the effect of lesion size and number (Amanat 1976; Sherwood 1983; Forbes and Jeger 1987). The illusions discussed by Sherwood et al. (1983) that influenced visual judgement and resulted in greater overestimation obtained using SADs were particularly profound at low disease severities. Our results corroborate this finding; a significant departure from zero was estimated for the overall intercept, an effect that was greatest for the numerous and noncoalescent lesion category. Lesion coalescence may give the impression of larger single lesions that are easier for raters to accurately estimate disease severity. Interestingly, we found maximum severity influenced accuracy of the unaided estimates, but this was closely associated with lesion characteristics.

We were not able to detect a significant effect of the number of illustrations in a SAD set affecting the precision and bias of rater estimates, which may be related to maximum severity. We found that larger gains in precision occurred when maximum severity was less than 50%, a range where the number of diagrams tends to be fewer than with severities greater than 50%. When using maximum severity divided by the number of diagrams (data not shown), the effect was more pronounced, as this takes into account the effect of having a smaller severity range if maximum disease is low, which was the case in several of the SAD sets. Dividing the maximum severity by the number of illustrations in the SADs avoids the possibility of a false effect of fewer SADs where maximum severity is low—a disease with a low maximum severity with the same number of illustrations as a disease that attains 100% may provide more guidance for estimate interpolation. In this regard, the small effect we noted corroborates the results of three other studies where the number of illustrations in a SAD set was compared, and there was little effect of the number of illustrations in the SAD sets studied (Braido et al. 2014, 2015; Bock et al. 2015). Our results and those of the earlier studies are indicative that more research is needed to specifically investigate the effects of the number of illustrations

in a SAD. Interestingly, newer updated and refined SADs with greater numbers of illustrations have resulted in more accurate data (Franceschi et al. 2020). The Franceschi et al. (2020) result indicates a compelling reason to revisit some of the early-developed SADs and determine whether a new set should be developed.

We were unable to determine if there was an effect of image type (photo or drawing of diseased specimens) in this study. This may be due to any effect being particularly subtle. Schwanck and Del Ponte (2014) noted an effect of SADs color on both measures of bias, which were significantly improved using black and white images, although numerically the effect was slight. However, the greater accuracy achieved from new SAD sets (Franceschi et al. 2020) may in part be due to the image type used, although several other factors we have noted above could be involved too. There will inevitably be some factors that will have major effects on accuracy (including lesions numbers/size), and rater effects, but many other factors may play minor roles, although collectively we should strive to optimize these as well. Combined together these factors can offer further improvements to the accuracy of SAD-aided visual assessments. However, even with SADs, we are limited by the innate ability of raters to accurately estimate a value, and improvements in accuracy will have diminishing returns as the major sources of error are understood and resolved through optimized SAD design, appropriate training, and instruction.

In conclusion, we applied a meta-analysis to disease severity assessment data to explore how characteristics of the pathosystem and SADs affect precision and constant and systematic biases. Based on the results, we affirm the value of SADs for reducing bias and increasing precision of visual assessments of disease severity and also identify some of the characteristics of disease symptoms where the tool has greater or lesser value as an assessment aid. Specifically, we determined that the number of illustrations in SADs should take the maximum severity into account and more biased and less precise severity estimates are associated with diseases characterized by small and numerous lesions.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40858-021-00479-5>.

Author's contribution EMD conceptualized the work, performed the analysis, and wrote the manuscript; CHB conducted experiments and wrote the manuscript; LIC and KAS collected the data, conducted experiments, and revised the manuscript; and SJP conducted experiments and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding Clive H. Bock was funded by the USDA-ARS National Programs through CRIS project 6042-21220-014-00. Sarah J. Pethybridge was supported by the United States Department of Agriculture National Institute of Food and Agriculture (USDA-NIFA) Hatch project

NYG-625424. Emerson M. Del Ponte was supported by the National Council for Scientific and Technological Development (CNPq) through a Productivity Research Fellowship (PQ) project 310208/2019-0.

Data availability All data and R scripts generated for the analysis of the data and production of the figures are publicly available at: <https://osf.io/t2yfw/>.

Declarations

Conflict of interest The authors declare no competing interests.

References

- Amanat P (1976) Stimuli effecting disease assessment. *Agriculturae Conspectus Scientificus* 39:27–31
- Araújo ER, Resende RS, Krezanoski CE, Duarte HSS (2019) A standard area diagram set for severity assessment of botrytis leaf blight of onion. *European Journal of Plant Pathology* 153:273–277
- Arias SM, Álvarez GEG, Patiño PAG (2020) Diagrammatic scale for measuring severity of gray mould in thornless Castilla blackberry (*Rubus glaucus* Benth). *Ciência Rural* 50:11
- Belan LL, Belan LL, da Matta RA, Gonçalves Gomes CA, Alves FR, de Jesus C, Junior W, Moraes WB (2020) Standard area diagram with color photographs to estimate the severity of coffee leaf rust in *Coffea canephora*. *Crop Protection* 130:105077
- Bock CH, Parker PE, Cook AZ, Riley T, Gottwald TR (2009) Comparison of assessment of citrus canker foliar symptoms by experienced and inexperienced raters. *Plant Disease* 93:412–424
- Bock CH, Pethybridge SJ, Barbedo JGA, Esker PD, Mahlein AK, Del Ponte EM (2021a) A phytometry glossary for the twenty-first century: towards consistency and precision in intra- and interdisciplinary dialogues. *Trop Plant Pathol*. <https://doi.org/10.1007/s40858-021-00454-0>
- Bock CH, Chiang KS, Del Ponte EM (2021b) Plant disease severity estimated visually: a century of research, best practices, and opportunities for improving methods and practices to maximize accuracy. *Trop Plant Pathol*. <https://doi.org/10.1007/s40858-021-00439-z>
- Bock CH, Poole GH, Parker PE, Gottwald TR (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 CH, El Jarroudi M, Kouadio LA, Mackels C, Chiang K-S, Delfosse P (2015) Disease severity estimates - effects of rater accuracy and assessment methods for comparing treatments. *Plant Disease* 99:1104–1112
- Bock C, Chiang KS, Del Ponte EM (2016) Accuracy of plant specimen disease severity estimates: concepts, history, methods, ramifications and challenges for the future. *CAB Reviews* 11:1–21
- Bock CH, Barbedo JGA, Del Ponte EM, Bohnenkamp D, Mahlein A-K (2020) From visual estimates to fully automated sensor-based measurements of plant disease severity: status and challenges for improving accuracy. *Phytopathology Research* 2:9
- Braido R, Gonçalves-Zuliani AMO, Janeiro V, Carvalho SA, Junior JB, Bock CH, Nunes WMC (2014) Development and validation of standard area diagrams as assessment aids for estimating the severity of citrus canker on unripe oranges. *Plant Disease* 98:1543–1550
- Braido R, Gonçalves-Zuliani AMO, Nocchi PTR, Junior JB, Janeiro V, Bock CH, Nunes WMC (2015) A standard area diagram set to aid estimation of the severity of Asiatic citrus canker on ripe sweet orange fruit. *European Journal of Plant Pathology* 141:327–337
- Brás VV, Rios JA, Silva LC, Eduardo C, Santos M, Nascimento RSM, Rodrigues FÁ (2020) Standard area diagrams to assess scab severity of entire margined leaves of sour passion fruit. *Journal of Phytopathology* 2020:1–12
- Camara G de R, Busato LM, Almeida BF, Moraes WB (2018) Elaboration and validation of diagrammatic scale for lettuce powdery mildew. *Summa Phytopathologica* 44:116–121
- Castellar C, Jauch F, Moreira RR, da Silva Silveira Duarte H, De Mio LLM (2021) Standard area diagram set for assessment of severity and temporal progress of apple blotch. *European Journal of Plant Pathology* 160:599–609
- Chiang K-S, Liu S-C, Bock CH, Gottwald TR (2014) What interval characteristics make a good categorical disease assessment scale? *Phytopathology* 104:575–585
- Cobb N (1892) Contributions to an economic knowledge of the Australian rusts (Uredineae). *Agricultural Gazette of New South Wales* 3:44–48
- de Melo VP, Mendonça ACS, de Souza HS, Gabriel LC, Bock CH, Eaton MJ, Schwan-Estrada KRF, Nunes WMC (2020) Reproducibility of the development and validation process of standard area diagram by two laboratories: an example using the *Botrytis cinerea* / *Gerbera jamesonii* pathosystem. *Plant Disease* 104:2440–2448
- Del Ponte EM, Pethybridge SJ, Bock CH, Michereff SJ, Machado FJ, Spolti P (2017) Standard area diagrams for aiding severity Estimation: scientometrics, pathosystems, and methodological trends in the last 25 Years. *Phytopathology* 107:1161–1174
- Domiciano GP, Duarte HSS, Moreira EN, Rodrigues FA (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
- dos Santos PHD, Mussi-Dias V, Freire M Das GM, Carvalho BM, da Silveira SF (2017) Diagrammatic scale of severity for postharvest black rot (*Ceratocystis paradoxa*) in coconut palm fruits. *Summa Phytopathologica* 43:269–275
- Forbes G, Jeger M (1987) Factors affecting the estimation of disease intensity in simulated plant structures. *Journal of Plant Diseases and Protection* 94:113–120
- Franceschi VT, Alves KS, Mazaro SM, Godoy CV, Duarte HSS, Ponte EMD (2020) A new standard area diagram set for assessment of severity of soybean rust improves accuracy of estimates and optimizes resource use. *Plant Pathology* 69:495–505
- Hjellbrekke J (2018) Multiple correspondence analysis for the social sciences. Routledge
- Hock J, Kranz J, Renfro BL (1992) Tests of standard diagrams for field use in assessing the tar spot disease complex of maize (*Zea mays*). *Tropical Pest Management* 38:314–318
- James WC (1974) Assessment of plant diseases and losses. *Annual Review of Phytopathology*. 12:27–48
- Kranz J (1977) A study on maximum severity in plant disease. *Travaux dédiés à G. Viennot-Bourgin, Société Française de Phytopathologie, Paris 1977*:169–173
- Kublik G, Henkemeier NP, Lorenzetti E, Stangarlin JR, Kuhn OJ, Assi L (2020) Desenvolvimento e validação de escala diagramática para avaliação da severidade da bacteriose em mandioca. *Revista de Ciências Agroambientais* 18:75–81
- Large EC (1966) Measuring Plant Disease. *Annual Review of Phytopathology* 4:9–26
- Lin LIK (1989) A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 45:255–268
- Madden LV, Hughes G, van den Bosch F (2007) The study of plant disease epidemics. APS Press, St. Paul
- Madden LV, Piepho H-P, Paul PA (2016) Statistical models and methods for network meta-analysis. *Phytopathology* 106:792–806

- Montero YM, Rivera LR, Gutiérrez LNG (2021) Diagrammatic scales for the estimation of black node disease severity in common bean. *Ciência Rural* 51:3
- Nascimento FA, Duarte HS, Souza FF, Ishikawa FH, Capucho AS (2020) Development and validation of a standard area diagram set to assess powdery mildew severity on watermelon leaves. *Ciência Rural* 50:10
- Nita M, Ellis MA, Madden LV (2003) Reliability and accuracy of visual estimation of phomopsis leaf blight of strawberry. *Phytopathology* 93:995–1005
- Nutter FW Jr (1993) Assessing the accuracy, intra-rater repeatability, and inter-rater reliability of disease assessment systems. *Phytopathology* 83:806
- Nutter JF, Teng P, Shokes FM (1991) Disease assessment terms and concepts. *Plant Disease* 75:1187–1188
- Pereira WEL, de Andrade SMP, Del Ponte EM, Esteves MB, Canale MC, Takita MA, Coletta-Filho HD, De Souza AA (2020) Severity assessment in the *Nicotiana tabacum-Xylella fastidiosa* subsp. *pauca* pathosystem: design and interlaboratory validation of a standard area diagram set. *Tropical Plant Pathology* 45:710–722
- R Core Team (2021) R: A language and environment for statistical computing. R Foundation for statistical computing, Vienna
- Rivera JF, Duarte HSS, Furtado EB, Dallagnol LJ (2020) A standard area diagram set for severity assessment of eyespot on rice. *Australasian Plant Pathology* 49:367–371
- Robaina RR, Longhi TV, Zeffa DM, Gonçalves LSA, Leite RP (2020) Development of a protocol and a diagrammatic scale for quantification of bacterial leaf streak disease on young plants of maize. *Plant Disease* 104:2921–2927
- Schwanck AA, Del Ponte EM (2014) Accuracy and reliability of severity estimates using linear or logarithmic disease diagram sets in true colour or black and white: a study case for rice brown spot. *Journal of Phytopathology* 162:670–682
- Sherwood RT (1983) Illusions in visual assessment of stagonospora leaf spot of orchardgrass. *Phytopathology* 73:173
- Sparks AH, Del Ponte EM, Alves KS, Foster ZSL, Grünwald NJ (2021) Reproducibility in plant pathology: where do we stand and a way forward. Preprint. <https://doi.org/10.31220/agriRxiv.2021.00082>
- Spolti P, Schneider L, Sanhueza RMV, Batzer JC, Gleason ML, Del Ponte EM (2011) Improving sooty blotch and flyspeck severity estimation on apple fruit with the aid of standard area diagrams. *European Journal of Plant Pathology* 129:21–29
- Teng PS (1981) Validation of computer models of plant disease epidemics: a review of philosophy and methodology. *Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz* 88:1981
- Trojan DG, Pria MD (2018) Validação de escala diagramática para quantificação da severidade da antracnose da folha do milho. *Summa Phytopathologica* 44:56–64
- Viechtbauer W (2010) Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* 36:1–48
- Yadav NVS, de Vos SM, Bock CH, Wood BW (2013) Development and validation of standard area diagrams to aid assessment of pecan scab symptoms on fruit: standard area diagrams for pecan scab. *Plant Pathology* 62:325–335

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.