



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

POSTGRADO DE FITOSANIDAD

FITOPATOLOGÍA

**CARACTERIZACIÓN DE LOS FACTORES QUE INFLUYEN EN LA
EVALUACIÓN VISUAL DE ENFERMEDADES EN PLANTAS Y SU
COMPARACIÓN CON ESCALAS LOGARÍTMICAS**

LAURO SOTO ROJAS

T E S I S

PRESENTADA COMO REQUISITO PARCIAL

PARA OBTENER EL GRADO DE:

DOCTOR EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MÉXICO

2017

La presente tesis titulada: CARACTERIZACIÓN DE LOS FACTORES QUE INFLUYEN EN LA EVALUACIÓN VISUAL DE ENFERMEDADES EN PLANTAS Y SU COMPARACIÓN CON ESCALAS LOGARÍTMICAS, realizada por el alumno: LAURO SOTO ROJAS, bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

DOCTOR EN CIENCIAS
FITOSANIDAD
FITOPATOLOGÍA

CONSEJO PARTICULAR

CONSEJERO:



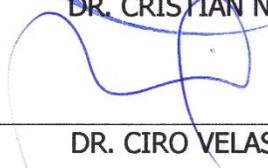
DR. DANIEL NIETO ÁNGEL

ASESOR:



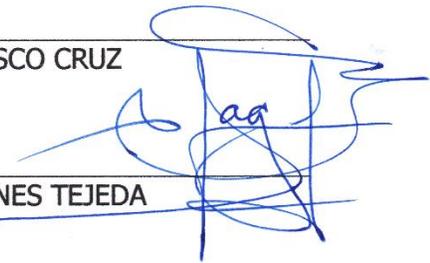
DR. CRISTIAN NAVA DÍAZ

ASESOR:



DR. CIRO VELASCO CRUZ

ASESOR:



DR. ÁNGEL LAGUNES TEJEDA

ASESOR:



DR. ARMANDO LOZANO RODRÍGUEZ

Montecillo, Texcoco, Estado de México, junio de 2017

CARACTERIZACIÓN DE LOS FACTORES QUE INFLUYEN EN LA EVALUACIÓN VISUAL DE ENFERMEDADES EN PLANTAS Y SU COMPARACIÓN CON ESCALAS LOGARÍTMICAS

Lauro Soto Rojas, Dr.

Colegio de Postgraduados, 2017

RESUMEN

La evaluación visual es el método más utilizado para estimar severidad; sus resultados varían en función de las diferencias innatas de los evaluadores, incluyendo su habilidad cognitiva. En este estudio se analizó el efecto del estilo cognitivo reflexividad-impulsividad, la experiencia del evaluador y el uso de diagramas de área estándar (SAD) en la concordancia y confiabilidad de estimaciones visuales. El material experimental estuvo constituido por hojas estilizadas de cinco tipos (lineal, elíptica, lobulada, triangular y palmeada), con severidad en el intervalo 0 a 70 %, y con lesiones en patrones agregados (PI-AG) y aleatorios (PI-AL). Se conformó un grupo de 26 evaluadores (13 reflexivos y 13 impulsivos, clasificados con el test MFFT); de este grupo, 14 tienen experiencia en evaluación visual y el resto recibió entrenamiento. Los componentes de la concordancia se estimaron con el índice de Lin (LCC, ρ_c); la confiabilidad se obtuvo con el coeficiente de correlación intra-clase (ρ) y para las medias de los parámetros se establecieron los intervalos de confianza de 95 % (CIs) con análisis bootstrap (2 000 repeticiones). Los evaluadores reflexivos obtuvieron una concordancia y confiabilidad significativamente mayor que los evaluadores impulsivos ($P < 0.0001$). Los resultados del estudio muestran que el estilo cognitivo es tan importante como la experiencia del evaluador; sin embargo, en relación con el tiempo de evaluación, se evidenció que los evaluadores reflexivos son más eficientes que los evaluadores experimentados. Los SADs que producen resultados más confiables y con mayor concordancia son derivados de escalas de proporción con tres a siete diagramas, dispuestos a intervalos lineales. La evidencia científica indica que la forma de la hoja y el patrón de las lesiones deben considerarse como fuentes de error adicionales a las ya conocidas.

Palabras clave: Evaluación visual, concordancia, reflexivo-impulsivo, estilo cognitivo, patrones de infección, SAD

CHARACTERIZATION OF FACTORS INFLUENCING VISUAL ASSESSMENT OF PLANT DISEASE SEVERITY AND COMPARISON WITH LOGARITHMIC SCALES

Lauro Soto Rojas, Dr.
Colegio de Postgraduados, 2017

ABSTRACT

In applied research, the most used method for estimating plant disease severity is visual assessment, whose results varies by the innate differences between individuals, such as cognitive ability. In this study, the effects of reflectivity-impulsivity cognitive style and standard area diagrams on accuracy, precision, concordance and reliability of visual estimations of severity in different leaf types (linear, elliptical, lobed, triangular and palmate) were analyzed. Additionally, the effect of the aggregate infection patterns (PI-AG) and random infection patterns (PI-AL) on leaves was studied. The visual estimation was performed by 26 raters, of whom 13 were reflective and 13 impulsive. In addition, the raters were classified for their experience in assessing severity. The visual estimations were compared with actual values obtained with image analysis. The agreement was calculated using Lin's concordance correlation (LCC, ρ_c). For all types of leaves and infection patterns, the reflective raters were more precise and accurate, with a significantly higher concordance than the impulsive raters ($P < 0.0001$). The results show that cognitive style is as important as the rater's experience. However, the analysis of the evaluation time and the agreement suggests that reflective raters are more efficient than experienced raters. Moreover, the leaf shape and the infection patterns are factors that determine precision and accuracy of visual estimations, and therefore, they should be considered as sources of error additional to those already known.

Keywords: visual assessment, agreement, reflective-impulsive, cognitive style, infection pattern, SAD

AGRADECIMIENTOS

A **DIOS**, por la porción de fe que pusiste en mi alma.

¡En tus propósitos eternos... me vi, entre tus planes para hoy... me encontré!

Al **Consejo Nacional de Ciencia y Tecnología**, por el apoyo que otorga a todos los estudiantes del país.

Al **Colegio de Postgraduados**, por darme la oportunidad de recibir una educación de excelencia. Gracias por prepararme para servir mejor a la sociedad.

A mi consejero **Dr. Daniel Nieto Angel**, por su invaluable amistad. Gracias por su gran calidad profesional y humana. Gracias por la confianza y el apoyo incondicional que siempre me ha brindado.

Al **Dr. Cristian Nava Díaz**, por el soporte científico, por los comentarios que me motivaron a continuar con esta investigación. Gracias por su gentileza y su aportación a mi formación como investigador.

Al **Dr. Ciro Velasco Cruz**, por su acertada visión científica y crítica. Gracias por su paciencia y por compartir conmigo sus conocimientos sobre Estadística. Gracias por orientarme más allá de los límites de esta investigación.

Al **Dr. Ángel Lagunes Tejeda**, por compartirme sus vastos conocimientos y experiencias. Gracias por sus valiosas observaciones y recomendaciones que enriquecieron la presente investigación.

Al **Dr. Armando Lozano Rodríguez**, por incorporarse con entusiasmo a nuestro equipo de trabajo. Gracias por su calidad profesional, su disposición y amabilidad.

Al **Dr. Néstor Bautista Martínez**, por su amistad y su confianza. Muchas gracias por apoyarme en muchos aspectos de mi vida.

A todos los colaboradores que, con muy buena disposición, me ayudaron en la evaluación visual: AAS, ABC, AMS, APH, CDL, CFS, CJM, CND, COA, CVG, DGP, DTO, ERR, ESR, EZM, FDI, FFN, GEV, GLM, GVP, HRG, IJR, IRG, ITS, JRR, JTS, KDA, LAP, LPC, LSR, LVM, MDS, MOV, MPE, MSR, MTR, NLC, NLV, OMG, PIR, RLG, RSM, RVA, SAO, SAR, SMA, SOL, VAE y VHA.

DEDICATORIA

A mi esposa, Elisa Ríos Ramírez. Eres mi compañía perfecta, la ayuda idónea. Gracias por construir nuestra felicidad. Te amo.

A mis hijos, Flor Dalí y Enoc Maximiliano. En humildad, espero ser un buen ejemplo para ustedes, y que al ojear este documento encuentren alguna inspiración que los ayude a edificar sus sueños... ¡Los amo! <Efesios 6:10-18>

A mi madre, Heriberta Rojas García. Mamita siempre has sido una bendición para mí. Te amo y estaré contigo todos los días de mi vida.

A mi padre Laurentino Soto Reyes. Por darme la extraordinaria oportunidad de vivir.

A mis hermanos: Pedro, Pepe, Luis, Flor, Laura y Adrián, por estar conmigo a cada instante. Los amo, que Dios los bendiga!

A mis sobrinos: Ángel, Ángelo, Bertha, Chaco, Cristo, Irving, Jonathan, Leslie, Lupita, Michel, Sergio y Sheyla. Porque son la alegría de nuestras vidas. Siempre estaré para ser su amigo.

A la familia Ríos Ramírez. Por adoptarme y brindarme su confianza.

A las personas que siempre me han acompañado, mis amigos y amigas:

Daniel García	Sol Millán	Berestravn Cruz	Alfredo Baltazar
Beto Ramírez	Martha Segundo	Mariana S. Roldán	Isabel Ruíz
Cecilia Velázquez	Karen Díaz	Chalo Espinoza	Luis A. Aguilar
Santo Ángel	Adrián Negrete	Edgar García	<i>El Toñazo</i>
Ibeth Jaimes	Rosy Quesada	Gris Pérez	Ever López

Gracias por su apoyo durante mi estancia en el Colegio de Postgraduados

*“Amo el canto del cenxontle, pájaro de cuatrocientas voces,
Amo el color del jade, y el enervante perfume de las flores;
Pero amo más a mi hermano el hombre”*

-Nezahualcóyotl-



CONTENIDO

CAPITULO 1. EFECTOS DEL ESTILO COGNITIVO REFLEXIVIDAD-IMPULSIVIDAD EN LA ESTIMACIÓN VISUAL DE SEVERIDAD.....	1
1.1. RESUMEN	1
1.2. INTRODUCCIÓN	3
1.3. MATERIALES Y MÉTODOS.....	5
1.3.1. Obtención de las hojas	5
1.3.2. Evaluación del estilo cognitivo impulsividad-reflexividad.....	9
1.3.3. Evaluación visual de la severidad	11
1.3.4. Análisis de datos.	12
1.4. RESULTADOS.....	16
1.4.1. Efecto del estilo cognitivo.....	16
1.4.2. Efecto del patrón de las lesiones.....	20
1.4.3. Efecto de la experiencia del evaluador.....	20
1.4.4. Efecto del tipo de hoja.....	23
1.4.5. Confiabilidad inter-evaluadores.....	27
1.4.6. Confiabilidad intra-evaluadores.....	27
1.5. DISCUSIÓN	30
1.6. LITERATURA CITADA	41

CAPITULO 2. EVALUACIÓN VISUAL DE SEVERIDAD, UN ANÁLISIS COMPARATIVO DE ESCALAS DE PROPORCIÓN Y CATEGÓRICAS CON INTERVALOS LOGARÍTMICOS Y LINEALES	45
2.1. RESUMEN	45
2.2. INTRODUCCIÓN	46
2.3. MATERIALES Y MÉTODOS.....	54
2.3.1. Obtención de las hojas.....	54
2.3.2. Desarrollo de las escalas	57
2.3.3. Estimación visual de la severidad	59
2.3.4. Análisis de datos.	62
2.4. RESULTADOS.....	66
2.4.1. Efecto del tipo de intervalo (logarítmico-lineal) para escalas categorías.....	66
2.4.2. Efecto del tipo de intervalo (logarítmico-lineal) para escalas de proporción.....	68
2.4.3. Efecto del número de diagramas	70
2.4.4. Comparación entre escalas categorías y de proporción.	70
2.5. DISCUSIÓN	74
2.6. LITERATURA CITADA	84

LISTA DE CUADROS

Cuadro 1.1. Relación de imágenes generadas a partir de las hojas patrón (PL) y los patrones de infección (PI).....	8
Cuadro 1.2. Clasificación de los evaluadores por estilo cognitivo y nivel de experiencia.....	10
Cuadro 1.3. Efecto del estilo cognitivo en el sesgo, precisión, concordancia y tiempo de evaluación en distintos tipos de hoja y patrones de infección	18
Cuadro 1.4. Efecto del patrón de las lesiones en el sesgo, precisión, concordancia y tiempo de evaluación en distintos tipos de hoja.....	19
Cuadro 1.5. Efecto de la experiencia del evaluador en el sesgo, precisión, concordancia y tiempo de evaluación en distintos tipos de hoja y patrones de infección.....	22
Cuadro 1.6. Confiabilidad de la estimación visual de la severidad en relación con el estilo cognitivo reflexividad-impulsividad.....	28
Cuadro 2.1. Escala logarítmica de Horsfall - Barrat para evaluación de severidad.....	62
Cuadro 2.2. Efecto del tipo de intervalo y el número de diagramas del SAD, en el sesgo, precisión y concordancia en escalas categóricas. Modo de uso: CLAS.....	67
Cuadro 2.3. Efecto del tipo de intervalo y el número de diagramas del SAD, en el sesgo, precisión y concordancia en escalas de proporción. Modo de uso: AV.....	69
Cuadro 2.4. Coeficiente de Correlación de Concordancia de Lin (ρ_c). Comparativo entre número de diagramas, intervalos de incremento y modos de uso.....	71

LISTA DE FIGURAS

Figura 1.1. Tipos de hoja y patrones de infección.....	7
Figura 1.2. Relación entre la severidad real y la severidad estimada para cada tipo de hoja.	25
Figura 1.3. Comparación de las mediciones de sesgo, precisión, concordancia y tiempo para cinco tipos de hoja (lineal, elíptica, triangular, lobulada y palmeada) y dos patrones de infección (agregado y aleatorio).....	26
Figura 1.4. Confiabilidad inter evaluadores medida a través de la frecuencia de los coeficientes de determinación (R^2) por pares. Comparación por estilo cognitivo, tipo de hoja y patrón de las lesiones.....	29
Figura 1.5. Gráficos Bland-Altman. Error absoluto vs severidad real.....	34
Figura 1.6. gráficos Bland-Altman. Error absoluto vs severidad real.....	37
Figura 2.1. Ejemplo de escala de proporción con SAD.	50
Figura 2.2. Hojas estilizadas a partir de una hoja patrón (PL) y patrones de infección (PI) obtenidos de cultivos infectados naturalmente.....	56
Figura 2.3. Determinación de los intervalos logarítmicos de incremento de severidad.....	58
Figura 2.4. Escala categórica con SAD de tres clases a intervalos logarítmicos.....	59
Figura 2.5. Escala categórica con SAD de tres clases a intervalos lineales..	60
Figura 2.6. Escala de proporción con SAD de cinco clases a intervalos logarítmicos.....	61
Figura 2.7. Escala de proporción con SAD de siete clases a intervalos lineales.....	63

Figura 2.8. Coeficiente de Correlación de Concordancia de Lin. Comparativo entre escalas categóricas y de proporción.....	72
Figura 2.9. Coeficiente de Correlación Intra-clase, comparativo entre escalas logarítmicas y lineales.....	73
Figura 2.10. Análisis de concordancia. Comparativo entre Evaluación Visual sin escala (EVD) y la transformación a puntos medios de la escala de Horsfall-Barrat.....	75
Figura 2.11. Análisis de concordancia para escalas de 3 diagramas. Comparativo del modo de uso entre escalas logarítmicas y lineales.....	76
Figura 2.12. Análisis de concordancia para escalas de 4 diagramas. Comparativo del modo de uso entre escalas logarítmicas y lineales.....	78
Figura 2.13. Análisis de concordancia para escalas de 5 diagramas. Comparativo del modo de uso entre escalas logarítmicas y lineales.....	79
Figura 2.14. Análisis de concordancia para escalas de 6 diagramas. Comparativo del modo de uso entre escalas logarítmicas y lineales.....	80
Figura 2.15. Análisis de concordancia para escalas de 7 diagramas. Comparativo del modo de uso entre escalas logarítmicas y lineales.....	82

Al principio vienen necesariamente a la mente la fantasía y la fábula.
Desfilan después los cálculos matemáticos, y sólo al final la realización corona el
pensamiento.

Konstantín E. Tsiolkovski

CAPITULO 1

EFFECTOS DEL ESTILO COGNITIVO REFLEXIVIDAD-IMPULSIVIDAD EN LA ESTIMACIÓN VISUAL DE SEVERIDAD

1.1. RESUMEN

En investigación aplicada, el método más utilizado para estimar la intensidad de enfermedades en plantas es la evaluación visual. Este método puede ser impreciso e inexacto y los resultados varían en relación con las diferencias innatas de los individuos, como la percepción visual y la habilidad cognitiva. En este estudio, se analizó el efecto del estilo cognitivo reflexividad-impulsividad en la exactitud, precisión, concordancia y confiabilidad de las estimaciones visuales de severidad en cinco tipos de hoja (lineal, elíptica, lobulada, triangular y palmeada); con el mismo objetivo, se analizó el efecto de la distribución de las lesiones en la hoja: patrón agregado (PI-AG) y aleatorio (PI-AL). En primer lugar, se escanearon hojas de diversos cultivos infectados de manera natural; posteriormente, con base en las hojas escaneadas, se usó el procesador de imágenes GIMP para generar 120 imágenes de hojas con severidad entre 0 y 70%. A los evaluadores se les aplicó el Matching Familiar Figures Test (MFFT) para clasificarlos por su estilo cognitivo, se tomó un grupo balanceado con 13 evaluadores reflexivos y 13 impulsivos; del grupo de evaluadores, 14 tienen experiencia en evaluación visual de severidad y el resto recibió entrenamiento. Las estimaciones visuales se compararon con los valores reales obtenidos con análisis de imágenes; el nivel de concordancia entre los dos enfoques se calculó con la correlación de concordancia de Lin (LCC, ρ_c).

Se obtuvo el coeficiente de determinación (R^2) como indicador de la confiabilidad; adicionalmente, la confiabilidad inter-evaluadores se midió a través del coeficiente de correlación intra-clase (ρ) obtenido a través de un ANOVA de dos vías con efectos aleatorios. Con remuestreo bootstrap se calcularon los intervalos de confianza de 95% (CIs) para la diferencia entre reflexivos - impulsivos y experimentados - inexperimentados.

Para todos los tipos de hoja y patrones de infección, los evaluadores reflexivos efectuaron estimaciones más precisas, con una concordancia significativamente más alta que la de los evaluadores impulsivos ($P < 0.0001$). La concordancia ρ_c tomó valores más cercanos a 1 cuando las imágenes tuvieron PI-AG en hojas de forma simple (lineal, elíptica y triangular); en contraste, las imágenes con PI-AL registraron niveles de concordancia más altos en hojas de forma compleja (lobulada y palmeada). La confiabilidad inter-evaluadores registró los valores más elevados con los evaluadores reflexivos ($0.80 < \rho < 0.89$), en comparación con los evaluadores impulsivos ($0.70 < \rho < 0.84$); sin embargo, Los evaluadores impulsivos mostraron mayor confiabilidad intra-evaluador. Los resultados del estudio muestran que el estilo cognitivo de los evaluadores es tan importante como su experiencia; sin embargo, al analizar la relación del tiempo de evaluación con la concordancia, los evaluadores reflexivos son más eficientes que los evaluadores experimentados. Además, la forma de la hoja y el patrón de las lesiones son factores determinantes de la precisión y la exactitud de las evaluaciones visuales, por lo que deben considerarse como fuentes de error adicionales a las ya conocidas.

1.2. INTRODUCCIÓN

La evaluación visual es el método más utilizado para estimar la intensidad de enfermedades en plantas. La variable con mayor aplicación es el porcentaje de severidad, definida como el área relativa que muestra síntomas de una enfermedad en la superficie de un órgano vegetal (hoja, fruto, tallo, etc.). La estimación de la severidad tiene muchas aplicaciones y propósitos en la investigación fitosanitaria; por ejemplo, para la estimación de pérdidas en el rendimiento en relación con la presencia de alguna enfermedad, la evaluación de la resistencia de distintos cultivares, la comparación de métodos de control, y en estudios epidemiológicos (Yadav *et al.*, 2013).

La estimación visual de la severidad no produce siempre los mismos resultados ya que depende de particularidades del individuo y del patosistema. Uno de los factores que se considera determinante en la calidad de la estimación visual es la experiencia del evaluador (Bock *et al.*, 2009; Schwanck y Del Ponte 2014; Costa Lage *et al.*, 2015). Además, Bock *et al.* (2010) sugieren que la calidad de las estimaciones varía en relación con las diferencias innatas de los individuos, desde la percepción visual hasta la habilidad cognitiva.

La cognición es el conjunto de subprocesos mentales a través de los cuales el individuo adquiere, procesa, acumula y utiliza la información. Uno de los enfoques más aplicados en la educación y en las ciencias médicas es el estilo cognitivo que dimensiona la reflexividad y la impulsividad (Buela y Riaño, 2015). El comportamiento impulsivo se define como la predisposición a reaccionar rápidamente, con una escasa o nula

planeación; en contraste, el comportamiento reflexivo se caracteriza por una predisposición al procesamiento metódico antes de reaccionar o emitir un juicio (Kertzman *et al.*, 2009). El estilo cognitivo conceptualiza a la reflexividad como un polo opuesto a la impulsividad; además, se sabe que existen infinidad de puntos intermedios entre ambos polos. Para fines prácticos, los individuos pueden clasificarse como impulsivos o reflexivos, dependiendo de su estilo predominante. El enfoque de estilos cognitivos se implementa en este estudio para analizar su relación con la exactitud, precisión, concordancia y confiabilidad de las estimaciones visuales.

Hay pocos estudios que analicen la calidad de la evaluación visual en función del tiempo invertido. Parker *et al.* (1995) indican que la precisión se reduce cuando la evaluación se realiza rápidamente. El estilo cognitivo reflexividad-impulsividad supone que cuando una tarea es efectuada rápidamente se tiene mayor probabilidad de cometer errores; en contraste, una mayor planeación y enfoque en el desempeño de la tarea conlleva a un mayor consumo de tiempo y menor probabilidad de error. Este fenómeno es conocido como compensación de velocidad y exactitud (*Speed-Accuracy Tradeoff*) (Kertzman *et al.*, 2009); en este trabajo se evalúa un enfoque similar, se analiza la relación entre el tiempo y la calidad de las estimaciones. Los efectos del estilo cognitivo y la velocidad de reacción no han sido estudiados en la evaluación de severidad, por lo que se plantearon los siguientes objetivos:

- i) Comparar la exactitud, precisión, concordancia y confiabilidad de las estimaciones de severidad efectuadas por evaluadores impulsivos y reflexivos.
- ii) Analizar el efecto del estilo cognitivo reflexividad-impulsividad en las estimaciones de severidad efectuadas sobre cinco tipos de hoja y dos patrones de infección.
- iii) Establecer la relación funcional entre el tiempo de evaluación y la calidad de las estimaciones.

1.3. MATERIALES Y MÉTODOS

1.3.1. Obtención de las hojas

Las hojas patrón (PL) se obtuvieron de cultivos con cinco tipos de hoja: lineal (trigo), elíptica (aguacate), lobulada (crisantemo), triangular (pepino) y palmeada (papaya). Para cada cultivo se digitalizaron 20 hojas, a las que se les midió el largo y ancho con una regla de precisión; con estos datos se construyeron histogramas y se seleccionó como patrón a la hoja que tenía las dimensiones más similares a las medias de la muestra. Se obtuvo una PL para cada cultivo. Las PL se vectorizaron usando el GNU Image Manipulation Program (GIMP).

Los patrones de infección (PI) se obtuvieron de hojas de distintos cultivos infectados naturalmente, con lo que se representó la diversidad de patrones de lesiones e intensidades que ocurren en el campo. Las hojas infectadas se digitalizaron y posteriormente se procesaron con GIMP; en cada imagen se eliminó el tejido sano y el

fondo, conservando únicamente las lesiones y su distribución, las imágenes resultantes se denominaron PI. A partir de cada PI se crearon pinceles en GIMP (archivos GBR), los cuales se usan de manera similar a un sello. Adicionalmente, cada PI se analizó con el paquete EImage en R, en este programa se creó una rutina para obtener el índice de agregación de Morisita (Morisita, 1959), a través del conteo y distribución de píxeles. En total se obtuvieron 320 PI, de los cuales 170 fueron aleatorios (PI-AL) y 150 agregados (PI-AG).

A partir de las PL y los pinceles de PI, se generaron 120 imágenes de hojas con severidad entre 0 y 70%; los niveles de severidad que se representaron en las imágenes se definieron generando tres números aleatorios dentro de cada intervalo: 0-10, 10-20, ..., 60-70. En el intervalo de 0-10 se obtuvieron los números aleatorios: 1.65, 4.38 y 7.58, cada uno de estos números se representó con los dos patrones (PI-AG y PI-AL), mientras que los números aleatorios mayores a 10 se representaron alternando los patrones (Cuadro 1.1). Para generar el valor exacto del porcentaje de severidad requerido, en cada imagen se probaron distintos pinceles de PI hasta encontrar el que produjera el valor de severidad más aproximado, posteriormente se ajustó aumentando o reduciendo el tamaño de una lesión. El porcentaje de severidad real se calculó usando GIMP, mediante el conteo de píxeles de tejido sano y enfermo. De las 120 imágenes generadas, se obtuvieron 24 para cada tipo de hoja; la mitad de las imágenes tiene PI-AL y la otra mitad tiene PI-AG. En la Figura 1.1 se muestra un ejemplo de las imágenes generadas.

Para entrenamiento de los evaluadores sin experiencia, se generaron 30 imágenes adicionales, con el mismo procedimiento que se ha descrito antes.

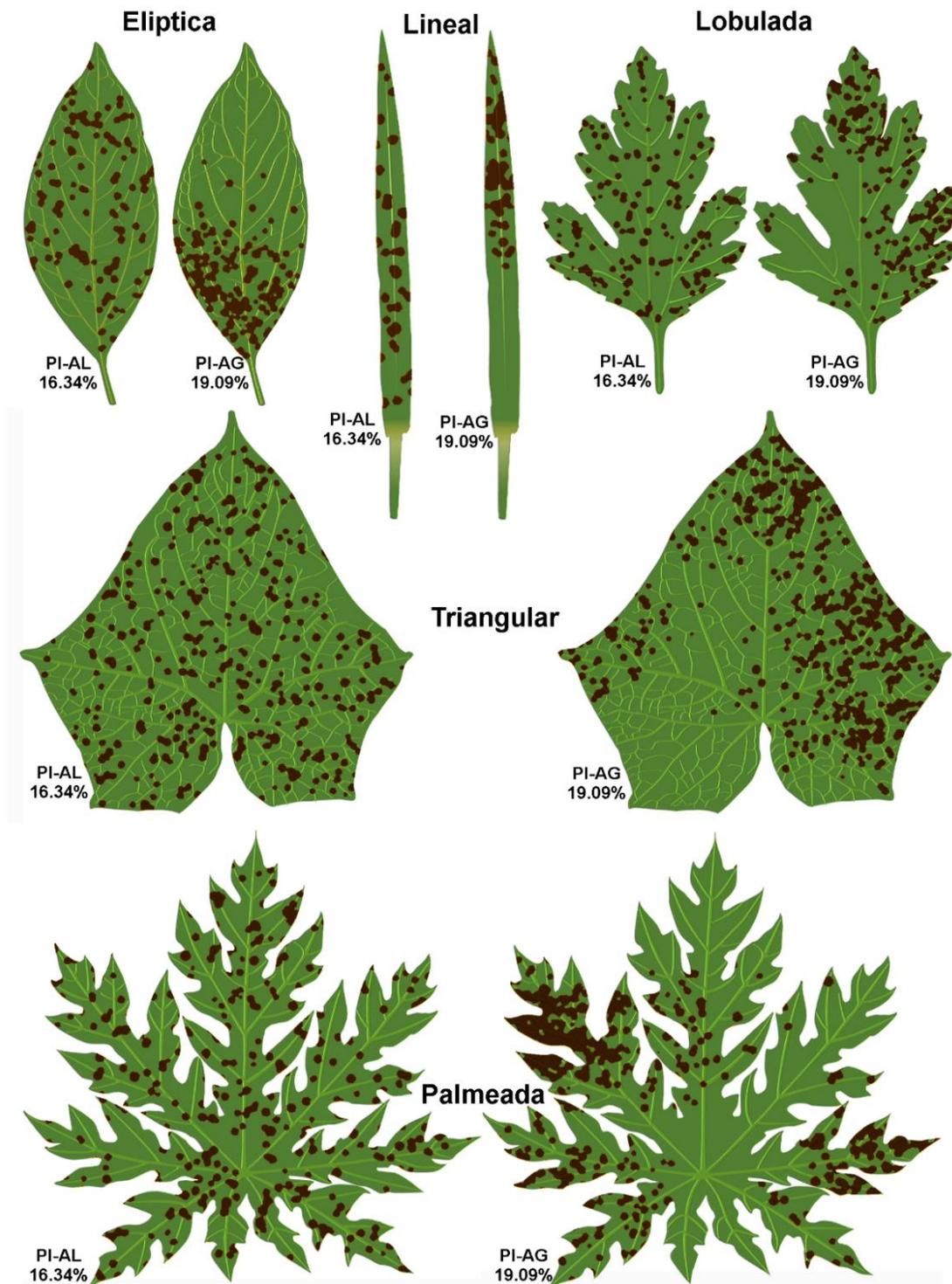


Figura 1.1. Tipos de hoja y patrones de infección. En este estudio se eligieron hojas de distintos cultivos con el objetivo de incluir diversidad de formas y tamaños. El estudio incluye el análisis de los patrones de infección agregado (PI-AG) y aleatorio (PI-AL), los cuales son muy comunes en la naturaleza.

Cuadro 1.1. Relación de imágenes generadas a partir de las hojas patrón (PL) y los patrones de infección (PI). La severidad representada en las imágenes se determinó a través de números aleatorios dentro del rango 0-70%

Número de imagen ^z	Severidad Real	Patrón de infección ^y
1	1.65	PI-AG
2	1.65	PI-AL
3	4.38	PI-AG
4	4.38	PI-AL
5	7.58	PI-AG
6	7.58	PI-AL
7	12.71	PI-AG
8	16.34	PI-AL
9	19.09	PI-AG
10	23.28	PI-AL
11	25.44	PI-AG
12	29.00	PI-AL
13	33.50	PI-AG
14	35.85	PI-AL
15	37.22	PI-AG
16	40.14	PI-AL
17	43.95	PI-AG
18	48.19	PI-AL
19	52.40	PI-AG
20	55.64	PI-AL
21	59.72	PI-AG
22	62.90	PI-AL
23	65.95	PI-AG
24	68.27	PI-AL

^z: Para cada uno de los cinco tipos de hoja se generaron 24 imágenes con la severidad y los patrones de infección indicados en este cuadro. En total se obtuvieron 120 imágenes para evaluación y 30 imágenes para entrenamiento.

^y: Patrones de infección agregado (PI-AG) y aleatorio (PI-AL)

1.3.2. Evaluación del estilo cognitivo impulsividad-reflexividad

Para la caracterización del estilo cognitivo de los participantes se aplicó la versión computarizada del Matching Familiar Figures Test MFFT (AnimaScan Ltd., Ashdod, Israel, 2000). Este test está constituido por dos ejercicios de entrenamiento y 12 ítems de medición. En cada ítem se presenta una figura modelo del lado izquierdo de la pantalla, y del lado derecho se muestra un arreglo de seis figuras similares, de las cuales únicamente una es igual a la figura modelo; la tarea consiste en que el observador debe seleccionar la figura del arreglo que es igual a la del modelo. Si la selección es correcta, inmediatamente se despliega el siguiente ítem; de otra manera, el sistema solicita al observador que intente otra vez. En cada ítem se puede registrar un máximo de 6 intentos. Para todos los individuos se aplicó el test en dos ocasiones (test-retest), con un total de 24 ítems distintos. El sistema registró la latencia (tiempo que tarda el observador en dar la primera respuesta) y el número de errores cometidos. El test se aplicó usando una computadora personal HP con un monitor de 23 pulgadas (frame rate 100 Hz) con sistema operativo Windows 10. La distancia entre el observador y la pantalla fue de 60 cm. Con los datos obtenidos se calculó el *impulsivity score* (*I-score*), el cual ha sido utilizado en diversos estudios relacionados con el estilo cognitivo impulsividad-reflexividad (Kertzman *et al.*, 2009; Quednow *et al.*, 2007). El *I-score* se calculó para cada individuo, a través de la resta de los valores estandarizados de la latencia media a los valores estandarizados del número total de errores: $I\text{-score} = Z_{\text{error}} - Z_{\text{latencia}}$. Los valores positivos de *I-score* indican que el individuo es predominantemente impulsivo, mientras que los valores negativos sugieren que el individuo es predominantemente reflexivo. El MFFT se aplicó a 28 participantes, de los cuales 15 resultaron reflexivos y

13 impulsivos; para balancear el experimento se descartaron los dos participantes reflexivos que obtuvieron el *I*-score más cercano a cero. De esta manera se conformó un grupo de 26 participantes para la evaluación visual de la severidad, se incluyeron 13 por cada estilo cognitivo (Cuadro 1.2).

Cuadro 1.2. Clasificación de los evaluadores por estilo cognitivo y nivel de experiencia

EVALUADOR	<i>I</i>-score	ESTILO COGNITIVO	EXPERIENCIA
1	-2.452	Reflexivo (R1)	Experimentado
2	-0.175	Reflexivo (R2)	Experimentado
3	-0.547	Reflexivo (R3)	Experimentado
4	-0.565	Reflexivo (R4)	Experimentado
5	-1.278	Reflexivo (R5)	Inexperimentado
6	-0.296	Reflexivo (R6)	Inexperimentado
7	-1.062	Reflexivo (R7)	Inexperimentado
8	-0.747	Reflexivo (R8)	Experimentado
9	-0.309	Reflexivo (R9)	Experimentado
10	-4.258	Reflexivo (R10)	Inexperimentado
11	-0.091	Reflexivo (R11)	Experimentado
12	-1.504	Reflexivo (R12)	Experimentado
13	-2.413	Reflexivo (R13)	Inexperimentado
14	1.387	Impulsivo (I1)	Inexperimentado
15	0.581	Impulsivo (I2)	Experimentado
16	3.557	Impulsivo (I3)	Inexperimentado
17	1.258	Impulsivo (I4)	Experimentado
18	1.537	Impulsivo (I5)	Inexperimentado
19	0.097	Impulsivo (I6)	Experimentado
20	0.422	Impulsivo (I7)	Inexperimentado
21	3.524	Impulsivo (I8)	Inexperimentado
22	0.830	Impulsivo (I9)	Experimentado
23	0.100	Impulsivo (I10)	Experimentado
24	1.803	Impulsivo (I11)	Inexperimentado
25	0.358	Impulsivo (I12)	Inexperimentado
26	0.117	Impulsivo (I13)	Experimentado

El estudio incluyó inicialmente a 28 participantes; sin embargo, únicamente se consideraron 13 individuos por cada estilo cognitivo. Los 26 participantes concluyeron todas las etapas del estudio.

1.3.3. Evaluación visual de la severidad

Para la evaluación visual, las imágenes se ordenaron aleatoriamente en una presentación de 120 diapositivas y se desplegaron en un monitor de 23 pulgadas, procurando conservar las proporciones reales de las hojas; la transición entre diapositivas estuvo acompañada por una notificación sonora (*beep*). La estimación visual se realizó por 26 evaluadores, los cuales tienen formación agronómica y están familiarizados con síntomas de enfermedades en plantas. Del grupo de evaluadores, 14 tienen experiencia en evaluación visual de severidad, el resto recibió entrenamiento usando el conjunto de 30 imágenes generadas de la forma descrita anteriormente. El entrenamiento se efectuó en dos ocasiones, una semana antes de la evaluación definitiva.

La evaluación de severidad se realizó de manera individual, a los evaluadores se le instruyó para que hicieran la estimación visual y manifestaran verbalmente su resultado. Los datos se almacenaron mediante una grabación de audio; las variables registradas fueron: Porcentaje de severidad estimado y Tiempo de evaluación por hoja. Adicionalmente, para medir la confiabilidad intra-evaluador, todos participantes hicieron una evaluación adicional con un subconjunto de 25 imágenes (cinco por cada tipo de hoja).

1.3.4. Análisis de datos.

El análisis de los datos se hizo con el programa R (R Core Team 2016) y con Excel 2016 (Microsoft Corp.). Los parámetros que miden la confiabilidad, precisión, sesgo y concordancia se calcularon para hacer comparaciones entre los estilos cognitivos (impulsividad-reflexividad), los tipos de hoja (lineal, elíptica, lobulada, triangular y palmeada), los patrones de infección (aleatorio y agregado) y la experiencia de los evaluadores.

El nivel de concordancia de las estimaciones se obtuvo a través de la correlación de concordancia de Lin (LCC, Lin 1989). Con este método se calcula el grado en que los pares de observaciones (severidad estimada y severidad real) se aproximan a la línea de concordancia de 45° (pendiente=1, intercepto=0, *i.e.* concordancia perfecta) (El Jarroudi *et al.*, 2015). El análisis LCC considera a los componentes esenciales de la concordancia (precisión y exactitud), y de su cálculo se derivan varios parámetros que miden el sesgo de ubicación (μ), sesgo sistemático (v), sesgo generalizado (C_b) y exactitud (ρ_c). El parámetro ρ_c es el Coeficiente de Correlación de Concordancia de Lin (LCCC) y toma valores entre 0 y 1. Adicionalmente, como medida de la precisión, se calculó el coeficiente de correlación (r).

El LCCC es el producto del sesgo generalizado (C_b) y el coeficiente de correlación (r). Los cálculos se hicieron como se ha descrito previamente (Lin, 1989; Nita *et al.*, 2003; Madden *et al.*, 2007; Yadav *et al.*, 2013; El Jarroudi *et al.*, 2015). El máximo nivel de concordancia se obtiene cuando los pares de observaciones están sobre la línea de

concordancia (i.e., $\mu = 0$, $v = 1$, $C_b = 1$, $r = 1$, lo que implica $\rho_c = 1$). Cualquier desviación de esos valores indica sesgo, imprecisión y pérdida de exactitud (Lin, 1989; Nita *et al.*, 2003; Madden *et al.*, 2007; Bock *et al.*, 2015)

Para estimar la concordancia entre X (severidad real) y Y (severidad estimada), el parámetro ρ_c se obtuvo con las siguientes fórmulas:

$$\rho_c = rC_b$$

El parámetro C_b mide la distancia entre la línea mejor ajustada y la línea de concordancia (45°), por lo que es una estimación de la exactitud. El parámetro r es el coeficiente de correlación y mide la precisión. El valor de C_b se obtiene de la siguiente manera:

$$C_b = \left[\left(v + \frac{1}{v} + u^2 \right) / 2 \right]^{-1}$$

Donde

$v = \frac{\sigma_y}{\sigma_x}$ es una medición del sesgo sistemático y $u = \frac{(\mu_y - \mu_x)}{\sqrt{2\sigma_y\sigma_x}}$ es una medición del sesgo de ubicación. σ y μ representan la varianza y media, respectivamente.

Para medir la confiabilidad intra-evaluador se hizo un análisis de regresión con las estimaciones repetidas (25 imágenes) de los 26 evaluadores. De este análisis se obtuvo el coeficiente de determinación (R^2) como indicador de la confiabilidad, se calcularon los promedios por estilo cognitivo, tipo de hoja y patrón de infección. Con el mismo método, se midió la confiabilidad inter-evaluadores para cada combinación de pares de evaluadores; se registró la frecuencia de los valores de R^2 para cada tipo de hoja, patrón de infección y estilo cognitivo. Cuando $R^2=1$ se obtiene el nivel más alto de confiabilidad

(Nutter *et al.*, 1993; Bock *et al.*, 2015). Adicionalmente, la confiabilidad inter-evaluadores se midió a través del coeficiente de correlación intra-clase (ρ); para lo anterior, la base de datos se subdividió por estilo cognitivo, tipo de hoja y patrón de infección. Se realizó un análisis de cada conjunto de datos a través de un ANOVA de dos vías con efectos aleatorios, tal como lo describen Nita *et al.* (2003). El coeficiente de correlación intra-clase (ρ) se obtuvo a través de los componentes de la varianza σ_{sev}^2 , σ_{eval}^2 y σ_{error}^2 ; de tal manera que $\rho = \sigma_{sev}^2 / (\sigma_{sev}^2 + \sigma_{eval}^2 + \sigma_{error}^2)$. Los efectos aleatorios fueron la severidad y los evaluadores; lo anterior se debe a que los niveles de severidad representados en las hojas se determinaron a través de números aleatorios, además de que el grupo de evaluadores representa una muestra aleatoria. Antes del análisis, los datos se sometieron a la transformación angular: $Z = \arcsin(\sqrt{\% \text{ estimado} / 100})$.

El efecto del estilo cognitivo (reflexivo-impulsivo) en los parámetros indicadores de la exactitud, precisión y concordancia (μ , v , C_b , r , y ρ_c) y en el tiempo de evaluación (t), se estudió a través de un análisis bootstrap; con esta técnica de remuestreo se calcularon los intervalos de confianza de 95% (CIs) para la diferencia entre los parámetros estimados de ambos estilos cognitivos. Cuando los CIs incluyeron al cero, se concluyó que no hubo diferencia significativa entre evaluadores reflexivos e impulsivos. De la misma forma, se analizó el efecto de los patrones de lesiones (aleatorio y agregado) y de la experiencia de los evaluadores (experimentados e inexperimentados). Para desarrollar las pruebas anteriores, se realizó un análisis para cada tipo de hoja, generando 2 000 muestras balanceadas con el paquete bootstrap de R. Adicionalmente, para la distribución de los parámetros estimados se aplicó una prueba de normalidad

Shapiro-Wilk (SW) (función shapiro.test). Cuando los datos de los parámetros (μ , v , C_b , r , ρ_c y t) tuvieron distribución normal se aplicó una prueba de t (función t.test) para comparar las medias de dos grupos (reflexivos vs impulsivos, PI-AG vs PI-AL y experimentados vs inexperimentados). Cuando los datos no siguieron una distribución normal se realizó un análisis no paramétrico, aplicando la prueba de Kolmogorov-Smirnov (KS) (función ks.test) para comparar las distribuciones de los grupos antes mencionados.

Los errores se analizaron a través de gráficos de Bland-Altman; este es un método para examinar el grado de concordancia entre los valores estimados y los valores de severidad real (Bland y Altman, 1999). En este análisis visual se hace un gráfico de dispersión XY , en el cual el eje X muestra la severidad real, mientras que Y representa el error absoluto (severidad estimada- severidad real). Los límites de concordancia de 95% se establecen a dos desviaciones estándar ($2s$) de la media de los errores absolutos. Este método asume una distribución normal de los errores absolutos; aunque este supuesto no se cumple a bajos niveles de severidad (0-10%) (Bock *et al.*, 2010), los gráficos Bland-Altman permiten visualizar algunos componentes del error como el sesgo o la imprecisión.

1.4. RESULTADOS

1.4.1. Efecto del estilo cognitivo

Los componentes estadísticos que miden el sesgo (v , u y C_b) variaron dependiendo del tipo de hoja y el patrón de las infecciones; aunque en la mayoría de estos parámetros se registraron diferencias significativas entre los estilos cognitivos, ningún estilo mostró resultados predominantemente mejores en todos los tipos de hoja o patrones de infección (Cuadro 1.3). Adicionalmente, el componente estadístico que mide la precisión (r) proporcionó valores significativamente más altos para los evaluadores reflexivos; para este parámetro, la diferencia media entre ambos estilos cognitivos estuvo entre 0.048 y 0.098 para todos los tipos de hoja.

El tipo de hoja lineal se diferenció porque fue el único en el que el sesgo sistemático mostró valores mayores que 1 (v ; reflexivos=1.232, impulsivos=1.295) (Figura 1.2 A, B); además, para este tipo de hoja, el sesgo de ubicación fue mayor que 0 en ambos estilos cognitivos (u ; reflexivos =0.202, impulsivos=0.162). En general, el sesgo de ubicación es distinto en magnitud y signo dependiendo del tipo de hoja; para las hojas elíptica, lobulada y palmeada, el valor de u es menor que 0, con mejores resultados para los evaluadores reflexivos, mientras que para las hojas lineal y triangular, el valor de u es mayor que 0 y los mejores resultados se obtuvieron con los evaluadores impulsivos (Cuadro 1.3).

La concordancia general (ρ_c), entre la severidad estimada y la severidad real, fue distinta para cada tipo de hoja y cada patrón de infección; los resultados de los evaluadores reflexivos estuvieron más cercanos a 1, en comparación con los evaluadores impulsivos. El análisis bootstrap, la prueba *KS* y la prueba *t* indican niveles de concordancia significativamente más altos para los evaluadores reflexivos.

Los evaluadores impulsivos dedicaron más tiempo a la evaluación de hojas de forma simple (lineal = 6.693 s, elíptica = 7.186 s), en comparación con los evaluadores reflexivos (lineal = 6.619 s, elíptica = 6.918 s). En contraste, los evaluadores reflexivos tomaron más tiempo para la evaluación de hojas de forma compleja (lobulada = 7.488 s, palmeada = 8.087 s), en comparación con los evaluadores impulsivos (lobulada = 7.253 s, palmeada = 7.983 s). Las pruebas estadísticas mostraron un efecto significativo del estilo cognitivo en el tiempo de evaluación de los cuatro tipos de hoja indicados. Finalmente, el tiempo de evaluación para la hoja triangular fue indistinto para ambos estilos cognitivos (Cuadro 1.3).

Para los patrones de infección de las lesiones, los evaluadores reflexivos mostraron una precisión más alta (r ; PI-AG=0.899, PI-AL=0.892), en comparación con los evaluadores impulsivos (r ; PI-AG=0.837, PI-AL=0.819). Adicionalmente, los PI-AG están asociados a sesgos sistemáticos menores que 1 (v ; reflexivos=0.950, impulsivos=0.977), mientras que los PI-AL están asociados a un sesgo sistemático mayor que 1 (v ; reflexivos =1.035, impulsivos=1.094).

Cuadro 1.3. Efecto del estilo cognitivo en el sesgo, precisión, concordancia y tiempo de evaluación en distintos tipos de hoja y patrones de infección

VARIABLE ^a	MEDIA ^b		DIFERENCIA ENTRE MEDIAS ^c	95% CIs ^d DE LA DIFERENCIA		KS value (P)	t - value (P)
	Evaluadores Reflexivos	Evaluadores Impulsivos					
Lineal							
Sesgo sistemático (v)	1.232 (0.024)	1.295 (0.031)	-0.063 (0.039)	-0.0650	-0.0616	0.771 (<0.0001)	
Sesgo de ubicación (μ)	0.202 (0.024)	0.162 (0.03)	0.041 (0.037)	0.0390	0.0423	0.567 (<0.0001)	
Sesgo generalizado (C_b)	0.959 (0.007)	0.955 (0.009)	0.004 (0.011)	0.0037	0.0047	0.22 (<0.0001)	
Coef. de correlación (r)	0.924 (0.008)	0.876 (0.015)	0.048 (0.018)	0.0475	0.0490	0.96 (<0.0001)	
LCCC (ρ_c)	0.887 (0.011)	0.837 (0.018)	0.05 (0.021)	0.0491	0.0509	0.921 (<0.0001)	
Tiempo ^e (segundos)	6.619 (0.195)	6.693 (0.198)	-0.074 (0.277)	-0.0862	-0.0619	0.162 (<0.0001)	
Elíptica							
Sesgo sistemático (v)	0.911 (0.019)	0.939 (0.028)	-0.027 (0.034)	-0.0288	-0.0258	0.47 (<0.0001)	
Sesgo de ubicación (μ)	-0.116 (0.024)	-0.132 (0.032)	0.017 (0.04)	0.0148	0.0183	0.258 (<0.0001)	
Sesgo generalizado (C_b)	0.989 (0.004)	0.988 (0.006)	0.0002 (0.007)	-0.0001	0.0005	0.092 (<0.0001)	
Coef. de correlación (r)	0.914 (0.01)	0.848 (0.023)	0.067 (0.025)	0.0658	0.0680	0.967 (<0.0001)	
LCCC (ρ_c)	0.904 (0.011)	0.838 (0.022)	0.066 (0.024)	0.0653	0.0674	0.96 (<0.0001)	
Tiempo (segundos)	6.918 (0.237)	7.186 (0.216)	-0.268 (0.325)	-0.2822	-0.2536		-37.386 (<0.0001)
Triangular							
Sesgo sistemático (v)	0.938 (0.02)	0.963 (0.035)	-0.025 (0.04)	-0.0266	-0.0231	0.391 (<0.0001)	
Sesgo de ubicación (μ)	0.050 (0.023)	0.012 (0.034)	0.038 (0.041)	0.0360	0.0396	0.517 (<0.0001)	
Sesgo generalizado (C_b)	0.996 (0.002)	0.998 (0.002)	-0.002 (0.002)	-0.0019	-0.0017	0.508 (<0.0001)	
Coef. de correlación (r)	0.917 (0.009)	0.819 (0.019)	0.098 (0.021)	0.0972	0.0991	0.999 (<0.0001)	
LCCC (ρ_c)	0.914 (0.009)	0.817 (0.019)	0.096 (0.021)	0.0954	0.0972	0.999 (<0.0001)	
Tiempo (segundos)	8.048 (0.283)	8.039 (0.227)	0.008 (0.358)	-0.0074	0.0240	0.057 (0.0001)	
Lobulada							
Sesgo sistemático (v)	0.935 (0.023)	0.971 (0.036)	-0.035 (0.042)	-0.0371	-0.0334	0.478 (<0.0001)	
Sesgo de ubicación (μ)	-0.116 (0.027)	-0.171 (0.036)	0.055 (0.045)	0.0530	0.0569	0.626 (<0.0001)	
Sesgo generalizado (C_b)	0.990 (0.004)	0.984 (0.007)	0.007 (0.008)	0.0062	0.0069	0.455 (<0.0001)	
Coef. de correlación (r)	0.894 (0.012)	0.819 (0.020)	0.076 (0.023)	0.0746	0.0767	0.984 (<0.0001)	
LCCC (ρ_c)	0.886 (0.013)	0.806 (0.020)	0.080 (0.024)	0.0793	0.0813	0.986 (<0.0001)	
Tiempo (segundos)	7.488 (0.27)	7.253 (0.244)	0.235 (0.365)	0.2192	0.2511	0.358 (<0.0001)	
Palmeada							
Sesgo sistemático (v)	0.894 (0.025)	0.957 (0.036)	-0.063 (0.042)	-0.0653	-0.0616	0.718 (<0.0001)	
Sesgo de ubicación (μ)	-0.065 (0.028)	-0.095 (0.035)	0.031 (0.045)	0.0290	0.0329	0.387 (<0.0001)	
Sesgo generalizado (C_b)	0.991 (0.004)	0.993 (0.004)	-0.002 (0.006)	-0.0027	-0.0022	0.274 (<0.0001)	
Coef. de correlación (r)	0.890 (0.011)	0.826 (0.017)	0.064 (0.021)	0.0628	0.0646	0.98 (<0.0001)	
LCCC (ρ_c)	0.882 (0.011)	0.821 (0.017)	0.061 (0.021)	0.0601	0.0620	0.968 (<0.0001)	
Tiempo (segundos)	8.087 (0.292)	7.983 (0.251)	0.104 (0.382)	0.0876	0.1211	0.165 (<0.0001)	
PI-AG							
Sesgo sistemático (v)	0.950 (0.014)	0.977 (0.019)	-0.027 (0.024)	-0.0282	-0.0261	0.588 (<0.0001)	
Sesgo de ubicación (μ)	-0.039 (0.017)	-0.079 (0.021)	0.04 (0.026)	0.0383	0.0407	0.718 (<0.0001)	
Sesgo generalizado (C_b)	0.998 (0.001)	0.996 (0.002)	0.001 (0.002)	0.0013	0.0015	0.352 (<0.0001)	
Coef. de correlación (r)	0.899 (0.007)	0.837 (0.012)	0.061 (0.013)	0.0608	0.0620	0.999 (<0.0001)	
LCCC (ρ_c)	0.897 (0.007)	0.834 (0.011)	0.062 (0.014)	0.0618	0.0630	0.999 (<0.0001)	
Tiempo (segundos)	7.274 (0.163)	7.143 (0.139)	0.131 (0.213)	0.1216	0.1403	0.345 (<0.0001)	
PI-AL							
Sesgo sistemático (v)	1.035 (0.018)	1.094 (0.023)	-0.059 (0.03)	-0.0600	-0.0574	0.844 (<0.0001)	
Sesgo de ubicación (μ)	0.032 (0.017)	0.0001 (0.022)	0.032 (0.027)	0.0310	0.0334	0.602 (<0.0001)	
Sesgo generalizado (C_b)	0.999 (0.001)	0.996 (0.002)	0.003 (0.002)	0.0030	0.0032	0.717 (<0.0001)	
Coef. de correlación (r)	0.892 (0.007)	0.819 (0.012)	0.073 (0.014)	0.0726	0.0738	0.999 (<0.0001)	
LCCC (ρ_c)	0.891 (0.007)	0.815 (0.013)	0.076 (0.015)	0.0750	0.0762	0.999 (<0.0001)	
Tiempo (segundos)	7.598 (0.168)	7.708 (0.153)	-0.112 (0.228)	-0.1201	-0.1001	0.280 (<0.0001)	

^a El máximo nivel de concordancia se obtiene con $\mu = 0$, $v = 1$, $C_b = 1$, $r = 1$, lo que implica $\rho_c = 1$

^b Desviación estándar entre paréntesis

^c Promedio de la diferencia de cada parámetro, desviación estándar entre paréntesis. Basado en 2 000 muestras bootstrap.

^d Intervalo de confianza (CIs). Basado en 2 000 muestras bootstrap. Si CIs incluye al cero, la diferencia no es significativa.

^e Tiempo de evaluación por hoja.

El análisis bootstrap se desarrolló con el conjunto combinado de datos de todos los evaluadores; para cada sección, los datos se dividieron por estilo cognitivo, tipo de hoja y por patrón de infección.

Cuadro 1.4. Efecto del patrón de las lesiones en el sesgo, precisión, concordancia y tiempo de evaluación en distintos tipos de hoja.

VARIABLE ^a	MEDIA ^b		DIFERENCIA ENTRE MEDIAS ^c	95% CIs ^d DE LA DIFERENCIA		KS value (P)	t - value (P)
	Patrones Agregados	Patrones Aleatorios					
Lineal							
Sesgo sistemático (v)	1.137 (0.023)	1.357 (0.027)	-0.219 (0.036)	-0.2208	-0.2176	0.9998 (<0.0001)	
Sesgo de ubicación (μ)	0.097 (0.026)	0.262 (0.029)	-0.165 (0.040)	-0.1670	-0.1636	0.9945 (<0.0001)	
Sesgo generalizado (C_b)	0.987 (0.004)	0.925 (0.010)	0.062 (0.011)	0.0617	0.0626	0.9998 (<0.0001)	
Coef. de correlación (r)	0.901 (0.013)	0.906 (0.011)	-0.005 (0.017)	-0.0056	-0.0041	0.1725 (<0.0001)	
LCCC (ρ_c)	0.889 (0.014)	0.838 (0.015)	0.052 (0.021)	0.0506	0.0524	0.9230 (<0.0001)	
Tiempo ^e (segundos)	6.474 (0.188)	6.832 (0.209)	-0.358 (0.276)	-0.3701	-0.3459	0.6470 (<0.0001)	
Elíptica							
Sesgo sistemático (v)	0.919 (0.023)	0.929 (0.024)	-0.01 (0.033)	-0.0114	-0.0085	0.1815 (<0.0001)	
Sesgo de ubicación (μ)	-0.133 (0.028)	-0.115 (0.03)	-0.019 (0.041)	-0.0204	-0.0169		-20.466 (<0.0001)
Sesgo generalizado (C_b)	0.987 (0.005)	0.990 (0.005)	-0.003 (0.007)	-0.0033	-0.0027	0.2480 (<0.0001)	
Coef. de correlación (r)	0.886 (0.016)	0.874 (0.020)	0.013 (0.025)	0.0114	0.0137	0.2775 (<0.0001)	
LCCC (ρ_c)	0.875 (0.016)	0.865 (0.019)	0.01 (0.025)	0.0087	0.0109	0.2255 (<0.0001)	
Tiempo (segundos)	6.815 (0.214)	7.302 (0.237)	-0.487 (0.318)	-0.5005	-0.4726	0.7160 (<0.0001)	
Triangular							
Sesgo sistemático (v)	0.915 (0.025)	0.979 (0.029)	-0.064 (0.039)	-0.0656	-0.0622	0.7565 (<0.0001)	
Sesgo de ubicación (μ)	-0.006 (0.029)	0.066 (0.029)	-0.071 (0.041)	-0.0732	-0.0695		-78.189 (<0.0001)
Sesgo generalizado (C_b)	0.995 (0.003)	0.997 (0.002)	-0.001 (0.003)	-0.0016	-0.0014	0.2735 (<0.0001)	
Coef. de correlación (r)	0.872 (0.015)	0.864 (0.016)	0.008 (0.021)	0.0075	0.0094	0.2335 (<0.0001)	
LCCC (ρ_c)	0.868 (0.015)	0.861 (0.016)	0.007 (0.022)	0.0061	0.0080	0.1915 (<0.0001)	
Tiempo (segundos)	7.983 (0.269)	8.079 (0.241)	-0.096 (0.36)	-0.1122	-0.0806	0.1675 (<0.0001)	
Lobulada							
Sesgo sistemático (v)	0.931 (0.029)	0.971 (0.032)	-0.040 (0.043)	-0.0414	-0.0376	0.4965 (<0.0001)	
Sesgo de ubicación (μ)	-0.175 (0.034)	-0.114 (0.031)	-0.061 (0.046)	-0.0630	-0.0590	0.6660 (<0.0001)	
Sesgo generalizado (C_b)	0.982 (0.007)	0.992 (0.004)	-0.011 (0.008)	-0.0110	-0.0102	0.6710 (<0.0001)	
Coef. de correlación (r)	0.853 (0.017)	0.857 (0.016)	-0.004 (0.023)	-0.0051	-0.0031	0.1060 (<0.0001)	
LCCC (ρ_c)	0.837 (0.019)	0.850 (0.016)	-0.013 (0.024)	-0.0141	-0.0120	0.2965 (<0.0001)	
Tiempo (segundos)	7.096 (0.247)	7.656 (0.276)	-0.560 (0.371)	-0.5760	-0.5434	0.7115 (<0.0001)	
Palmeada							
Sesgo sistemático (v)	0.871 (0.026)	0.971 (0.03)	-0.1 (0.039)	-0.1012	-0.0978	0.934 (<0.0001)	
Sesgo de ubicación (μ)	-0.095 (0.031)	-0.065 (0.031)	-0.03 (0.044)	-0.0320	-0.0281		-30.691 (<0.0001)
Sesgo generalizado (C_b)	0.985 (0.006)	0.996 (0.003)	-0.011 (0.006)	-0.0115	-0.0110	0.8235 (<0.0001)	
Coef. de correlación (r)	0.862 (0.014)	0.853 (0.015)	0.009 (0.021)	0.0080	0.0098	0.2540 (<0.0001)	
LCCC (ρ_c)	0.849 (0.014)	0.850 (0.016)	-0.0009 (0.021)	-0.0018	0.0001	0.0410 (0.06933)	
Tiempo (segundos)	7.679 (0.234)	8.391 (0.283)	-0.713 (0.371)	-0.7289	-0.6964	0.8355 (<0.0001)	

^a El máximo nivel de concordancia se obtiene con $\mu = 0$, $v = 1$, $C_b = 1$, $r = 1$, lo que implica $\rho_c = 1$

^b Desviación estándar entre paréntesis

^c Promedio de la diferencia de cada parámetro, desviación estándar entre paréntesis. Basado en 2 000 muestras bootstrap.

^d Intervalo de confianza (CIs). Basado en 2 000 muestras bootstrap. Si CIs incluye al cero, la diferencia no es significativa.

^e Tiempo de evaluación por hoja.

El análisis bootstrap se desarrolló con el conjunto combinado de datos de todos los evaluadores; para cada sección, los datos se dividieron por tipo de hoja y por patrón de infección.

1.4.2. Efecto del patrón de las lesiones

El análisis bootstrap de los componentes del LCC mostró que los patrones de infección (PI) tienen un efecto significativo en el sesgo (v , u y C_b) y en la precisión (r) de la evaluación visual. Al igual que en los resultados anteriores, el tipo de hoja es un factor determinante en la concordancia de las estimaciones, principalmente en los parámetros que describen la magnitud del sesgo (Cuadro 1.4). De esta forma, la hoja lineal fue la única que de manera reiterada registró un sesgo sistemático mayor que 1 (v ; PI-AG=1.137, PI-AL=1.357) y el sesgo de ubicación fue mayor que 0 para ambos patrones de infección (u ; PI-AG=0.097, PI-AL=0.262). El estadístico que mide la concordancia (ρ_c) mostró los valores más cercanos a 1 cuando las imágenes tuvieron PI-AG en hojas de forma simple (lineal, elíptica y triangular); en contraste, las imágenes con PI-AL registraron sus niveles más altos de concordancia en hojas de forma compleja (lobulada y palmeada). El tiempo de evaluación es significativamente mayor para las imágenes con PI-AL (KS test $P < 0.0001$), este resultado se verificó con todos los tipos de hoja evaluados.

1.4.3. Efecto de la experiencia del evaluador

Para desarrollar el remuestreo bootstrap, se tomaron los datos de los 10 evaluadores inexperimentados y los datos de 10 evaluadores experimentados, los cuales fueron seleccionados al azar; esta reducción en la base de datos se hizo para balancear la prueba. En el análisis de concordancia, la precisión (r) es el parámetro que marcó la diferencia entre evaluadores experimentados e inexperimentados.

Los evaluadores con experiencia registraron una precisión significativamente mayor en todos los tipos de hoja y patrones de infección, estos resultados se obtuvieron a través de los 95% CI y la prueba KS. Por su parte, los parámetros relacionados con el sesgo (v , u y C_b) variaron dependiendo del tipo de hoja y el patrón de las infecciones; en general, estos estimadores mostraron diferencias significativas entre los niveles de experiencia; sin embargo, los evaluadores experimentados no tuvieron una asociación predominante con los mejores valores (Cuadro 1.5)

En la concordancia general (ρ_c), los resultados de los evaluadores experimentados estuvieron más cercanos a 1, en comparación con los evaluadores inexperimentados. Las diferencias son significativas en todos los tipos de hoja y patrones de infección (prueba KS $P < 0.0001$). El análisis bootstrap, la prueba KS y la prueba t indican que los evaluadores con experiencia hacen evaluaciones de mayor calidad debido a su precisión, lo cual les concede una concordancia significativamente más alta. Adicionalmente, los evaluadores experimentados mostraron una mayor inversión de tiempo en el proceso de evaluación, en comparación con los evaluadores inexperimentados. Las diferencias de las medias fueron positivas para los evaluadores con experiencia que dedicaron entre 0.480 y 1.058 s más que los evaluadores sin experiencia. Todas las pruebas estadísticas mostraron un efecto significativo del nivel de experiencia en el tiempo de evaluación.

Cuadro 1.5. Efecto de la experiencia del evaluador en el sesgo, precisión, concordancia y tiempo de evaluación en distintos tipos de hoja y patrones de infección

VARIABLE ^a	MEDIA ^b		DIFERENCIA ENTRE MEDIAS ^c	95% CIs ^d DE LA DIFERENCIA		KS value (P)	t - value (P)
	Evalua-dores Experimentados	Evalua-dores Inexperimen-tados					
Lineal							
Sesgo sistemático (v)	1.197 (0.028)	1.335 (0.034)	-0.138 (0.043)	-0.1395	-0.1357		-140.78 (<0.0001)
Sesgo de ubicación (μ)	0.175 (0.026)	0.239 (0.036)	-0.064 (0.044)	-0.0664	-0.0625	0.7150 (<0.0001)	
Sesgo generalizado (C_b)	0.969 (0.007)	0.933 (0.012)	0.036 (0.014)	0.0350	0.0362	0.9415 (<0.0001)	
Coef. de correlación (r)	0.929 (0.009)	0.864 (0.018)	0.065 (0.02)	0.0639	0.0657	0.9800 (<0.0001)	
LCCC (ρ_c)	0.9 (0.012)	0.807 (0.021)	0.093 (0.024)	0.0924	0.0946	0.9965 (<0.0001)	
Tiempo ^e (segundos)	7.471 (0.239)	6.413 (0.22)	1.058 (0.329)	1.0438	1.0727	0.9825 (<0.0001)	
Elíptica							
Sesgo sistemático (v)	0.902 (0.022)	0.983 (0.032)	-0.082 (0.039)	-0.0833	-0.0799	0.8850 (<0.0001)	
Sesgo de ubicación (μ)	-0.094 (0.026)	-0.094 (0.039)	0.002 (0.047)	-0.0018	0.0023	0.1085 (<0.0001)	
Sesgo generalizado (C_b)	0.99 (0.004)	0.994 (0.004)	-0.004 (0.006)	-0.0047	-0.0042	0.4780 (<0.0001)	
Coef. de correlación (r)	0.933 (0.009)	0.83 (0.027)	0.103 (0.028)	0.1015	0.1040	0.9985 (<0.0001)	
LCCC (ρ_c)	0.924 (0.01)	0.826 (0.026)	0.098 (0.028)	0.0968	0.0992	0.9975 (<0.0001)	
Tiempo (segundos)	7.867 (0.235)	7.010 (0.286)	0.867 (0.371)	0.8510	0.8835	0.9035 (<0.0001)	
Triangular							
Sesgo sistemático (v)	0.893 (0.028)	1.034 (0.036)	-0.142 (0.047)	-0.1436	-0.1395	0.9705 (<0.0001)	
Sesgo de ubicación (μ)	0.046 (0.03)	0.046 (0.036)	-0.001 (0.048)	-0.0030	0.0012	0.0465 (0.01348)	
Sesgo generalizado (C_b)	0.992 (0.003)	0.997 (0.003)	-0.006 (0.004)	-0.0058	-0.0054	0.6845 (<0.0001)	
Coef. de correlación (r)	0.892 (0.012)	0.84 (0.021)	0.053 (0.025)	0.0515	0.0537	0.8880 (<0.0001)	
LCCC (ρ_c)	0.885 (0.012)	0.837 (0.022)	0.047 (0.026)	0.0463	0.0486	0.8470 (<0.0001)	
Tiempo (segundos)	8.759 (0.259)	8.253 (0.329)	0.506 (0.421)	0.4878	0.5247	0.6350 (<0.0001)	
Lobulada							
Sesgo sistemático (v)	0.922 (0.03)	1.001 (0.041)	-0.078 (0.05)	-0.0806	-0.0762	0.7315 (<0.0001)	
Sesgo de ubicación (μ)	-0.106 (0.03)	-0.139 (0.044)	0.033 (0.053)	0.0308	0.0355	0.3685 (<0.0001)	
Sesgo generalizado (C_b)	0.99 (0.005)	0.989 (0.007)	0.002 (0.008)	0.0012	0.0019	0.0930 (<0.0001)	
Coef. de correlación (r)	0.903 (0.012)	0.798 (0.024)	0.105 (0.026)	0.1036	0.1059	0.9985 (<0.0001)	
LCCC (ρ_c)	0.894 (0.012)	0.789 (0.024)	0.105 (0.026)	0.1038	0.1061	0.9985 (<0.0001)	
Tiempo (segundos)	8.134 (0.263)	7.488 (0.347)	0.646 (0.43)	0.6273	0.6651	0.7245 (<0.0001)	
Palmeada							
Sesgo sistemático (v)	0.915 (0.034)	0.943 (0.036)	-0.028 (0.05)	-0.0298	-0.0255	0.3090 (<0.0001)	
Sesgo de ubicación (μ)	-0.040 (0.033)	-0.102 (0.039)	0.061 (0.05)	0.0590	0.0634	0.6270 (<0.0001)	
Sesgo generalizado (C_b)	0.994 (0.004)	0.992 (0.006)	0.002 (0.007)	0.0020	0.0026	0.1995 (<0.0001)	
Coef. de correlación (r)	0.875 (0.013)	0.832 (0.02)	0.043 (0.024)	0.0421	0.0442	0.8190 (<0.0001)	
LCCC (ρ_c)	0.870 (0.012)	0.825 (0.02)	0.045 (0.024)	0.0438	0.0458	0.8430 (<0.0001)	
Tiempo (segundos)	8.748 (0.29)	8.269 (0.32)	0.480 (0.430)	0.4607	0.4984	0.5735 (<0.0001)	
PI-AG							
Sesgo sistemático (v)	0.907 (0.016)	1.022 (0.022)	-0.115 (0.027)	-0.1160	-0.1136	0.9975 (<0.0001)	
Sesgo de ubicación (μ)	-0.051 (0.018)	-0.028 (0.024)	-0.023 (0.031)	-0.0240	-0.0213	0.4090 (<0.0001)	
Sesgo generalizado (C_b)	0.994 (0.002)	0.999 (0.001)	-0.005 (0.002)	-0.0053	-0.0051	0.9170 (<0.0001)	
Coef. de correlación (r)	0.908 (0.007)	0.829 (0.014)	0.079 (0.015)	0.0784	0.0797	0.9990 (<0.0001)	
LCCC (ρ_c)	0.902 (0.007)	0.828 (0.014)	0.074 (0.016)	0.0736	0.0749	0.9990 (<0.0001)	
Tiempo (segundos)	7.998 (0.159)	7.154 (0.186)	0.843 (0.244)	0.8327	0.8541	0.9845 (<0.0001)	
PI-AL							
Sesgo sistemático (v)	1.033 (0.022)	1.127 (0.026)	-0.094 (0.034)	-0.0954	-0.0924		-124.58 (<0.0001)
Sesgo de ubicación (μ)	0.049 (0.019)	0.024 (0.026)	0.025 (0.032)	0.0237	0.0265	0.4375 (<0.0001)	
Sesgo generalizado (C_b)	0.998 (0.001)	0.992 (0.003)	0.006 (0.003)	0.0056	0.0059	0.8095 (<0.0001)	
Coef. de correlación (r)	0.890 (0.007)	0.811 (0.015)	0.079 (0.016)	0.0781	0.0795	0.9990 (<0.0001)	
LCCC (ρ_c)	0.888 (0.008)	0.805 (0.015)	0.083 (0.017)	0.0825	0.0840	0.9990 (<0.0001)	
Tiempo (segundos)	8.398 (0.167)	7.817 (0.207)	0.581 (0.269)	0.5694	0.5929	0.8770 (<0.0001)	

^a El máximo nivel de concordancia se obtiene con $\mu = 0$, $v = 1$, $C_b = 1$, $r = 1$, lo que implica $\rho_c = 1$

^b Desviación estándar entre paréntesis

^c Promedio de la diferencia de cada parámetro, desviación estándar entre paréntesis. Basado en 2 000 muestras bootstrap

^d Intervalo de confianza (CIs). Basado en 2 000 muestras bootstrap. Si CIs incluye al cero, la diferencia no es significativa.

^e Tiempo de evaluación por hoja. El análisis bootstrap se desarrolló con el conjunto combinado de datos de 10 evaluadores experimentados y 10 evaluadores inexperimentados; para cada sección, los datos se dividieron por nivel de experiencia, tipo de hoja y por patrón de infección.

Para los patrones de infección de las lesiones, los evaluadores experimentados mostraron una precisión más alta (r ; PI-AG=0.908, PI-AL=0.890), en comparación con los evaluadores sin experiencia (r ; PI-AG=0.829, PI-AL=0.811). Además, los PI-AG están asociados a sesgos de ubicación menores que 0 (u ; experimentados =-0.051, inexperimentados=-0.028), mientras que los PI-AL están asociados con valores mayores a 0 (u ; experimentado =0.049, inexperimentado=0.024). En general, la evaluación de PI-AL requiere mayor tiempo de la evaluación que PI-AG; sin embargo, los niveles más altos de concordancia se registraron en las evaluaciones de los PI-AG.

1.4.4. Efecto del tipo de hoja

El tipo de hoja es un factor determinante en el nivel de concordancia entre la severidad estimada y la severidad real. Algunas hojas de forma simple (lineal y elíptica) representan “tareas sencillas” en las que los evaluadores impulsivos están dispuestos a dedicar mayor tiempo que los evaluadores reflexivos; en contraparte, las hojas de forma compleja (lobulada y palmeada) representan “tareas difíciles” en las cuales los evaluadores impulsivos muestran menos disposición de tiempo que los evaluadores reflexivos; estas diferencias tienen significancia estadística (Cuadro 1.3). De esta manera, el tipo de hoja es determinante del tiempo de evaluación, el cual puede considerarse como una medida del nivel de compromiso del evaluador.

Las hojas de forma simple (lineal, elíptica y triangular) están asociadas con una evaluación más eficaz del PI-AG, en comparación con el PI-AL. Adicionalmente, en hojas de forma compleja (lobulada y palmeada) se obtuvieron mayores niveles de

concordancia (ρ_c) con PI-AL que con PI-AG (Cuadro 1.4). Los parámetros que miden el sesgo sistemático (v) y el sesgo de ubicación (u) varían en magnitud dependiendo del tipo de hoja (Figura 1.2); por ejemplo, las hojas de tipo lineal tienen una asociación recurrente con valores de v mayores que 1 y valores positivos de u (Cuadros 1.3, 1.4 y 1.5), lo cual es un indicativo de una fuerte sobreestimación ya que la varianza de las estimaciones es mayor a la varianza de la severidad real.

Con los datos de la distribución bootstrap de los parámetros, se observó que el promedio de precisión (r) de ambos patrones de infección fue: lineal=0.904, elíptica=0.880, triangular=0.868, lobulada=0.855 y palmeada=0.857. Adicionalmente, el promedio del estadístico LCCC (ρ_c) de ambos patrones de infección fue: lineal=0.863, elíptica=0.870, triangular=0.865, lobulada=0.843 y palmeada=0.849 (Cuadro 1.4). Por lo anterior, los resultados sugieren que el tipo de hoja es uno de los factores que determinan la precisión y la exactitud en el análisis de concordancia.

En cuanto a la confiabilidad intra-evaluador, las hojas de forma simple (lineal, elíptica y triangular) estuvieron asociadas a un mayor grado de confiabilidad ($0.97 < R^2 < 0.98$), en comparación con las hojas de forma compleja (lobulada y palmeada) ($0.90 < R^2 < 0.95$). De manera similar, la confiabilidad inter-evaluadores registró los valores más elevados para las hojas de forma simple, a través del coeficiente de correlación intra-clase (ρ). En la Cuadro 1.6 se observa que los resultados tuvieron la misma tendencia para los dos estilos cognitivos.

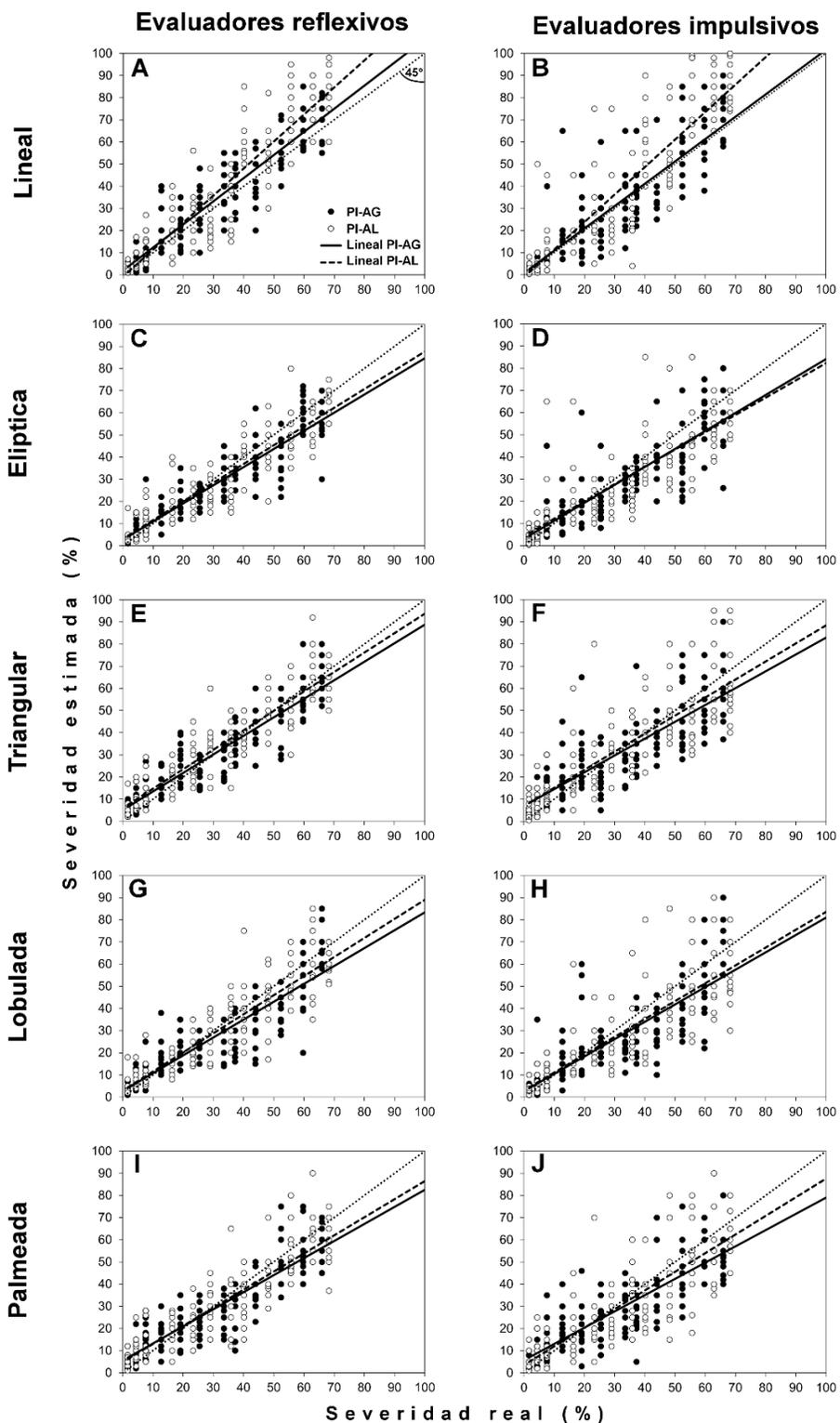


Figura 1.2. Relación entre la severidad real y la severidad estimada para cada tipo de hoja. A, C, E, G e I representa los datos de 13 evaluadores reflexivos; B, D, F, H y J representa los datos de 13 evaluadores impulsivos. La línea gris a 45° es la línea de concordancia.

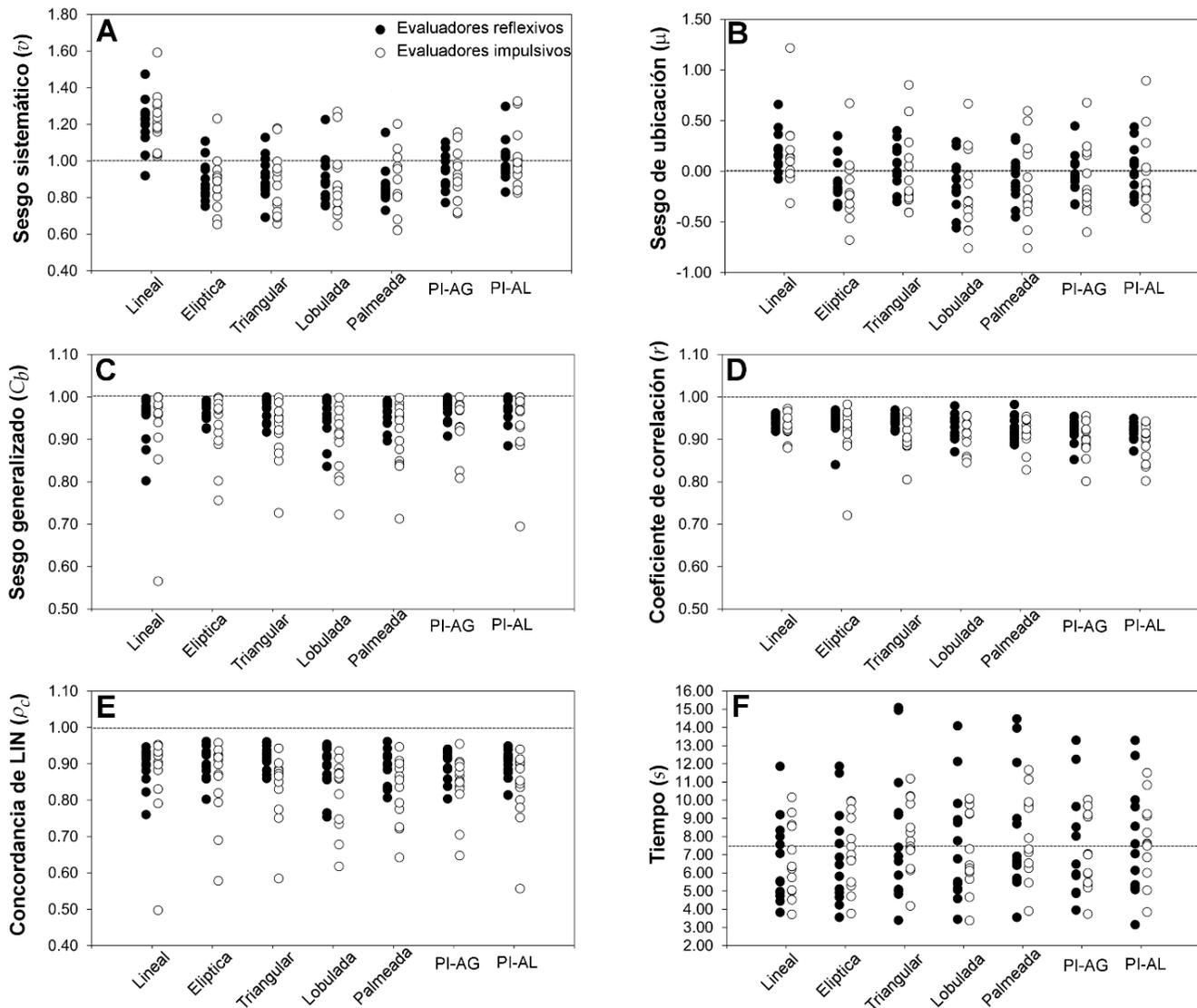


Figura 1.3. Comparación de las mediciones de sesgo, precisión, concordancia y tiempo para cinco tipos de hoja (lineal, elíptica, triangular, lobulada y palmeada) y dos patrones de infección (agregado y aleatorio). El máximo nivel de concordancia se obtiene con $\mu = 0$, $v = 1$, $C_b = 1$, $r = 1$, lo que implica $\rho_c = 1$

Para los resultados de los 26 evaluadores, los valores del sesgo sistemático (v) son consistentemente mayores que 1 para la hoja de tipo lineal (Figura 1.3 A); mientras que en las hojas lobulada y palmeada se registraron frecuentemente valores menores a 1,

siendo éste el nivel óptimo del parámetro. De manera similar, con el sesgo de ubicación (u), los valores para la hoja lineal estuvieron por arriba del nivel óptimo del parámetro (0), y las hojas lobulada y palmeada mostraron una alta frecuencia de estimaciones por abajo de dicho valor. Estas diferencias inciden directamente en el sesgo generalizado (C_b), que es una medida de la exactitud de las estimaciones (Figura. 1.3 A, B, C). En este sentido, los niveles bajos de concordancia en evaluadores impulsivos (Figura 1.3 E) están estrechamente relacionados con baja exactitud (Figura 1.3 C) e imprecisión (Figura 1.3 D).

1.4.5. Confiabilidad inter-evaluadores

La confiabilidad inter-evaluadores fue significativamente más alta con los evaluadores reflexivos (Cuadro 1.6). El coeficiente de correlación intraclase (ρ) y los promedios del coeficiente de determinación (R^2) por pares (Figura 1.4) muestran que en los evaluadores reflexivos es más uniforme la capacidad de evaluación de la enfermedad que en los evaluadores impulsivos; además, los PI-AG están asociados con evaluaciones más consistentes que los PI-AL. Finalmente, las hojas de forma simple (lineal, elíptica y triangular) mostraron mayor confiabilidad inter-evaluadores ($0.85 < \rho < 0.89$) que las hojas de forma compleja (lobulada y palmeada) ($0.80 < \rho < 0.86$).

1.4.6. Confiabilidad intra-evaluadores

Con el subgrupo de 25 imágenes evaluadas en dos ocasiones, se calculó la confiabilidad intra-evaluador a través del coeficiente de determinación (R^2) que, en este caso, es una medida del grado de acuerdo entre dos estimaciones realizadas en diferente momento

por el mismo evaluador. Los promedios de R^2 por estilo cognitivo muestran una tendencia, en la que las estimaciones de los evaluadores impulsivos tienen una mayor consistencia interna que las estimaciones de los evaluadores reflexivos. Este resultado manifiesta la existencia de diferencias en el proceso cognitivo, lo cual deriva en estimaciones menos minuciosas, pero más coincidentes en los evaluadores impulsivos.

Cuadro 1.6. Confiabilidad de la estimación visual de la severidad en relación con el estilo cognitivo reflexividad-impulsividad.

Confiabilidad Intra-evaluador			Confiabilidad Inter-evaluadores			
Coeficiente de determinación (R^2)			Coeficiente de Correlación Intraclase (ρ)			
	Reflexivos	Impulsivos			Reflexivos	Impulsivos
Lineal	0.973	0.979	Lineal	PI-AG	0.888	0.807
				PI-AL	0.886	0.832
Elíptica	0.974	0.976	Elíptica	PI-AG	0.880	0.797
				PI-AL	0.866	0.760
Triangular	0.978	0.979	Triangular	PI-AG	0.869	0.734
				PI-AL	0.857	0.691
Lobulada	0.905	0.934	Lobulada	PI-AG	0.853	0.774
				PI-AL	0.838	0.708
Palmeada	0.925	0.941	Palmeada	PI-AG	0.819	0.736
				PI-AL	0.807	0.712
PI-AG	0.882	0.902				
PI-AL	0.958	0.968				

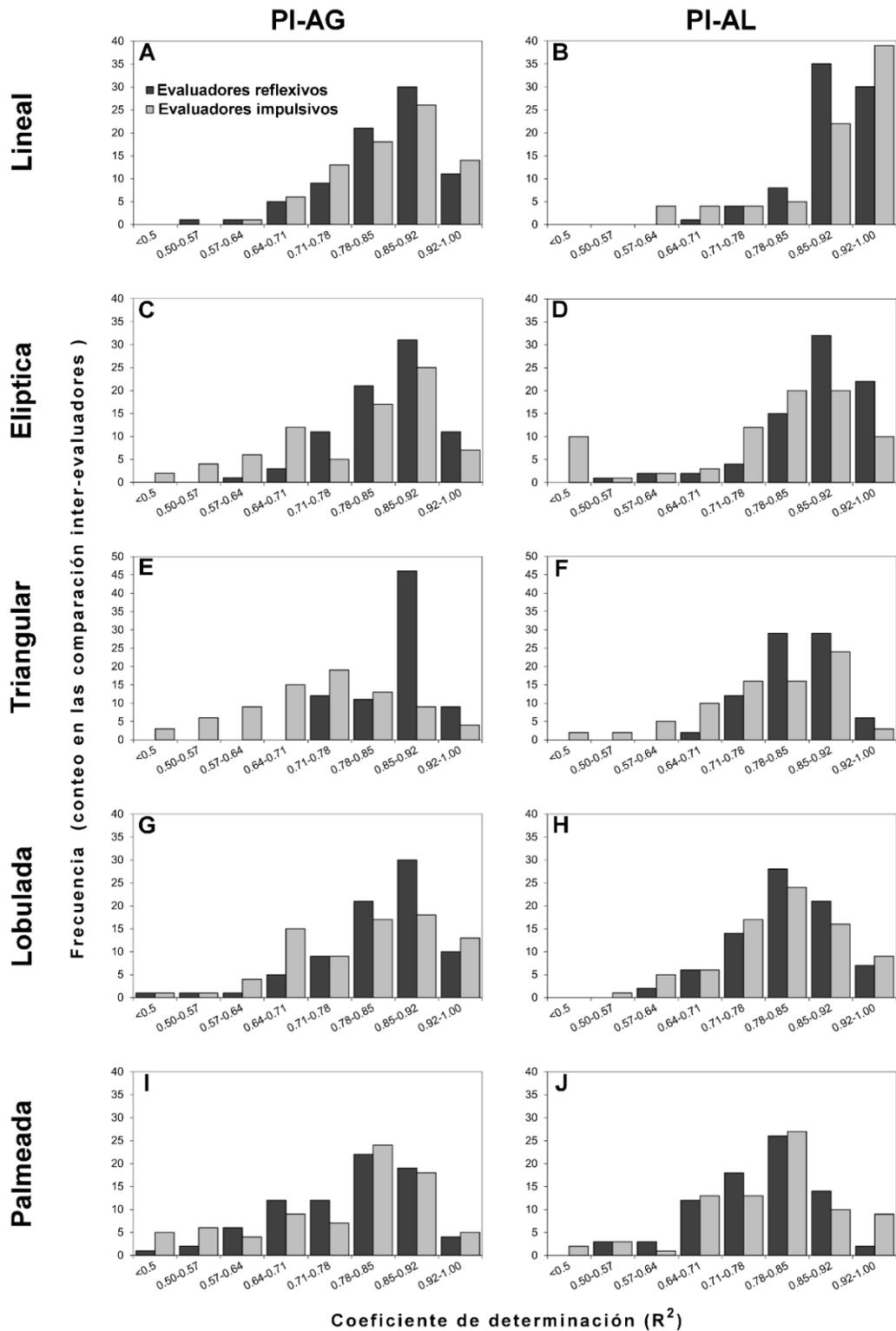


Figura 1.4. Confiabilidad inter evaluadores medida a través de la frecuencia de los coeficientes de determinación (R^2) por pares. Comparación por estilo cognitivo, tipo de hoja y patrón de las lesiones.

1.5. DISCUSIÓN

El estilo cognitivo es una característica inherente de los individuos y es independiente de otras características como la inteligencia o la personalidad. El estilo cognitivo reflexividad-impulsividad es un constructo que puede visualizarse como una variable continua con dos extremos: la reflexividad y la impulsividad (Cairns y Cammock 1978). De manera práctica, los individuos reflexivos se caracterizan por una disposición al análisis y al procesamiento metódico antes de reaccionar o emitir un juicio; en contraste, los individuos impulsivos poseen una predisposición a reaccionar rápidamente a estímulos internos o externos, esta reacción ocurre sin una planeación esmerada y en la mayoría de los casos no existe una valoración de las posibles consecuencias (Kertzman *et al.*, 2009).

El estudio del estilo cognitivo reflexividad-impulsividad se ha aplicado por más de 50 años en el campo de la educación de niños y adolescentes (Kagan, 1965; Buela y Riaño, 2015) además, en ciencias médicas se utiliza para el estudio de personas adultas con trastornos mentales, adicciones y comportamientos patológicos (Quednow *et al.*, 2007; Kertzman *et al.*, 2009).

La calidad de la evaluación visual de enfermedades depende de algunas características propias de los evaluadores, una de las más estudiadas es el nivel de experiencia (Bock *et al.*, 2009; Bock *et al.*, 2013; Bardsley y Ngugi 2013; Yadav *et al.*, 2013; Schwanck y Del Ponte 2014; Braido *et al.*, 2014; Costa Lage *et al.*, 2015). En contraparte, el estilo cognitivo reflexividad-impulsividad no se había estudiado antes, como uno de los factores

que determinan la calidad de la evaluación visual. Bock *et al.* (2009) sugieren que los mejores evaluadores, independientemente de su experiencia, podrían poseer una habilidad innata relacionada con sus competencias visuales. En relación con lo anterior, los resultados del presente estudio permiten hipotetizar que esa habilidad innata podría estar relacionada con el estilo cognitivo reflexividad-impulsividad.

Para todos los tipos de hoja y patrones de infección, los evaluadores reflexivos efectuaron estimaciones más precisas, lo que originó una concordancia significativamente más alta que la obtenida por los evaluadores impulsivos (95% CIs de la diferencia, *KS value* $P < 0.0001$). Para las hojas de forma simple (lineal, elíptica y triangular), los evaluadores impulsivos dedicaron más tiempo a la evaluación, en comparación con los evaluadores reflexivos. En contraste, en las hojas de forma compleja (lobulada y palmeada), los evaluadores impulsivos mostraron menos disposición de tiempo que los evaluadores reflexivos. Este comportamiento puede ser originado por un componente llamado: *urgencia negativa*, en el cual los evaluadores impulsivos tienden a actuar rápidamente cuando se enfrentan a una tarea difícil (Buela y Riaño, 2015) . Una situación similar ocurrió con los patrones de infección, ya que las imágenes con PI-AG incitan al evaluador a un balance de proporciones debido a que las lesiones están agregadas, a diferencia del PI-AL en el que la dispersión de las lesiones no sugiere alguna estrategia en específico. Por esta razón, los PI-AG podrían representar una tarea más elaborada, por lo que los evaluadores impulsivos actuaron más rápido.

Los evaluadores experimentados tienen un mayor grado de concordancia entre sus estimaciones y los valores reales de severidad (Costa *et al.*, 2015). Diversos estudios

demuestran diferencias significativas entre los evaluadores experimentados e inexperimentados, en relación con la precisión, exactitud y concordancia (Nita *et al.*, 2003; Bock *et al.*, 2009; Bardsley y Ngugi 2013); sin embargo, tales diferencias no son manifiestas en todos los estudios, porque eventualmente existen evaluadores inexperimentados que, inherentemente, son precisos y exactos, además de evaluadores experimentados que son imprecisos e inexactos (Bock *et al.*, 2013), este tipo de resultados podría tener una explicación a través del análisis del estilo cognitivo de los evaluadores. En el presente estudio, los estimadores de precisión (r) y concordancia (ρ_c) de los evaluadores experimentados estuvieron más cercanos a 1, en comparación con los evaluadores inexperimentados; las diferencias fueron significativas en todos los tipos de hoja y patrones de infección (KS test $P < 0.0001$). El análisis bootstrap, la prueba KS y la prueba t indican que los evaluadores con experiencia hacen evaluaciones de mayor calidad debido a su precisión, lo cual les concede una concordancia significativamente más alta que la de los evaluadores sin experiencia.

Para los patrones de infección de las lesiones, los evaluadores experimentados mostraron una precisión (r) que es significativamente mayor a la de los evaluadores sin experiencia. En general, el proceso de evaluación de PI-AL es menos eficiente que el de PI-AG; es decir, al evaluar imágenes con PI-AL se requirió mayor tiempo y se obtuvo menor concordancia con los valores reales de severidad (Cuadro 1.4).

Estudios previos describen el efecto del número de lesiones en la evaluación visual; sugieren que a mayor número de lesiones se incurre más frecuentemente en sobreestimación en relación con el área infectada (Bock *et al.*, 2008; Bock *et al.*, 2009;

Schwanck y Del Ponte 2014) . Este tipo de sobreestimación ocurre generalmente a bajos niveles de severidad y cuando las lesiones son pequeñas y tienen una distribución aleatoria (Bock *et al.*, 2010). En este estudio, a través del análisis y clasificación del patrón de las lesiones, se estudió el efecto de los patrones agregados y aleatorios, los cuales son predominantes en infecciones naturales. El análisis de la concordancia (ρ_c) mostró valores más cercanos al óptimo (1) cuando las imágenes tuvieron PI-AG en hojas de forma simple (lineal, elíptica y triangular); por el contrario, las imágenes con PI-AL registraron sus niveles más altos de concordancia en hojas de forma compleja (lobulada y palmeada). Lo anterior puede deberse a que en las hojas de forma simple los evaluadores tienen un escenario elemental en el que, al presentarse un PI-AG, con facilidad pueden desarrollar un balance de proporciones por la cercanía de las lesiones; esta tarea se complica con las hojas de forma compleja por lo intrincado de sus bordes. Por otro lado, los resultados indican que la evaluación de imágenes con PI-AL no registró una disminución importante en la calidad de las estimaciones cuando se comparan los resultados de hojas de forma simple contra hojas de forma compleja. El análisis conjunto de la concordancia y el tiempo evidenció que la evaluación de imágenes con PI-AG es un proceso más eficiente que produce estimaciones de mayor calidad con menor inversión de tiempo. Las diferencias entre PI-AG y PI-AL son significativas (95% CIs de la diferencia, *KS* value $P < 0.0001$, *t* - value $P < 0.0001$), este resultado se verificó con todos los tipos de hoja evaluados.

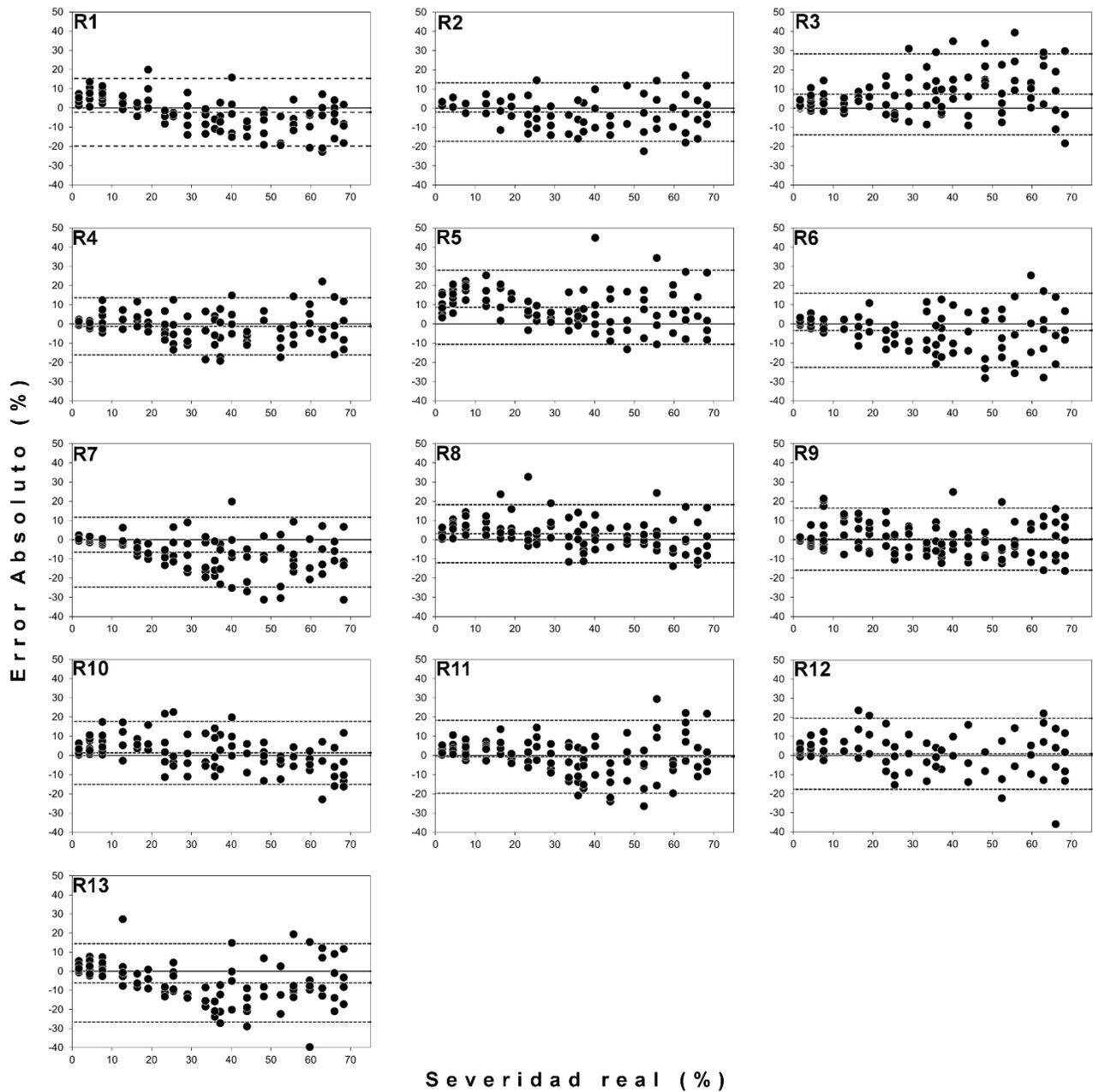


Figura 1.5. Gráficos Bland-Altman. Error absoluto (severidad estimada – severidad real) graficado contra la severidad real. La línea punteada del centro indica la media de los errores; las líneas de los extremos se ubican a dos desviaciones estándar.

En este estudio, el análisis comparativo de los cinco tipos de hoja mostró que la forma de la lámina foliar tiene un efecto directo sobre el sesgo sistemático (v); por ejemplo,

para la hoja lineal, las estimaciones tienen una variabilidad mayor que la asociada con los valores de severidad real; es decir, en la estimación de este tipo de hoja se registraron valores extremos que incrementaron la variabilidad; esos valores extremos fueron tan frecuentes que implicaron un problema de sesgo sistemático. De acuerdo con la Figura 1.2 A y B, tales valores extremos están conglomerados por arriba de la línea de concordancia por lo que el sesgo es por sobreestimación; este fenómeno se observó con mayor frecuencia cuando la severidad real fue mayor a 40 %, además la sobreestimación fue más acentuada para los evaluadores impulsivos. Este tipo de sesgo está asociado con valores de $v > 1$, el cual fue un resultado predominante en este estudio (Cuadro 1.3, 1.4 y 1.5, Figura 1.3 A). En el trigo (hojas de tipo lineal) se obtuvieron resultados similares en cuanto al sesgo sistemático (El Jarroudi *et al.*, 2015; Bock *et al.*, 2015).

Las estimaciones de las hojas elíptica, lobulada y palmeada estuvieron asociadas a un sesgo por subestimación. La razón de este sesgo es que los errores no son constantes en el intervalo de severidad estudiado; particularmente, cuando la severidad real varía en el rango de 30 a 60 %, las estimaciones tienen una variación “limitada” en el rango de 20 a 50 %, por lo que en ese intervalo se origina una acumulación de observaciones asociadas a subestimaciones (errores negativos). Este fenómeno es perceptible en Figura 1.2 C, D, G, H, J, aunque el sesgo por subestimación es más evidente con los evaluadores R1, R5, R7, R11, R13, I3, I5, I6, I7, I8, I9, I10, I12 e I13 en los gráficos Bland-Altman (Figuras 1.5 y 1.6). Un sesgo de esta naturaleza produce valores de $v < 1$, este resultado fue consistente (Cuadro 1.3, 1.4 y 1.5, Figura 1.3 A). El error de estimación no tiene una tendencia o comportamiento predecible; Bock *et al.* (2009) indican que la variabilidad de los errores puede estar en función de la severidad real, aunque esta

relación no ha sido suficientemente estudiada. La acumulación de subestimaciones en el rango de severidad real de 30 a 60%, podría estar asociado a conglomerados de observaciones, conocidos como nudos o *knots* (Bock *et al.*, 2010).

El sesgo de ubicación (u) registró valores positivos consistentemente en las hojas de tipo lineal y triangular. Esto sugiere un sesgo por sobreestimación ligado a la forma de la lámina foliar; sin embargo, el proceso de evaluación de hojas triangulares es menos variable ($v < 1$) que el proceso de evaluación de hojas lineales ($v > 1$) (Cuadro 1.3 y 1.5). La interpretación de los componentes del sesgo (u y v) tiene algunas excepciones, Nita *et al.* (2003) indican que puede ocurrir sobreestimación incluso con valores de u menores que 0; además sugieren que, en algunos casos, los valores de v mayores que 1 indican la existencia de un sesgo que cambia en función de la severidad real.

El análisis de concordancia tiene dos componentes fundamentales: la precisión y la exactitud (Madden *et al.*, 2007; Bock *et al.*, 2010). La precisión es medida por el coeficiente de correlación (r) y la exactitud es estimada a través del sesgo generalizado (C_b), que, a su vez, sintetiza al sesgo sistemático (v) y al de ubicación (u) (Yadav *et al.*, 2013; Costa Lage *et al.*, 2015) . Los resultados del presente estudio indican que el tipo de hoja y el patrón de las lesiones son factores determinantes de la precisión y la exactitud; de esta manera, la concordancia también depende de dichos factores. Por lo anterior, y dado que la estimación visual es el método de evaluación ampliamente usado en la actualidad, resulta necesario considerar al tipo de hoja y al patrón de las lesiones como una fuente de error adicional a las ya conocidas (Nita *et al.*, 2003; Bock *et al.*,

2010). El conocimiento las fuentes de error permite tomar medidas apropiadas para incrementar la exactitud y confiabilidad de las estimaciones (Yadav *et al.*, 2013).

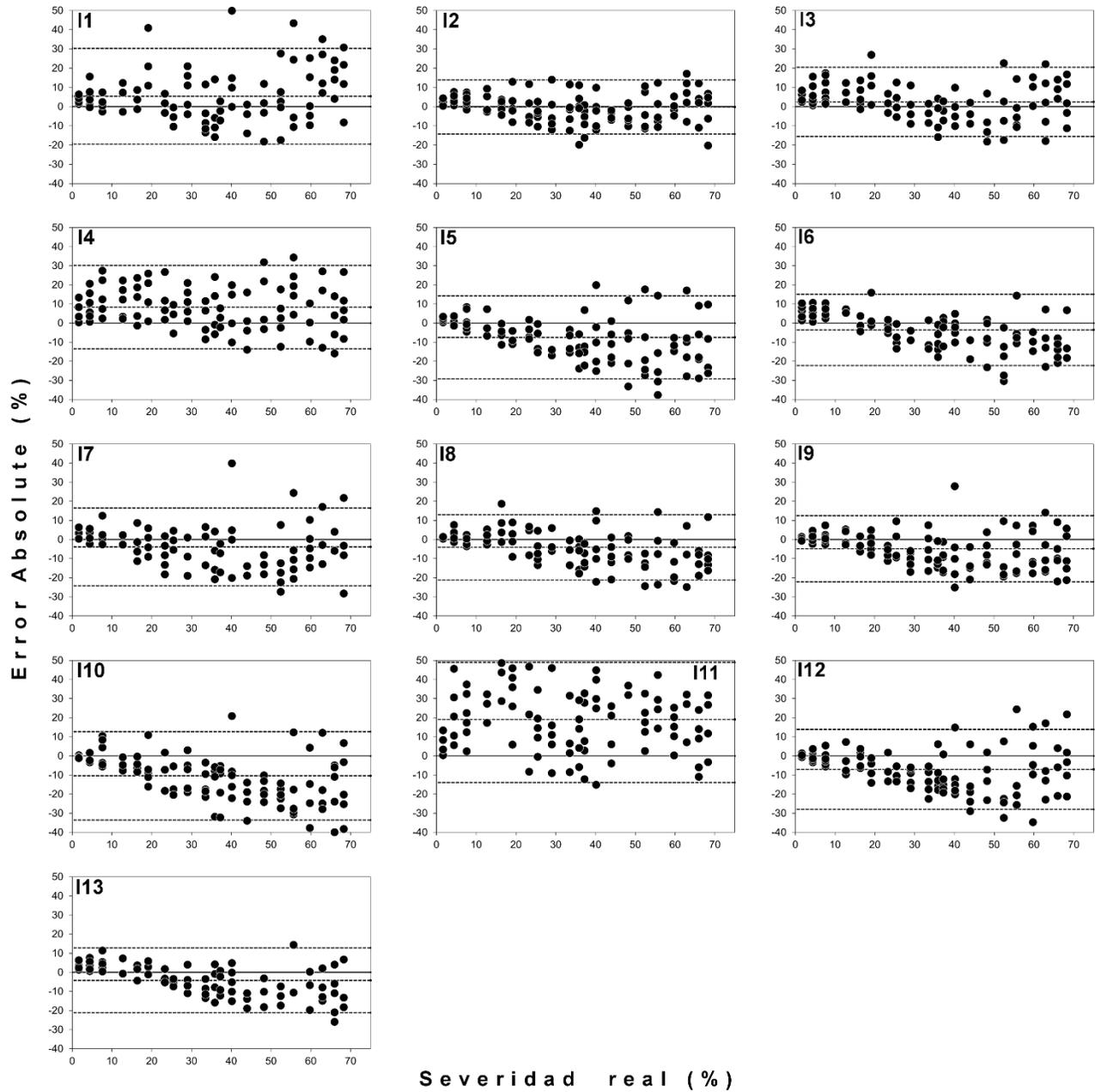


Figura 1.6. gráficos Bland-Altman. Error absoluto (severidad estimada – severidad real) graficado contra la severidad real. La línea punteada del centro indica la media de los errores; las líneas de los extremos se ubican a dos desviaciones estándar.

Los evaluadores reflexivos fueron consistentemente más precisos que los evaluadores impulsivos (Cuadro 1.3); sin embargo; los impulsivos mostraron una mayor confiabilidad intra-evaluador en todos los tipos de hoja y patrones de infección. Adicionalmente, la confiabilidad intra-evaluador estuvo en función del tipo de hoja; las hojas de forma simple (lineal, elíptica y triangular) estuvieron asociadas a un mayor grado de confiabilidad ($0.96 < R^2 < 0.98$), en comparación con las hojas de forma compleja (lobulada y palmeada) ($0.90 < R^2 < 0.95$). De manera complementaria, la confiabilidad inter-evaluadores registró los valores más elevados con los evaluadores reflexivos ($0.80 < \rho < 0.89$), en comparación con los evaluadores impulsivos ($0.70 < \rho < 0.84$). Estos resultados indican que el proceso de evaluación es más heterogéneo entre los evaluadores, y más homogéneo en el proceso cognitivo individual.

Existen diferentes métodos que son utilizados para incrementar la precisión y la exactitud de la estimación visual, incluyendo programas de cómputo para entrenamiento (Yadav *et al.*, 2013). Estos métodos de entrenamiento podrían ser potencializados si se implementa una diversidad de escenarios (tipos de hojas) con variantes en los patrones de infección, tamaño, número y color de las lesiones; cada patosistema podría representar un reto diferente. Diversos estudios indican que el uso de diagramas de área estándar (SADs) incrementa la precisión y la exactitud de la estimación (Spolti *et al.*, 2011; Chiang *et al.*, 2014; Costa Lage *et al.*, 2015), este incremento es más significativo en evaluadores inexperimentados (Yadav *et al.*, 2013). En estudios posteriores se puede analizar la calidad de las estimaciones realizadas con y sin SADs, relacionándola con el estilo cognitivo reflexividad-impulsividad.

En el presente estudio no se solicitó a los evaluadores que efectuara el conteo de lesiones, por lo que no es posible hacer alguna aseveración acerca del efecto de esta variable en la calidad de las estimaciones de la severidad. Estudios posteriores podrían incluir el análisis del número, el tamaño, la distribución y el color de las lesiones.

El comportamiento impulsivo es complejo y tiene tres rasgos fundamentales: un componente motor (actuar sin pensar), un componente atencional (inestabilidad atencional) y un componente de no planificación (falta de planificación) (Gilbert *et al.*, 2011). Todos estos componentes inciden en cierto grado en la precisión de las evaluaciones (r), por lo que este fue el parámetro que marcó la diferencia en el análisis de concordancia. Adicionalmente, la impulsividad podría estar asociada con la conocida tendencia de evaluación a través de puntos preferidos o nudos (knots, *i.e.* 5, 10, 15, ...).

El supuesto de la compensación en la concordancia (ρ_c), como consecuencia de un mayor el tiempo de evaluación, no se cumplió con los evaluadores impulsivos que, en hojas simples, hicieron estimaciones menos concordantes, aun cuando destinaron más tiempo que los evaluadores reflexivos. Los resultados sugieren que el tiempo de evaluación depende de la forma de la hoja y de la distribución de las lesiones; es decir, de la complejidad del tejido evaluado. En general, se concluye que la calidad de las estimaciones está relacionada con el estilo cognitivo de los evaluadores; sin embargo, no se obtuvo una evidencia clara de que un mayor tiempo de evaluación ocasione niveles más altos de concordancia.

El estilo cognitivo que dimensiona la reflexividad y la impulsividad es un factor importante en la evaluación de la severidad. Adicionalmente, se sabe que el ser humano es cognoscitivo, es decir, tiene la capacidad de modificar su estilo cognitivo para adoptar nuevas formas de procesar la información. Lo anterior podría lograrse a través de planes de entrenamiento diseñados para acentuar el estilo reflexivo. La implementación y validación de los planes de entrenamiento deberían ser objeto de estudios posteriores; adicionalmente, es de gran relevancia evaluar otros estilos cognitivos como la dependencia e independencia de campo.

1.6. LITERATURA CITADA

- Bardsley, S. J., and Ngugi, H. K. 2013. Reliability and accuracy of visual methods to quantify severity of foliar bacterial spot symptoms on peach and nectarine. *Plant Pathology*. 62:460–474
- Bland, M. J., and Altman, D. G. 1999. Measuring agreement in method comparison studies. *Statistical methods in medical research*. 8:135–160
- Bock, C. H., El Jarroudi, M., Kouadio, L. A., Mackels, C., Chiang, K. S., and Delfosse, P. 2015. Disease Severity Estimates—Effects of Rater Accuracy and Assessment Methods for Comparing Treatments. *Plant Disease*. 99:1104–1112
- Bock, C. H., Hotchkiss, M. W., and Wood, B. W. 2015. Assessing disease severity: accuracy and reliability of rater estimates in relation to number of diagrams in a standard area diagram set. *Plant Pathology*. DOI:10.1111/ppa.12403
- Bock, C. H., Parker, P. E., Cook, A. Z., and Gottwald, T. R. 2008. Characteristics of the perception of different severity measures of citrus canker and the relationships between the various symptom types. *Plant Disease*. 92:927–939
- Bock, C. H., Parker, P. E., Cook, A. Z., Riley, T., and Gottwald, T. R. 2009. Comparison of assessment of citrus canker foliar symptoms by experienced and inexperienced raters. *Plant Disease*. 93:412–424
- Bock, C. H., Poole, G. H., Parker, P. E., and 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., Den Bosch, F. Van, Parnell, S., and 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
- Bock, C. H., Wood, B. W., and Gottwald, T. R. 2013. Pecan Scab Severity-Effects of Assessment Methods. *Plant Disease*. 97:675–684
- Braido, R., Gonçalves-Zuliani, A. M., Janeiro, V., Carvalho, S. A., Belasque, J. J., Bock, C. H., *et al.* 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
- Buela, G., and Riaño, D. 2015. Conceptualización y evaluación de la impulsividad en adolescentes: una revisión sistemática. *Universitas Psychologica*. 14:1077–1090
- Cairns, ED, and Cammock, T. 1978. Development of a more reliable version of the Matching Familiar Figures Test. *Developmental Psychology*. 14:555 Available at: <http://psycnet.apa.org/journals/dev/14/5/555/>.
- Chiang, K. S., Liu, S. H., Bock, C. H., and Gottwald, T. R. 2014. What interval characteristics make a good categorical disease assessment scale? *Phytopathology*. 104:575–85
- Costa Lage, D. A., Marouelli, W. A., Duarte, H. S., and Café-Filho, A. C. 2015. Standard area diagrams for assessment of powdery mildew severity on tomato leaves and leaflets. *Crop Protection*. 67:26–34
- El Jarroudi, M., Kouadio, A. L., Mackels, C., Tychon, B., Delfosse, P., and Bock, C. H. 2015. A comparison between visual estimates and image analysis

measurements to determine septoria leaf blotch severity in winter wheat. *Plant Pathology*. 64:355–364

- Gilbert, K. E., Kalmar, J. H., Womer, F. Y., Markovich, P. J., Pittman, B., Hoeksema, S., *et al.* 2011. Impulsivity in adolescent bipolar disorder. *Acta Neuropsychiatrica*. 23:57–61
- Kagan, J. 1965. Reflection-impulsivity and reading ability in primary grade children. *Child development*. Available at: <http://www.jstor.org/stable/1126908>.
- Kertzman, S., Vainder, M., Vishne, T., Aizer, A., Kotler, M., and Dannon, P. N. 2009. Speed-accuracy tradeoff in decision-making performance among pathological gamblers. *European Addiction Research*. 16:23–30
- Lin, L. I. 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 45:255–268
- Madden, L. V., Hughes, G., and Den Bosch, F. Van. 2007. *The Study of Plant Disease Epidemics*. St. Paul, MN.: American Phytopathological Society.
- Morisita, M. 1959. Measuring of the dispersion of individuals and analysis of the distributional patterns. *Mem. Fac. Sci. Kyushu Univ. Ser. E*. 2:5–235
- Nita, M., Ellis, M. A., and Madden, L. V. 2003. Reliability and accuracy of visual estimation of phomopsis leaf blight of strawberry. *Phytopathology*. 93:995–1005
- Nutter Jr, F. W., Gleason, M. L., Jenco, J. H., and Christians, N. C. 1993. Assessing the accuracy, intra-rater repeatability, and inter-rater reliability of disease assessment systems. *Phytopathology*. 83:806–812
- Parker, SR, Whelan, MJ, and Royle, DJ. 1995. Reliable measurement of disease severity. *Aspects of Applied Biology* (United Available at:

<http://agris.fao.org/agris-search/search.do?recordID=GB9632690>.

- Quednow, B. B., Kühn, K. U., Hoppe, C., Westheide, J., Maier, W., Daum, I., *et al.* 2007. Elevated impulsivity and impaired decision-making cognition in heavy users of MDMA (“Ecstasy”). *Psychopharmacology*. 189:517–530
- Schwanck, A. A., and Ponte, E. M. Del. 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
- Spolti, P., Schneider, L., Sanhueza, R. M., Batzer, J. C., Gleason, M. L., and Ponte, E. 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 Available at: <http://link.springer.com/article/10.1007/s10658-010-9636-7>.
- Yadav, N. V., Vos, S. M. de, Bock, C. H., and 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

CAPITULO 2.

EVALUACIÓN VISUAL DE SEVERIDAD, UN ANÁLISIS COMPARATIVO DE ESCALAS DE PROPORCIÓN Y CATEGÓRICAS CON INTERVALOS LOGARÍTMICOS Y LINEALES

RESUMEN

Los diagramas de área estándar (SAD), también conocidos como escalas diagramáticas, tienen el objetivo incrementar la precisión, exactitud y confiabilidad de las evaluaciones visuales de severidad. Los SADs varían en cuanto al tipo (escalas de proporción y categóricas), el intervalo de incremento de severidad (logarítmico y lineal) y el número de diagramas. En la actualidad, la mayoría de los SADs son desarrollados con base en la ley de agudeza visual de Weber-Fechner (incrementos logarítmicos), aun cuando se ha refutado la ley de Fechner. En el presente estudio se hizo un análisis comparativo de la exactitud, precisión, concordancia y confiabilidad de las estimaciones de severidad efectuadas con escalas de proporción y categóricas, ambas desarrolladas con SADs; adicionalmente, se evaluó el efecto del número de diagramas y del intervalo de incremento (logarítmico y lineal). Las estimaciones visuales se compararon con los valores reales obtenidos con análisis de imágenes; el nivel de concordancia entre los enfoques se calculó con la correlación de concordancia de Lin (LCC, ρ_c); la confiabilidad inter-evaluadores se midió a través del coeficiente de correlación intra-clase (ρ). El efecto del tipo de escala, intervalo y número de diagramas se estudió a través de análisis bootstrap con 2 000 muestras.

El análisis de concordancia evidenció que los SADs con intervalos lineales proporcionan más exactitud, precisión y confiabilidad que los SADs con intervalos logarítmicos ($P<0.0001$). Además, las escalas de proporción producen estimaciones significativamente más concordantes que las escalas categóricas ($P<0.0001$). En este estudio, los instrumentos de evaluación más precisos, exactos y confiables fueron las escalas de proporción con SADs a intervalos lineales; para este tipo de instrumento, el número de diagramas puede variar entre 3 y 7 ya que sus niveles de concordancia son similares ($\rho_c>0.95$).

2.2. INTRODUCCIÓN

La estimación visual de severidad es el método más utilizado en Fitopatometría. Su campo de aplicación es el más vasto en agricultura protegida y en campo abierto; su implementación sigue vigente, a pesar de los avances tecnológicos en el análisis computarizado de imágenes que desempeñan una función similar (Bock *et al.*, 2010). La estimación visual es un proceso que no produce siempre los mismos resultados, esto depende de varios factores como la forma de la hoja, el color, la distribución y tamaño de las lesiones, además de la habilidad cognitiva de cada individuo (Bock *et al.*, 2010; Bardsley y Ngugi 2013). En los últimos 30 años se han registrado grandes progresos en la comprensión de los factores que afectan la calidad de la estimación visual de enfermedades; esta información se ha capitalizado para desarrollar “instrumentos de medición” como los diagramas de área estándar, que mejoran la capacidad de evaluación; sin embargo, el proceso de comprensión de este fenómeno continúa (Bock *et al.*, 2015; Bock *et al.*, 2010).

Existen diversas escalas para evaluar visualmente la severidad, algunas de ellas tienen una aplicación histórica tan arraigada que, a pesar de que se ha demostrado su limitada aportación a la precisión y exactitud, continúan siendo objeto de estudio en diversas regiones del mundo (Bock *et al.*, 2009). En contraste, las escalas de publicación reciente deben justificar su funcionalidad a través del incremento en la exactitud, la confiabilidad y la concordancia de las estimaciones (Del Ponte *et al.*, 2017). De acuerdo con Bock *et al.* (2010), las escalas más utilizadas en la evaluación visual de enfermedades son:

1. Escalas nominales. Estas son las escalas más simples y subjetivas. Las unidades de evaluación (hojas, folíolos, frutos, etc.) se clasifican en una serie de clases con términos que describen la intensidad de la enfermedad de la siguiente manera: "leve", "moderado" o "grave". Debido a la subjetividad involucrada y la falta de definición cuantitativa, estas escalas tienen una aplicación muy limitada (Nutter y Esker, 2006).

2. Escalas ordinales. Estas son escalas descriptivas de la enfermedad, con clases ordenadas que representan severidades cada vez mayores. Este tipo de instrumentos ordinales también son subjetivos, por ejemplo: 1 = poca enfermedad, 2 = enfermedad media y 3 = mucha enfermedad.

Las escalas ordinales tienen un campo de aplicación en ciertos grupos específicos de patógenos; destaca su uso en enfermedades virales, donde los síntomas no se pueden cuantificar fácilmente (Bock *et al.*, 2010).

3. Escalas categóricas. Este tipo de instrumento es también conocido como escala de intervalo, y está compuesto por una serie de categorías asociadas con

intervalos de porcentajes de severidad. En su desarrollo inicial, estas escalas únicamente presentaban una lista de las categorías y su porcentaje asociado; posteriormente, se implementó el uso de diagramas representativos de los porcentajes de severidad. Para el uso de este tipo de instrumentos, las unidades de evaluación se clasifican con base en las categorías de la escala.

4. Escalas de proporción. Estas escalas tienen una serie de clases asociadas con valores puntuales de porcentaje de severidad. Para su desarrollo, en la mayoría de los casos, se emplean diagramas representativos. Las unidades de evaluación se comparan con los diagramas, y el evaluador hace una estimación de la severidad por extrapolación o interpolación.

Los diagramas de área estándar (SAD), también conocidos como escalas diagramáticas, son arreglos ordenados de imágenes de plantas, o de alguno de sus órganos (hojas, folíolos, tallos o frutos), que en conjunto tienen el objetivo de representar los distintos niveles de severidad de una enfermedad (Figura 2.1). El desarrollo de los SADs está enfocado al robustecimiento de las escalas categóricas y de proporción (Schwanck y Del Ponte 2014). Los SADs se han utilizado desde hace más de 120 años con el objetivo de incrementar la exactitud y la confiabilidad de las evaluaciones; de acuerdo con una revisión exhaustiva que realizaron Del Ponte *et al.* (2017), en el periodo de 1990 a 2016 se han desarrollado más de 105 investigaciones que derivaron en 127 SADs para la evaluación de más de 100 enfermedades en 48 cultivos.

La escala de Cobb fue el primer SAD utilizado en fitopatología (Cobb, 1892). Para el uso de esta escala categórica, el evaluador comparaba las unidades de evaluación (hojas)

con los cinco diagramas que conformaban el SAD; cada hoja se clasificaba en relación con el diagrama más semejante. Esta escala representó un avance tecnológico que abrió un nuevo panorama en la investigación fitosanitaria, al crear un procedimiento que estandarizó la evaluación de la severidad. Desde entonces, muchos investigadores han desarrollado y aplicado SADs para incrementar la exactitud y confiabilidad de las estimaciones (Bock *et al.*, 2010).

Los primeros SADs estaban compuestos por ilustraciones en blanco y negro, diseñados regularmente con 5 a 8 diagramas a incrementos no lineales; en la actualidad, las escalas están conformadas por fotografías o representaciones estilizadas a color, el número de diagramas ha disminuido (4 a 7) y los incrementos pueden ser logarítmicos o lineales (Del Ponte *et al.*, 2017). Costa Lage *et al.* (2015) destacan las bondades de los SADs en blanco y negro, ya que por su alto contraste estos colores son idóneos para representar los tejidos sanos y enfermos (Figura 2.1); además, Schwanck y Del Ponte (2014) demostraron que la exactitud y confiabilidad no se reducen de manera significativa al utilizar SADs en blanco y negro, en comparación con SADs a color.

Para el uso de escalas de proporción con SADs, los evaluadores son instruidos para que, con base en los diagramas, hagan una estimación del porcentaje de severidad a través de interpolación o extrapolación (Yadav *et al.*, 2013). Este modo de uso es conocido como Ayuda Visual (AV), y las escalas desarrolladas con este fin presentan diagramas con números puntuales que indican el porcentaje de severidad representado (Figura 2.1).

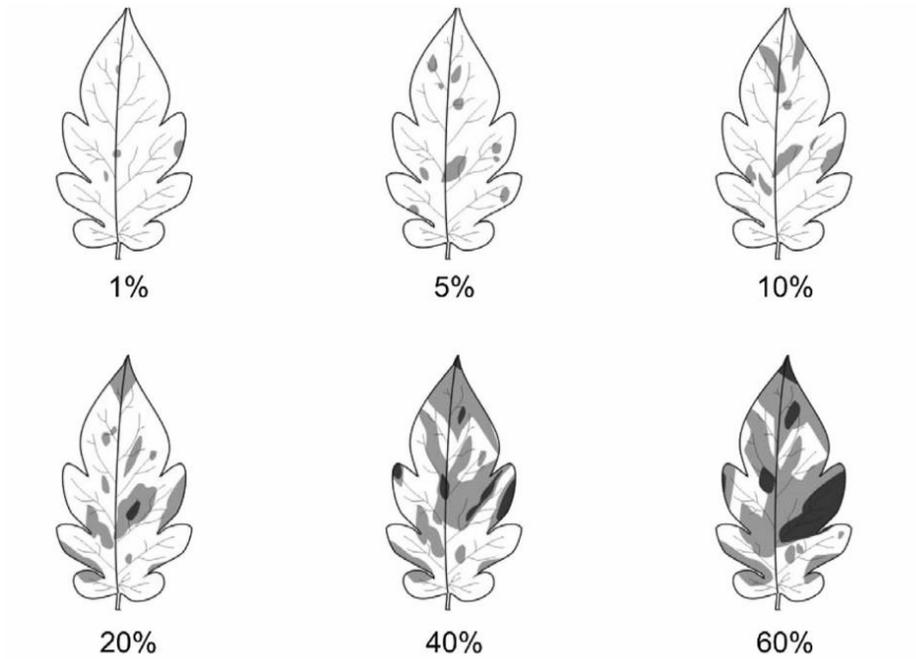


Figura 2.1. Escala de proporción con SAD. Tomada de Costa Lage *et al.* (2015)

Para el uso de escalas categóricas con SADs, los evaluadores comparan las unidades de evaluación (hojas) con los diagramas del SAD y clasifican en relación con el diagrama más semejante; este modo de uso será referido como CLAS, para los fines de este estudio. Los SADs de escalas categóricas presentan diagramas acompañados por un número de clase y un intervalo de severidad (Figuras 2.4 y 2.5); para el análisis de los datos es necesario convertir los números de clase registrados, a los puntos medios de los respectivos intervalos (Madden *et al.*, 2007)

En la actualidad, diversos autores sugieren que los SAD deben diseñarse con diagramas claramente distinguibles por el ojo del evaluador (Chiang *et al.*, 2014; Bock *et al.*, 2010; Schwanck y Del Ponte 2014). Una recomendación en el mismo sentido fue manifiesta hace más de 70 años por Horsfall y Barratt (1945), quienes desarrollaron la escala categórica más utilizada en la historia de la Fitopatología, la cual es conocida como escala Horsfall-Barratt o HB. Este instrumento se caracteriza porque sus intervalos de incremento de severidad son logarítmicos, de acuerdo con la ley de agudeza visual de Weber-Fechner. Al respecto, Nutter y Esker (2006) estimaron la "diferencia apenas perceptible" (JND) entre dos estímulos (hojas, una de referencia y otra de comparación), en la que el evaluador debía calificar a las hojas como iguales o diferentes en relación con su porcentaje de severidad. De esta manera, los autores demostraron que la JND dividida entre el estímulo de referencia es una fracción constante, tal como lo afirma la ley de Weber; y no es proporcional a la magnitud del estímulo de referencia, como lo enuncia la ley de Fechner. Por lo anterior, se concluyó que la capacidad del ojo humano para discernir diferencias en los niveles de severidad es mejor de lo que supusieron Horsfall y Barratt (1945).

La escala HB tiene intervalos de amplitud simétrica a ambos lados del 50% (Cuadro 2.1), esta simetría se basa en el siguiente supuesto: cuando la severidad es menor que 50%, el ojo del evaluador analiza en el tejido enfermo; en contraste, cuando la severidad es mayor a 50%, el ojo del evaluador analiza en el tejido sano. De esta manera, Horsfall y Barratt (1945) sugirieron que, de forma natural, el ojo humano identifica el tejido menos abundante y lo toma como base para el análisis de proporciones (estimación de severidad). Por su parte, Bock *et al.* (2010) afirman que no hay elementos científicos que

sustenten tal hipótesis, por lo que las escalas de evaluación pueden diseñarse sin ser simétricas con relación al 50%.

De acuerdo con el análisis cuantitativo realizado por Del Ponte *et al.* (2017), los SADs de publicaciones científicas varían en cuanto al número de diagramas y los intervalos de incremento de severidad; además, indican que es frecuente que los investigadores definan los intervalos de incremento con base en la ley de agudeza visual de Weber-Fechner (incrementos logarítmicos), aun cuando se ha refutado la ley de Fechner; Bock *et al.* (2010) desarrollaron un análisis de las consecuencias del uso de la escala HB y relacionadas, a través del cual sugirieron que los principios de la escala HB no deberían ser utilizados para definir los intervalos de incremento de severidad en los SADs. Nita *et al.* (2003) sugieren que los intervalos de incremento de severidad sean lineales y recomiendan añadir diagramas adicionales en puntos críticos en los que frecuentemente se registra sesgo; por ejemplo, en el intervalo 0 a 10% de severidad en el que las evaluaciones tienden a sobreestimar.

Los diagramas de área estándar pueden desarrollarse de diversas maneras, a través de dibujos, patrones digitales y fotografías. El programa SEVERITY.PRO tuvo su mayor difusión en la década 1990, se empleó para entrenamiento y para el desarrollo de SADs, sus principales funcionalidades permiten elegir el tipo de hoja, síntoma y tamaño de las lesiones (Nutter y Worawitlikit, 1989). Por su parte, el programa 2LOG (Osada y Mora, 1997; Mora *et al.*, 2003) es único en su tipo ya que está especialmente diseñado para calcular intervalos logarítmicos con base en la escala HB.

Para el análisis de la severidad en imágenes existen programas especializados como APS-Assess®, QUANT o medidores de área foliar LI-COR®; además, se ha registrado un aumento en el uso de *software* libre como los analizadores ImageJ® e Image Tool® (Del Ponte *et al.*, 2017). Incluso editores de imágenes como GIMP® o Photoshop® se han empleado para el desarrollo de SADs, ya que tienen funciones como: selección, posterización, segmentación y cuantificación de áreas.

Los SADs con un número reducido de diagramas permiten un uso fluido al momento de la evaluación (Chiang *et al.*, 2014; Bock *et al.*, 2010). Por otro lado, Bock *et al.* (2015) registraron un incremento en la exactitud de evaluaciones con SADs, en comparación con la evaluación sin SADs; pero tal incremento no tuvo una relación clara determinada por el número de diagramas. Adicionalmente, Chiang *et al.* (2014) hicieron un análisis comparativo de escalas categóricas lineales con incrementos de 5 y 10%, las cuales tuvieron un desempeño similar al de la evaluación visual sin escalas. Por su parte, Block *et al.* (2016) demostraron que las escalas categóricas afectan la precisión de las estimaciones en ciertas condiciones, y con frecuencia conduce a errores de Tipo II (no rechazar la hipótesis nula cuando es falsa).

La evaluación de la severidad tiene gran relevancia en la investigación fitosanitaria, Campbell y Neher (1994) indican que la medición de la enfermedad es *la piedra angular del análisis epidemiológico*. En tanto se logren mediciones/estimaciones más exactas, menor será la magnitud de los errores en el análisis y, por tanto, menor será el riesgo de tomar decisiones incorrectas (Block *et al.*, 2016).

En el presente estudio se hizo un análisis comparativo de la exactitud, precisión, concordancia y confiabilidad de las estimaciones de severidad efectuadas con escalas de proporción y categóricas, ambas desarrolladas con SADs. Adicionalmente, se evaluó el efecto del número de diagramas y de los intervalos de incremento de severidad (logarítmicos y lineales).

Los objetivos de este estudio fueron:

- iv) Comparar la exactitud, precisión, concordancia y confiabilidad de las estimaciones visuales de severidad efectuadas con SADs de escalas categóricas y de proporción.
- v) Comparar el efecto de los SADS con intervalos de incremento de severidad logarítmicos y lineales.
- vi) Establecer un número óptimo de diagramas para SADS de escalas categóricas y de proporción.

2.3. MATERIALES Y MÉTODOS

2.3.1. Obtención de las hojas

Se obtuvieron hojas estilizadas en color. Para obtener la hoja patrón (PL) se colectó una muestra de 20 hojas de aguacatero, completamente desarrolladas, las cuales se digitalizaron para la medición del largo y ancho con la regla calibrada del programa ImageJ (NIH Image). Con estos datos se construyeron histogramas y se seleccionó como

patrón a la hoja que tenía las dimensiones más similares a las medias de la muestra. La PL se vectorizó usando el GNU Image Manipulation Program (GIMP).

Los patrones de infección (PI) se obtuvieron de hojas de distintos cultivos infectados naturalmente, la muestra de tejido vegetal se conformó incluyendo la diversidad de patrones de lesiones e intensidades que ocurren en el campo. Las hojas infectadas se escanearon y posteriormente se procesaron con GIMP; en cada imagen se eliminó el tejido sano y el fondo, conservando únicamente las lesiones y su distribución, las imágenes resultantes se denominaron PI. A partir de cada PI se crearon pinceles (archivos GBR) que se usan de manera similar a un sello en el programa GIMP. En total se obtuvieron 320 PI.

A partir de la PL y los pinceles de PI, se generaron 300 imágenes de hojas con severidad entre 0 y 100 %; los niveles de severidad que se representaron en las imágenes se definieron generando números aleatorios dentro del intervalo de 0 a 100%. Para generar el valor exacto del porcentaje de severidad requerido, en cada imagen se probaron distintos pinceles de PI hasta encontrar el que produjera el valor de severidad más aproximado, posteriormente se ajustó aumentando o reduciendo el tamaño de una lesión. El porcentaje de severidad real se calculó usando GIMP, mediante el conteo de píxeles de tejido sano y enfermo. De las 300 imágenes generadas, se seleccionaron las hojas que, por su porcentaje de severidad, fueron propicias para ser incluidas en algún SAD; las hojas restantes se utilizaron para la validación y el análisis comparativo de las escalas. Para entrenamiento de los evaluadores sin experiencia, se generaron 30 imágenes adicionales, con el procedimiento descrito anteriormente.

En la Figura 2.2 se presenta un ejemplo de las imágenes generadas, las cuales representaron la diversidad de patrones de infección que se encuentran comúnmente en la naturaleza.

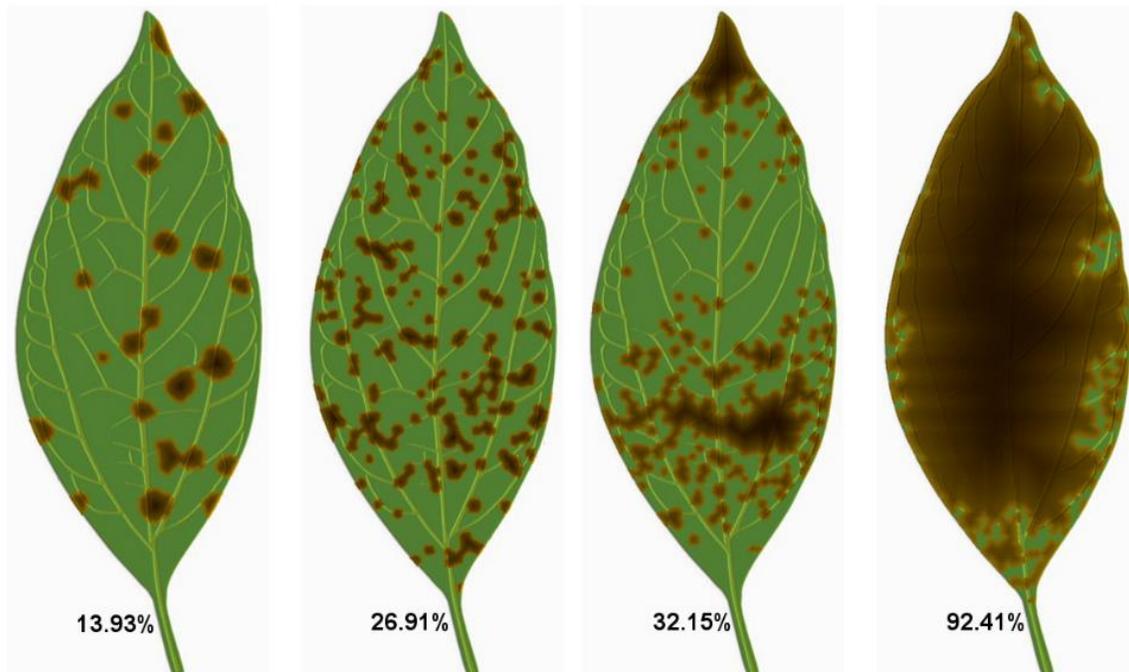


Figura 2.2. Hojas estilizadas a partir de una hoja patrón (PL) y patrones de infección (PI) obtenidos de cultivos infectados naturalmente.

2.3.2. Desarrollo de las escalas

Para obtener los intervalos logarítmicos de incremento de severidad, se siguió el procedimiento descrito a continuación:

1. Se identificó a la hoja con el máximo porcentaje de severidad (s_{max}), en este caso fue 95.57 %.
2. El valor s_{max} se graficó en un diagrama de dispersión en el programa Excel (Microsoft Corp). En este gráfico, el eje vertical se convirtió a escala logarítmica base 10.
3. Se trazó una línea recta (lr) entre el origen y el valor s_{max} .
4. El eje horizontal se dividió en k segmentos iguales, donde k es el número de intervalos requeridos. En cada división se trazó una línea vertical (lv)
5. Las intercepciones entre la línea lr y las líneas lv son los intervalos logarítmicos de incremento de severidad.
6. Para validar los puntos, se hizo un diagrama de dispersión y se ajustó un modelo exponencial (Figura 2.3).

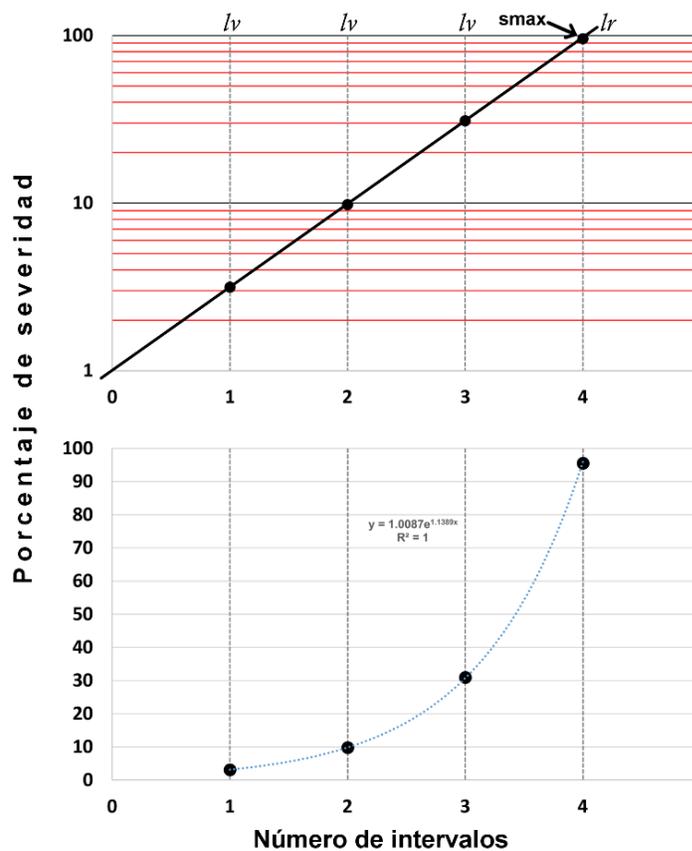


Figura 2.3. Determinación de los intervalos logarítmicos de incremento de severidad.

Por su parte, para obtener los intervalos lineales de incremento de severidad se dividió el valor smax entre el número de intervalos requeridos.

Para desarrollar los SADs, se consideraron como representativas las hojas con severidad más próxima a los puntos medios de los intervalos; en las escalas categóricas cada diagrama se identificó con un número de clase y un intervalo de severidad, mientras que en las escalas de proporción cada diagrama se identificó por el porcentaje de

severidad representado en la hoja. En todos los SADs se incluyó un diagrama de una hoja sana, para referencia del evaluador.

En total se diseñaron 20 SADs, de los cuales 10 fueron escalas categóricas y 10 escalas de proporción; en ambos casos, se obtuvieron 5 escalas con intervalos logarítmicos y 5 con intervalos lineales. Además, en cada combinación se desarrollaron SADs con 3, 4, 5, 6 y 7 diagramas.

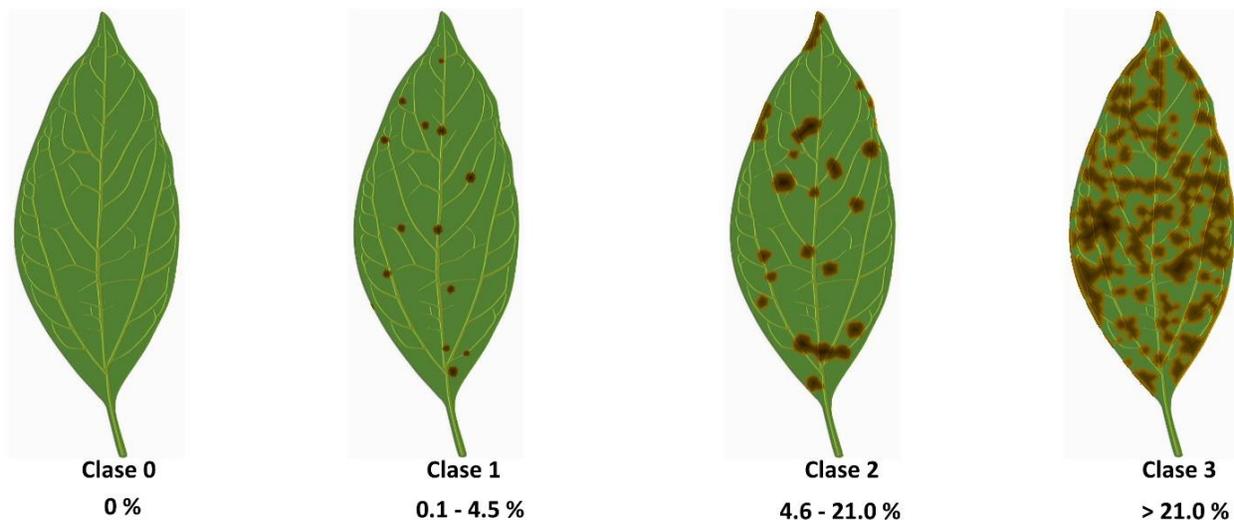


Figura 2.4. Escala categórica con SAD de tres clases a intervalos logarítmicos

2.3.3. Estimación visual de la severidad

Para la evaluación visual, se seleccionaron 140 imágenes con severidad entre 0 y 100%; este conjunto se dividió en dos colecciones con 70 imágenes cada una. Las colecciones se consideran equivalentes porque contienen imágenes con porcentajes de severidad similares. Estas colecciones se evaluaron alternadamente durante 21 sesiones

realizadas a intervalos semanales. Para cada sesión, las imágenes se ordenaron aleatoriamente en una presentación con 70 diapositivas; además, se giraron horizontalmente (efecto espejo), para reducir la posible memorización de los participantes. Las imágenes se desplegaron en un monitor LCD de 23 pulgadas, conservando las proporciones reales del tejido vegetal representado.

La estimación visual se realizó por ocho evaluadores, los cuales tienen formación agronómica y están familiarizados con síntomas de enfermedades en plantas. Del grupo de evaluadores, cuatro tienen experiencia en evaluación visual de severidad, el resto recibió entrenamiento usando el conjunto de 30 imágenes generadas con ese objetivo. El entrenamiento se efectuó una semana antes del inicio de las evaluaciones.

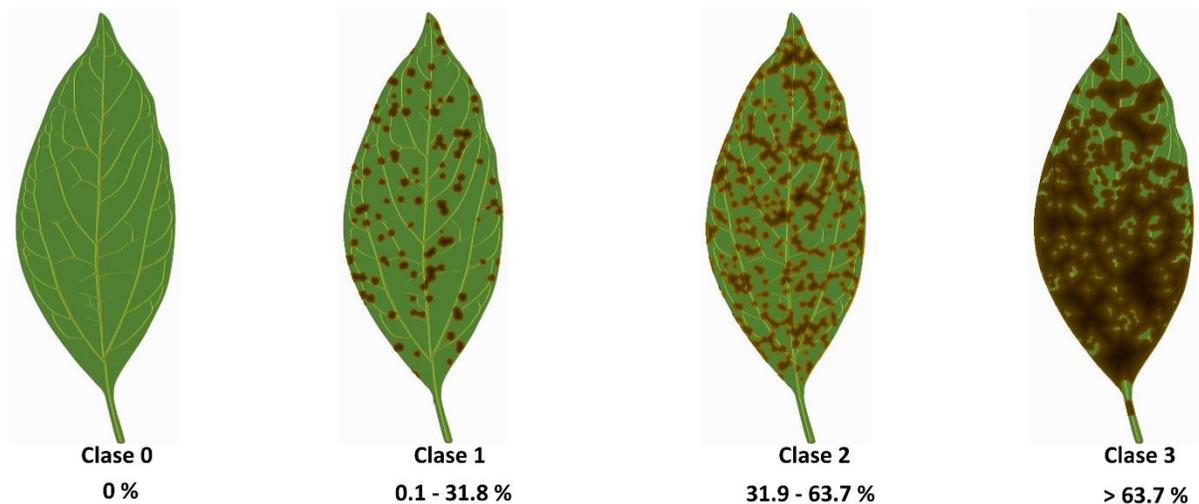


Figura 2.5. Escala categórica con SAD de tres clases a intervalos lineales

Las evaluaciones se realizaron en el periodo de febrero a agosto de 2016; de manera individual, los participantes evaluaron la severidad de 70 hojas cada semana, utilizando una escala diferente para cada sesión. Durante la primera sesión se hizo una evaluación

visual directa (EVD) sin escalas. Para las sesiones posteriores, a los participantes se les instruyó para que hicieran la evaluación visual con escalas, de la siguiente manera:

Escala categórica con SAD: Comparar cada una de las hojas en evaluación con el SAD y asignar el valor de clase correspondiente al diagrama más similar.

Escala de proporción con SAD: Comparar cada una de las hojas en evaluación con el SAD y estimar un porcentaje de severidad a través de la interpolación o extrapolación de los valores representados en los diagramas.

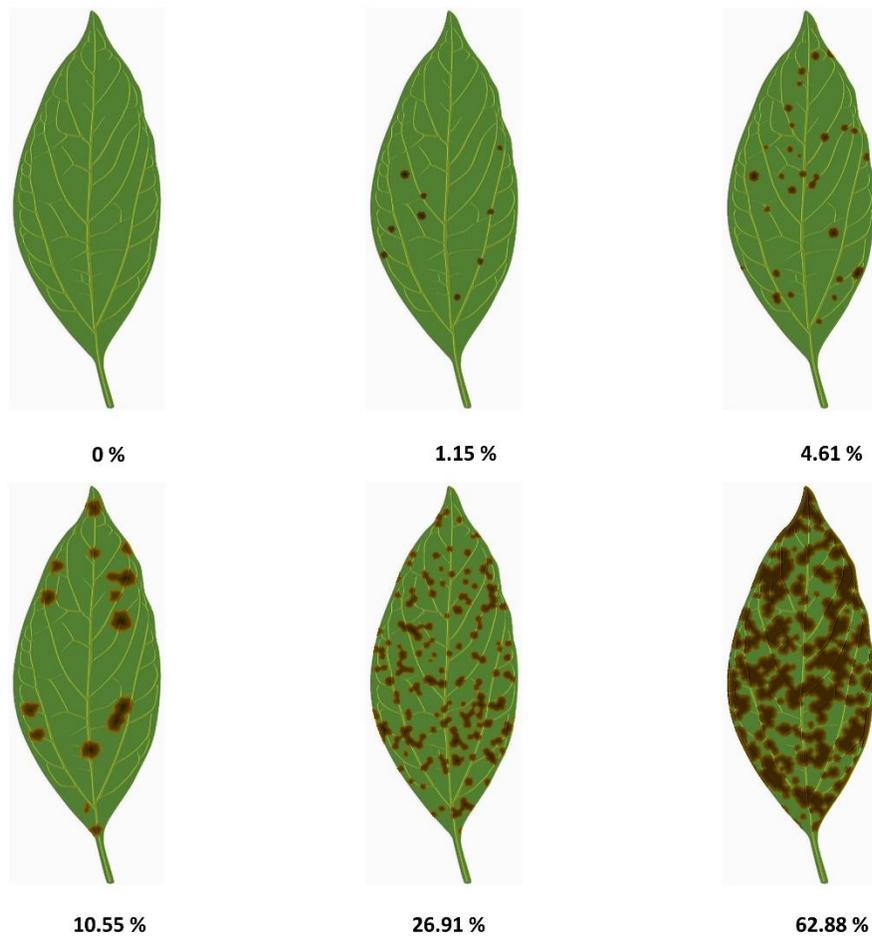


Figura 2.6. Escala de proporción con SAD de cinco clases a intervalos logarítmicos

Las variables registradas fueron: porcentaje de severidad estimado (escalas de proporción, AV) y número de clase (escalas categóricas, CLAS). Para la EVD también se registró la variable porcentaje de severidad estimado; adicionalmente, los valores de EVD se transformaron a puntos medios de la escala HB (Cuadro 2.1), para fines comparativos.

Cuadro 2.1. Escala logarítmica de Horsfall - Barrat para evaluación de severidad

Calificación	Intervalo porcentual	Punto medio
0	0	0.0
1	(0 - 3)	1.5
2	[3 - 6)	4.5
3	[6 - 12)	9.0
4	[12 - 25)	18.5
5	[25 - 50)	37.5
6	[50 - 75)	62.5
7	[75 - 88)	81.5
8	[88 - 94)	91.0
9	[94 - 97)	95.5
10	[97 - 100)	98.5
11	100	100

2.3.4. Análisis de datos.

El análisis de los datos se hizo con el programa R (R Core Team, 2016) y con Excel 2016 (Microsoft Corp.). Los parámetros que miden la confiabilidad, precisión, sesgo y concordancia se calcularon para hacer comparaciones entre los SADs y sus tipos de intervalos de incremento (logarítmico - lineal), los modos de uso (AV-CLAS) y el número de diagramas (3, 4, 5, 6 y 7).

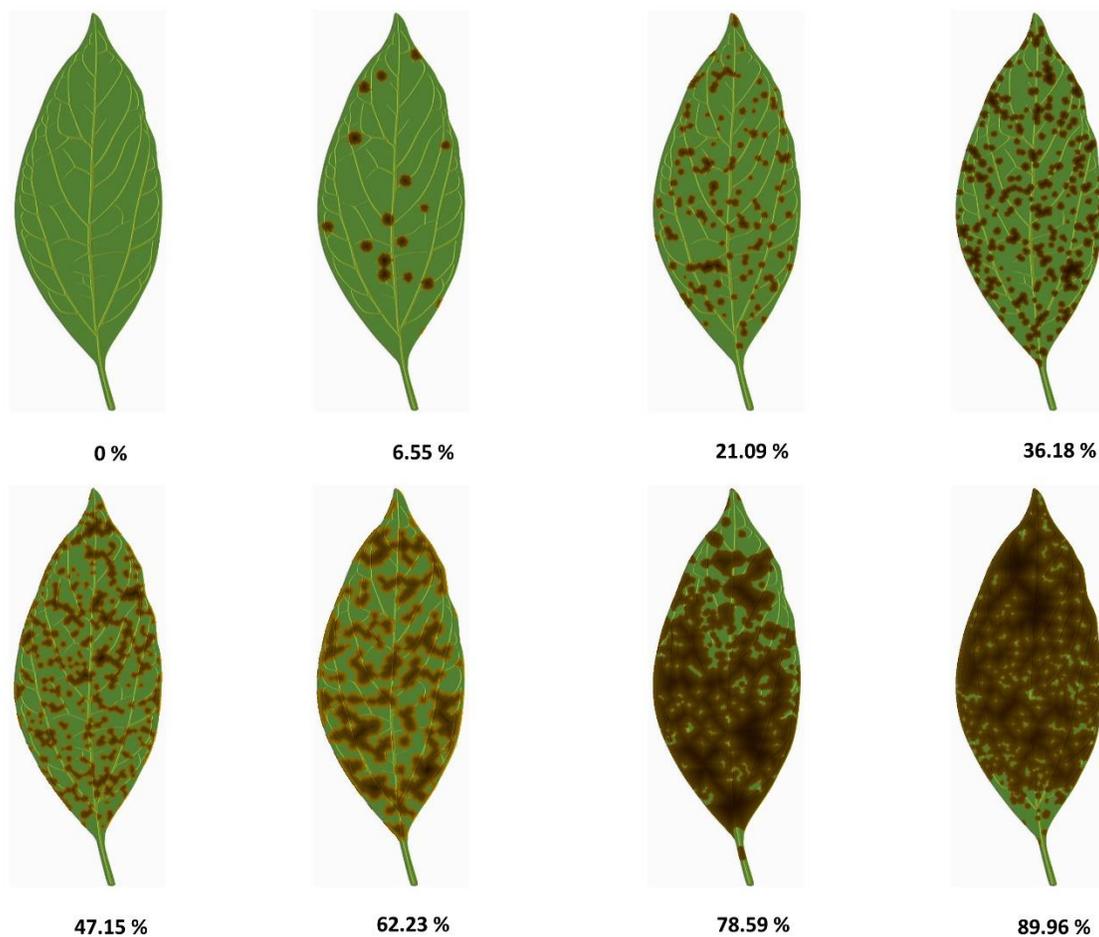


Figura 2.7. Escala de proporción con SAD de siete clases a intervalos lineales.

El nivel de concordancia de las estimaciones se obtuvo con la correlación de concordancia de Lin (Lin 1989; LCC). Con este método se calcula el grado en que los pares de observaciones (severidad estimada y severidad real) se aproximan a la línea de concordancia de 45° (pendiente=1, intercepto=0, i.e. concordancia perfecta) (El Jarroudi *et al.*, 2015). El análisis LCC considera a los componentes esenciales de la concordancia (precisión y exactitud), y de su cálculo se derivan varios parámetros que miden el sesgo constante o de localización (μ), sesgo sistemático (ν), sesgo generalizado (C_b) y concordancia (ρ_c). El parámetro ρ_c es el Coeficiente de Correlación de

Concordancia de Lin (LCCC) y toma valores entre 0 y 1. Adicionalmente, como medida de la precisión, se calculó el coeficiente de correlación (r).

El LCCC es el producto del sesgo generalizado (C_b) y el coeficiente de correlación (r). Los cálculos se hicieron como se ha descrito en el capítulo previo de este documento, siguiendo la metodología de trabajos publicados (Lin, 1989; Nita *et al.*, 2003; Madden *et al.*, 2007; Yadav *et al.*, 2013; El Jarroudi *et al.*, 2015). El máximo nivel de concordancia se obtiene cuando los pares de observaciones (severidad estimada y severidad real) están sobre la línea de concordancia (i.e., $\mu = 0$, $v = 1$, $C_b = 1$, $r = 1$, lo que implica $\rho_c = 1$). Cualquier desviación de esos valores indica sesgo, imprecisión y pérdida de exactitud (Lin, 1989; Madden *et al.*, 2007; Bock *et al.*, 2015).

La confiabilidad inter-evaluadores se midió a través del coeficiente de correlación intra-clase (ρ); para lo anterior, la base de datos se subdividió por tipo de intervalo, modo de uso y número de diagramas. Se realizó un análisis de cada conjunto de datos a través de un ANOVA de dos vías con efectos aleatorios, tal como lo describen Nita *et al.* (2003). El coeficiente de correlación intra-clase (ρ) se obtuvo a través de los componentes de la varianza σ_{sev}^2 , σ_{eval}^2 y σ_{error}^2 ; de tal manera que $\rho = \sigma_{sev}^2 / (\sigma_{sev}^2 + \sigma_{eval}^2 + \sigma_{error}^2)$. Los efectos aleatorios fueron la severidad y los evaluadores; lo anterior se debe a que los niveles de severidad representados en las hojas se determinaron a través de números aleatorios, además de que el grupo de evaluadores representa una muestra aleatoria. Antes del ANOVA, los datos se sometieron a la transformación angular: $Z = \arcsin(\sqrt{\%estimado / 100})$.

El efecto del tipo de intervalo (logarítmico-lineal) en los parámetros indicadores de la exactitud, precisión y concordancia (μ , v , C_b , r , y ρ_c), se estudió a través del análisis bootstrap; con esta técnica de remuestreo se calcularon los intervalos de confianza de 95% (CIs) para la diferencia entre los parámetros estimados de ambos tipos de intervalo. Cuando los CIs incluyeron al cero, se concluyó que no hubo diferencia significativa entre escalas logarítmicas y escalas lineales. De la misma manera se analizó el efecto de los modos de uso (escala de proporción AV – escala categórica CLAS). Para desarrollar las pruebas anteriores, se realizó un análisis diferenciando a las escalas por su número de diagramas (3, 4, 5, 6 y 7); en cada caso se generaron 2 000 muestras balanceadas con el paquete bootstrap de R. Adicionalmente, para la distribución de los parámetros estimados se aplicó una prueba de normalidad Shapiro-Wilk (SW) (función shapiro.test). Cuando los datos de los parámetros (μ , v , C_b , r , y ρ_c) tuvieron distribución normal se aplicó una prueba de t (función t.test) para comparar las medias de dos grupos (logarítmico vs lineal y AV vs CLAS). Cuando los datos no siguieron una distribución normal se realizó un análisis no paramétrico, aplicando la prueba de Kolmogorov-Smirnov (KS) (función ks.test) para comparar las distribuciones de los grupos antes referidos.

Los errores se analizaron a través de gráficos de dispersión XY , en donde se trazó la línea de concordancia y los parámetros estimados del análisis de concordancia.

2.4. RESULTADOS

2.4.1. Efecto del tipo de intervalo (logarítmico-lineal) para escalas categóricas

En el análisis de concordancia, las estimaciones de los parámetros que miden el sesgo (v , u y C_b) variaron dependiendo del número de diagramas del SAD; aunque en la mayoría de estos parámetros se registraron diferencias significativas entre sus estimaciones, los SAD con intervalos logarítmicos no mostraron resultados predominantemente mejores que los SAD con intervalos lineales (Cuadro 2.2). Adicionalmente, el estimador de la precisión (r) proporcionó valores significativamente más altos para SAD con intervalos lineales, este resultado se registró en todos los comparativos del número de diagramas. En general, la estimación de r resultó en valores más altos en razón del incremento en el número de diagramas; además, se obtuvo una precisión mayor a 0.9 para SADs lineales con 4 o más diagramas, mientras que únicamente el SAD logarítmico de 7 diagramas obtuvo una precisión mayor a 0.9.

Los SADs con intervalos lineales se distinguieron porque registraron un sesgo de ubicación con valores positivos ($0.144 < \mu < 0.324$), mientras que los SADs con intervalos logarítmicos registraron valores negativos ($-0.188 < \mu < -0.099$); en general, para las escalas categóricas se registró un sesgo sistemático menor que 1 para los SADs con intervalos lineales ($0.899 < v < 0.971$) y logarítmicos ($0.884 < v < 0.958$).

Las pruebas estadísticas mostraron una diferencia significativa a favor de los SADs lineales, este resultado coincidió para todos los números de diagramas.

Cuadro 2.2. Efecto del tipo de intervalo y el número de diagramas del SAD, en el sesgo, precisión y concordancia en escalas categóricas. Modo de uso: CLAS

VARIABLE ¹	MEDIA ²		DIFERENCIA ³ ENTRE MEDIAS	95% CIs ⁴ DE LA DIFERENCIA		KS value (P)	t - value (P)
	INTERVALO LINEAL	INTERVALO LOGARÍTMICO					
3 DIAG							
Sesgo sistemático (v)	0.971 (0.026)	0.899 (0.035)	0.072 (0.044)	0.0701	0.0739	0.8205 (<0.0001)	
Sesgo de localización (μ)	0.150 (0.029)	-0.099 (0.037)	0.250 (0.048)	0.2478	0.2520		-234.25 (<0.0001)
Sesgo generalizado (C_b)	0.988 (0.004)	0.988 (0.006)	-0.001 (0.007)	-0.0010	-0.0003	0.183 (<0.0001)	
Coefficiente de Correl. (r)	0.884 (0.012)	0.802 (0.016)	0.083 (0.02)	0.0819	0.0836	0.9975 (<0.0001)	
LCCC (ρ_c)	0.873 (0.014)	0.792 (0.016)	0.081 (0.02)	0.0803	0.0821	0.9955 (<0.0001)	
4 DIAG							
Sesgo sistemático (v)	0.904 (0.024)	0.958 (0.029)	-0.054 (0.038)	-0.0557	-0.0523	0.753 (<0.0001)	
Sesgo de localización (μ)	0.324 (0.03)	-0.188 (0.032)	0.512 (0.044)	0.5103	0.5142	0.9999 (<0.0001)	
Sesgo generalizado (C_b)	0.945 (0.009)	0.981 (0.007)	-0.036 (0.011)	-0.0364	-0.0354	0.984 (<0.0001)	
Coefficiente de Correl. (r)	0.913 (0.009)	0.845 (0.014)	0.068 (0.017)	0.0668	0.0683	0.9965 (<0.0001)	
LCCC (ρ_c)	0.862 (0.014)	0.829 (0.014)	0.034 (0.02)	0.0327	0.0345	0.7845 (<0.0001)	
5 DIAG							
Sesgo sistemático (v)	0.920 (0.018)	0.916 (0.03)	0.004 (0.035)	0.0025	0.0055	0.1945 (<0.0001)	
Sesgo de localización (μ)	0.144 (0.022)	-0.157 (0.031)	0.303 (0.039)	0.3009	0.3043	0.9999 (<0.0001)	
Sesgo generalizado (C_b)	0.986 (0.003)	0.983 (0.007)	0.003 (0.007)	0.0023	0.0029	0.276 (<0.0001)	
Coefficiente de Correl. (r)	0.943 (0.007)	0.851 (0.017)	0.092 (0.018)	0.0913	0.0929	0.9999 (<0.0001)	
LCCC (ρ_c)	0.930 (0.008)	0.837 (0.017)	0.093 (0.019)	0.0922	0.0939	0.9999 (<0.0001)	
6 DIAG							
Sesgo sistemático (v)	0.944 (0.017)	0.946 (0.028)	-0.002 (0.034)	-0.0035	-0.0005		2.4096 (0.01602)
Sesgo de localización (μ)	0.182 (0.022)	-0.103 (0.031)	0.286 (0.039)	0.2846	0.2880	0.9999 (<0.0001)	
Sesgo generalizado (C_b)	0.982 (0.004)	0.992 (0.004)	-0.011 (0.006)	-0.0109	-0.0104	0.8185 (<0.0001)	
Coefficiente de Correl. (r)	0.945 (0.007)	0.862 (0.014)	0.082 (0.016)	0.0814	0.0828	0.9999 (<0.0001)	
LCCC (ρ_c)	0.927 (0.01)	0.856 (0.014)	0.071 (0.017)	0.0707	0.0722	0.9965 (<0.0001)	
7 DIAG							
Sesgo sistemático (v)	0.899 (0.015)	0.884 (0.025)	0.016 (0.03)	0.0147	0.0173	0.4005 (<0.0001)	
Sesgo de localización (μ)	0.144 (0.021)	-0.184 (0.024)	0.329 (0.032)	0.3281	0.3309		-452.22 (<0.0001)
Sesgo generalizado (C_b)	0.984 (0.003)	0.975 (0.006)	0.009 (0.007)	0.0082	0.0088	0.6625 (<0.0001)	
Coefficiente de Correl. (r)	0.951 (0.007)	0.924 (0.01)	0.027 (0.012)	0.0265	0.0276	0.891 (<0.0001)	
LCCC (ρ_c)	0.936 (0.008)	0.902 (0.009)	0.035 (0.013)	0.0340	0.0351	0.956 (<0.0001)	

¹ El máximo nivel de concordancia se obtiene con $\mu = 0$, $v = 1$, $C_b = 1$, $r = 1$, lo que implica $\rho_c = 1$

² Desviación estándar entre paréntesis

³ Promedio de la diferencia de cada parámetro, desviación estándar entre paréntesis. Basado en 2 000 muestras bootstrap.

⁴ Intervalo de confianza (CIs). Basado en 2 000 muestras bootstrap. Si CIs incluye al cero, la diferencia es no significativa.

El análisis bootstrap se desarrolló con el conjunto combinado de datos de todos los evaluadores.

La concordancia general (ρ_c), entre la severidad estimada y la severidad real, fue estadísticamente mayor para los SAD con intervalos lineales ($0.873 < \rho_c < 0.936$), en comparación con los de intervalos logarítmicos ($0.792 < \rho_c < 0.902$); se observó que el valor del parámetro varió dependiendo del número de diagramas; los SADs lineales con 5, 6 y 7 diagramas registraron una concordancia en torno a 0.93; en contraste, el SAD logarítmico de 7 diagramas fue el único que resultó con valores cercanos a 0.90. El parámetro ρ_c mostró consistentemente que los SADs con intervalos lineales estuvieron más cercanos al valor máximo (1), en comparación con los SADs con intervalos logarítmicos. El análisis bootstrap, la prueba *KS* y la prueba *t* indican niveles de concordancia significativamente más altos para las escalas categóricas con SADs a intervalos lineales.

2.4.2. Efecto del tipo de intervalo (logarítmico-lineal) para escalas de proporción

El análisis bootstrap de los componentes del LCC mostró que el intervalo de incremento (logarítmico - lineal) tiene un efecto altamente significativo en el sesgo (ν , μ y C_b) y en la precisión (r) de la evaluación visual para escalas de proporción. Al igual que en los resultados anteriores, los SADs lineales registraron niveles más cercanos a los valores óptimos de los parámetros, en comparación con los SADs logarítmicos. De esta forma, los SADs lineales tuvieron un sesgo generalizado C_b mayor que 0.99, además de una precisión que registró niveles superiores a 0.96. El estadístico que mide la concordancia (ρ_c) proporcionó los valores más cercanos a 1 cuando los SADs tuvieron intervalos lineales ($0.96 < \rho_c < 0.97$); en contraste, los SADs logarítmicos obtuvieron valores estadísticamente menores ($0.93 < \rho_c < 0.95$).

Cuadro 2.3. Efecto del tipo de intervalo y el número de diagramas del SAD, en el sesgo, precisión y concordancia en escalas de proporción. Modo de uso: AV

VARIABLE ¹	MEDIA ²		DIFERENCIA ³ ENTRE MEDIAS	95% CIs ⁴ DE LA DIFERENCIA		KS value (P)	t - value (P)
	INTERVALO LINEAL	INTERVALO LOGARÍTMICO					
3 DIAG							
Sesgo sistemático (v)	0.959 (0.012)	0.962 (0.019)	-0.003 (0.023)	-0.0040	-0.0020	0.183 (<0.0001)	
Sesgo de localización (μ)	-0.008 (0.015)	0.078 (0.02)	-0.087 (0.026)	-0.0885	-0.0862	0.9901 (<0.0001)	
Sesgo generalizado (C_b)	0.999 (0.001)	0.996 (0.002)	0.003 (0.002)	0.0030	0.0032	0.871 (<0.0001)	
Coefficiente de Correl. (r)	0.968 (0.004)	0.938 (0.007)	0.030 (0.008)	0.0298	0.0305	0.9935 (<0.0001)	
LCCC (ρ_c)	0.967 (0.004)	0.934 (0.007)	0.033 (0.009)	0.0327	0.0334	0.997 (<0.0001)	
4 DIAG							
Sesgo sistemático (v)	0.954 (0.012)	0.978 (0.019)	-0.024 (0.023)	-0.0250	-0.0230	0.61 (<0.0001)	
Sesgo de localización (μ)	-0.009 (0.014)	0.062 (0.019)	-0.073 (0.025)	-0.0738	-0.0716	0.9645 (<0.0001)	
Sesgo generalizado (C_b)	0.999 (0.001)	0.997 (0.001)	0.001 (0.001)	0.0012	0.0014	0.5415 (<0.0001)	
Coefficiente de Correl. (r)	0.97 (0.004)	0.942 (0.007)	0.028 (0.008)	0.0277	0.0284	0.9945 (<0.0001)	
LCCC (ρ_c)	0.969 (0.004)	0.94 (0.007)	0.029 (0.008)	0.0289	0.0295	0.9975 (<0.0001)	
5 DIAG							
Sesgo sistemático (v)	0.965 (0.012)	0.977 (0.017)	-0.012 (0.022)	-0.0130	-0.0110	0.3575 (<0.0001)	
Sesgo de localización (μ)	-0.017 (0.014)	0.066 (0.019)	-0.084 (0.024)	-0.0852	-0.0831	0.9865 (<0.0001)	
Sesgo generalizado (C_b)	0.999 (0.001)	0.997 (0.001)	0.002 (0.001)	0.0018	0.0019	0.703 (<0.0001)	
Coefficiente de Correl. (r)	0.969 (0.003)	0.95 (0.006)	0.019 (0.007)	0.0188	0.0194	0.966 (<0.0001)	
LCCC (ρ_c)	0.968 (0.004)	0.948 (0.006)	0.021 (0.007)	0.0205	0.0211	0.9765 (<0.0001)	
6 DIAG							
Sesgo sistemático (v)	0.954 (0.012)	0.958 (0.016)	-0.003 (0.02)	-0.0039	-0.0021	0.146 (<0.0001)	
Sesgo de localización (μ)	0.05 (0.014)	0.147 (0.018)	-0.096 (0.024)	-0.0975	-0.0954	0.995 (<0.0001)	
Sesgo generalizado (C_b)	0.997 (0.001)	0.988 (0.003)	0.009 (0.003)	0.0093	0.0095	0.996 (<0.0001)	
Coefficiente de Correl. (r)	0.971 (0.004)	0.955 (0.005)	0.016 (0.006)	0.0154	0.0160	0.9325 (<0.0001)	
LCCC (ρ_c)	0.968 (0.004)	0.944 (0.006)	0.025 (0.007)	0.0243	0.0249	0.9965 (<0.0001)	
7 DIAG							
Sesgo sistemático (v)	0.955 (0.011)	0.952 (0.016)	0.003 (0.02)	0.0021	0.0039	0.144 (<0.0001)	
Sesgo de localización (μ)	0.064 (0.015)	0.161 (0.017)	-0.097 (0.024)	-0.0979	-0.0958	0.998 (<0.0001)	
Sesgo generalizado (C_b)	0.997 (0.001)	0.986 (0.003)	0.011 (0.003)	0.0109	0.0111	0.9985 (<0.0001)	
Coefficiente de Correl. (r)	0.969 (0.004)	0.959 (0.005)	0.01 (0.007)	0.0097	0.0103	0.7300 (<0.0001)	
LCCC (ρ_c)	0.966 (0.005)	0.945 (0.006)	0.021 (0.007)	0.0202	0.0208	0.9595 (<0.0001)	

¹ El máximo nivel de concordancia se obtiene con $\mu = 0$, $v = 1$, $C_b = 1$, $r = 1$, lo que implica $\rho_c = 1$

² Desviación estándar entre paréntesis

³ Promedio de la diferencia de cada parámetro, desviación estándar entre paréntesis. Basado en 2 000 muestras bootstrap.

⁴ Intervalo de confianza (CIs). Basado en 2 000 muestras bootstrap. Si CIs incluye al cero, la diferencia es no significativa.

El análisis bootstrap se desarrolló con el conjunto combinado de datos de todos los evaluadores.

2.4.3. Efecto del número de diagramas

Para las escalas de proporción con SADs, el valor de ρ_c no muestra una tendencia definida en relación con el número de diagramas. Únicamente en las escalas con 6 y 7 diagramas se observa un sesgo de localización positivo ($\mu > 0$); resultados similares se observaron para todos los SADs logarítmicos (Cuadro 2.3).

En el caso de las escalas categóricas con SADs, en el análisis de concordancia, la precisión (r) es el parámetro que marcó la diferencia entre escalas lineales y logarítmicas. Los SADs lineales registraron una precisión significativamente mayor que los SADs logarítmicos, además el parámetro mostró incrementos a medida que se aumentó el número de diagramas. Estos resultados se obtuvieron a través de los 95% CI y la prueba *KS*. Por su parte, los parámetros relacionados con el sesgo (v , u y C_b) variaron dependiendo del número de diagramas; un resultado destacable es que las escalas logarítmicas tuvieron una asociación predominante con los valores negativos de μ (Cuadro 2.2)

2.4.4. Comparación entre escalas categóricas y de proporción.

Con los resultados del estadístico ρ_c se obtuvo una agrupación de medias para diferenciar a los 22 métodos de evaluación (20 escalas + EVD + HB). Para lograr lo anterior, se estableció un criterio denominado *diferencia mínima operativa* (dmo), cuyo valor se estableció en 0.01; los métodos de evaluación cuyas medias tuvieron una diferencia mayor a 0.01, son considerados diferentes, por lo que se les asignan letras distintas. El criterio anterior se implementó debido a que los CIs, la prueba *KS* y la prueba *t* tienen la fortaleza de detectar y evidenciar pequeñas diferencias entre las poblaciones

de la distribución bootstrap; con base en estas pruebas, todos los métodos son estadísticamente diferentes. Sin embargo, para fines práctico se desarrolló la agrupación de la Cuadro 2.4, con base en la $dmo=0.01$.

Cuadro 2.4. Coeficiente de Correlación de Concordancia de Lin (ρ_c). Comparativo entre número de diagramas, intervalos de incremento y modos de uso.

ESCALA Y USO ¹	CONCORDANCIA ² (ρ_c)	AGRUPACIÓN ³
4_LIN_AV	0.969	A
5_LIN_AV	0.968	A
6_LIN_AV	0.968	A
3_LIN_AV	0.967	A
7_LIN_AV	0.966	A
5_LOG_AV	0.948	B
7_LOG_AV	0.945	B C
6_LOG_AV	0.944	B C D
4_LOG_AV	0.940	B C D E
7_LIN_CLAS	0.936	C D E F
3_LOG_AV	0.934	D E F G
5_LIN_CLAS	0.930	E F G
6_LIN_CLAS	0.927	F G
EVD	0.925	G
7_LOG_CLAS	0.902	H
HB	0.874	I
3_LIN_CLAS	0.873	I
4_LIN_CLAS	0.862	J
6_LOG_CLAS	0.856	J
5_LOG_CLAS	0.837	K
4_LOG_CLAS	0.829	K
3_LOG_CLAS	0.792	L

¹ El número indica la cantidad de diagramas. LIN: intervalo lineal, LOG: intervalo logarítmico. AV: Ayuda Visual (escala de proporción), CLAS: Clasificación (escala categórica). EVD: Evaluación Visual Directa. HB: Escala Horsfall-Barratt.

² Parámetro estimado con base en 2 000 muestras bootstrap.

³ Para fines prácticos, la diferencia mínima operativa para considerar distintos a los estimadores es de 0.01.

El nivel más alto se concordancia se obtiene con escalas de proporción (AV), sobresalen los SADs a intervalos lineales (Figura 2.8); para este tipo de instrumentos, los resultados indican que el uso de 3 a 5 diagramas proporciona resultados tan concordantes como la

evaluación con 7 diagramas ($\rho_c > 0.96$). Continuando con las escalas de proporción (AV), los SADs con intervalos logarítmicos (4 a 7 diagramas) proporcionan con concordancia ($0.94 < \rho_c < 0.95$) que, en términos operativos y estadísticos, es mayor a la EVD.

La escala HB y las categóricas con SADs a intervalos logarítmicos (3 a 7 diagramas) o lineales (3 a 4 diagramas), no evidencian mayores niveles de concordancia, en comparación con la evaluación visual directa EVD.

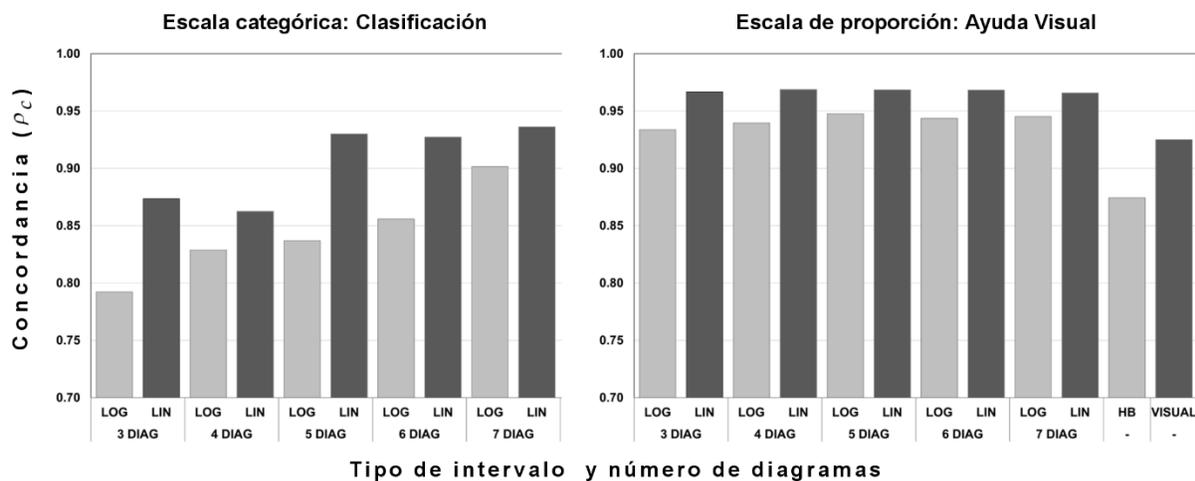


Figura 2.8. Coeficiente de Correlación de Concordancia de Lin. Comparativo entre escalas categóricas y de proporción. Estimadores calculados con base en 2 000 muestras bootstrap.

El coeficiente de correlación intraclase (ρ) indica que las evaluaciones son más consistentes con SADs de escalas de proporción (AV) ($0.88 < \rho < 0.97$), en comparación con los SADs de escalas categóricas (CLAS) ($0.83 < \rho < 0.93$). Además, dentro de las escalas de proporción, los SADs con incrementos lineales proporcionan la mayor

confiabilidad inter-evaluadores ($0.95 < \rho < 0.97$); mientras que, con las escalas categóricas, la confiabilidad más baja se obtuvo con los SADS con incrementos logarítmicos ($0.84 < \rho < 0.93$).

La confiabilidad inter-evaluadores (ρ) registró incrementos conforme se aumentó el número de diagramas; esta relación fue notable para las escalas de categóricas con intervalos lineales. Para las escalas de proporción, el número de diagramas no tiene un efecto claro sobre la confiabilidad inter-evaluadores; sin embargo, con este tipo de instrumento se obtuvieron los niveles más altos de ρ .

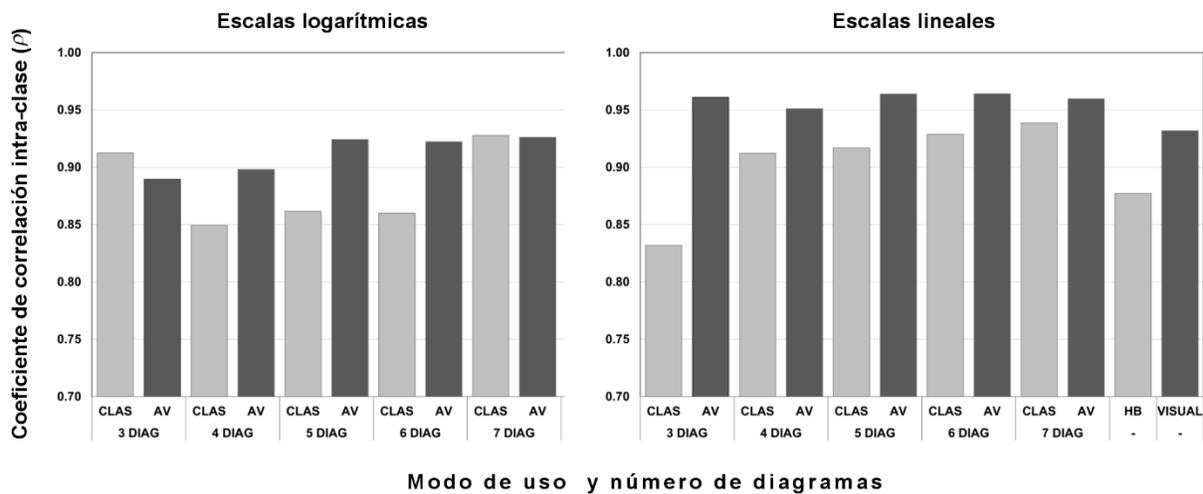


Figura 2.9. Coeficiente de Correlación Intra-clase, comparativo entre escalas logarítmicas y lineales. Estimadores calculados con base en 2 000 muestras bootstrap.

2.5. DISCUSIÓN

La escala HB ha sido objeto de investigaciones que han evidenciado el incumplimiento de su objetivo principal: incrementar la exactitud y la precisión de las estimaciones visuales (Bock *et al.*, 2009; Bock *et al.*, 2013; Bock *et al.*, 2010). Esta escala, por mucho tiempo justificó su uso con el supuesto de que la agudeza visual era extremadamente limitada en el intervalo 25-75%; en la actualidad se ha demostrado que los evaluadores tienen capacidad visual para discernir niveles de severidad dentro del intervalo antes mencionado (Nutter y Esker, 2006). La EVD es afectada por la falta de precisión, exactitud y confiabilidad (El Jarroudi *et al.*, 2015); sin embargo, la escala HB no incrementa los estimadores de esos parámetros, y en muchas ocasiones los reduce (Nita *et al.*, 2003) (Figura 2.10). Adicionalmente, se ha documentado que los evaluadores no categorizan correctamente con la escala HB, lo cual se debe al gran número de clases, y pone en manifiesto la reducida funcionalidad de los incrementos logarítmicos de este instrumento (Bock, *et al.*, 2013) .

En el presente estudio, la escala HB no redujo el sesgo, tampoco mejoró la precisión y la exactitud; y en términos generales, disminuyó la concordancia de las estimaciones, en comparación con la EVD. Estos resultados coinciden con lo reportado en estudios previos (Bardsley y Ngugi, 2013; Nita *et al.*, 2003; Bock *et al.*, 2013)

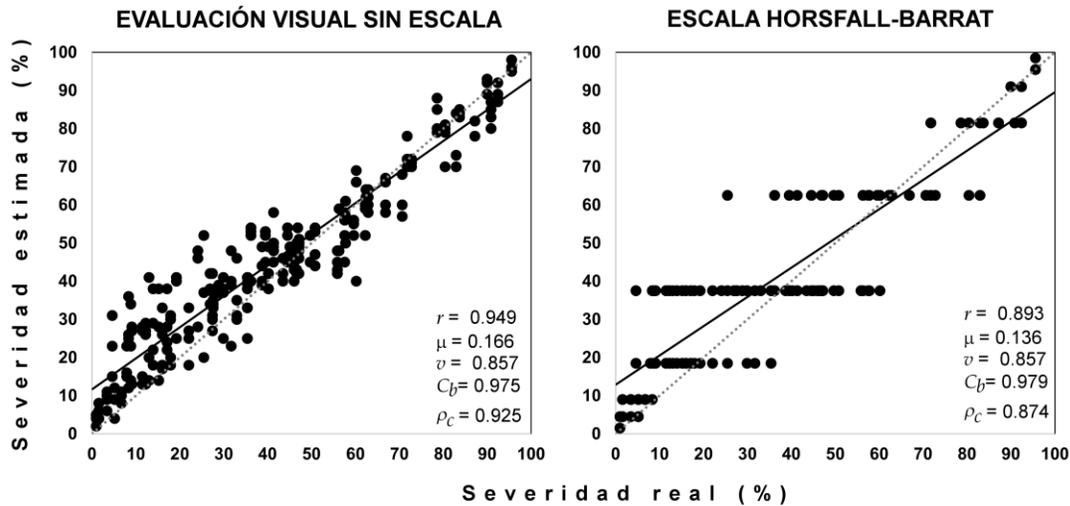


Figura 2.10. Análisis de concordancia. Comparativo entre Evaluación Visual sin escala (EVD) y la transformación a puntos medios de la escala de Horsfall-Barrat. Estimadores calculados con base en 2 000 muestras bootstrap.

Los parámetros del análisis de concordancia no muestran que la escala de HB ofrezca alguna ventaja en comparación con EVD (Figura 2.10). Además, la reducción en la precisión (r) sugiere que la escala HB aporta una fuente de error adicional al proceso de evaluación, además de las ya conocidas para la EVD.

Al comparar los SADs con 3 diagramas (Figura 2.11), se observa que las escalas categóricas (CLAS) simplifican la evaluación visual al punto de que las unidades de evaluación se clasificaron en solo tres categorías; por lo anterior, el uso de este tipo de instrumentos “obliga” al evaluador a asumir errores de hasta 47 puntos porcentuales (máximo error registrado). En contraste, las escalas de proporción (AV) proveen mayor libertad al usuario, con la ventaja de la estandarización del criterio de evaluación a través del SAD, el error máximo registrado con AV fue de 26 puntos porcentuales.

Las escalas de proporción con SADs lineales incrementan de manera significativa la precisión ($r > 0.96$), siendo éste el parámetro que produce una concordancia mayor en relación con las otras escalas (Bardsley y Ngugi 2013).

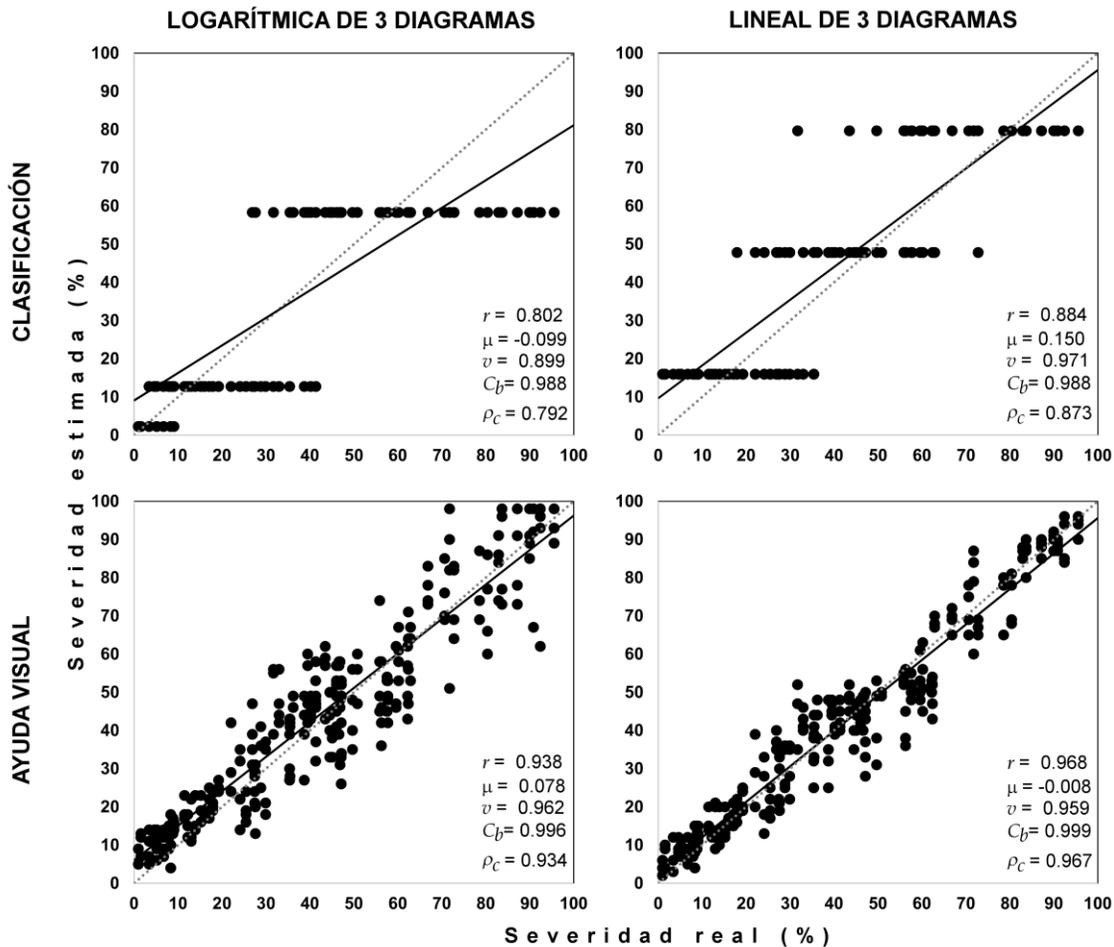


Figura 2.11. Análisis de concordancia para escalas de 3 diagramas. Comparativo del modo de uso entre escalas logarítmicas y lineales. Estimadores calculados con base en 2 000 muestras bootstrap.

Las escalas categóricas (CLAS) pueden ser útiles en experimentos en los que se tienen que clasificar numerosos tratamientos; por ejemplo, en estudios para seleccionar

germoplasma para resistencia (Subrahmanyam *et al.*, 1982). Pero no son apropiadas para el estudio cuantitativo de las epidemias (Bock *et al.*, 2010).

En este estudio, las escalas categóricas con SADs lineales registraron un sesgo de localización positivo ($0.324 < \mu < 0.144$), mientras que los SADs con intervalos logarítmicos registraron valores negativos ($-0.188 < \mu < -0.099$). Esto es originado por una sobreestimación de las escalas lineales, sobre todo, a bajos niveles de severidad (<10%). Para reducir la magnitud de estas sobreestimaciones, Nita *et al.* (2003) recomiendan incorporar diagramas adicionales en el intervalo 0 a 10%, principalmente cuando la enfermedad produzca frecuentemente severidades en ese rango.

Diversas investigaciones han confirmado que las estimaciones de severidad están linealmente relacionadas con los valores reales de severidad (Bock *et al.*, 2008; Nita, *et al.*, 2003; Nutter Jr *et al.*, 1991) . A pesar de eso, en la actualidad las escalas con intervalos logarítmicos siguen en uso; Del Ponte *et al.* (2017), a través de un estudio cuantitativo, detectaron distintas variantes de escalas desarrolladas con intervalos logarítmicos. Schwanck y Del Ponte (2014) compararon escalas de proporción con incrementos lineales y logarítmicos, en el intervalo de 0 a 36%, y encontraron niveles de concordancia similares para ambos tipos de intervalo. Sin embargo, en el presente estudio se evaluó el intervalo de 0 a 100%, con diferente número de diagramas, y en todos los casos los intervalos lineales proporcionaron niveles de precisión (r) y concordancia (ρ_c) más cercanos al óptimo (es decir, igual a 1).

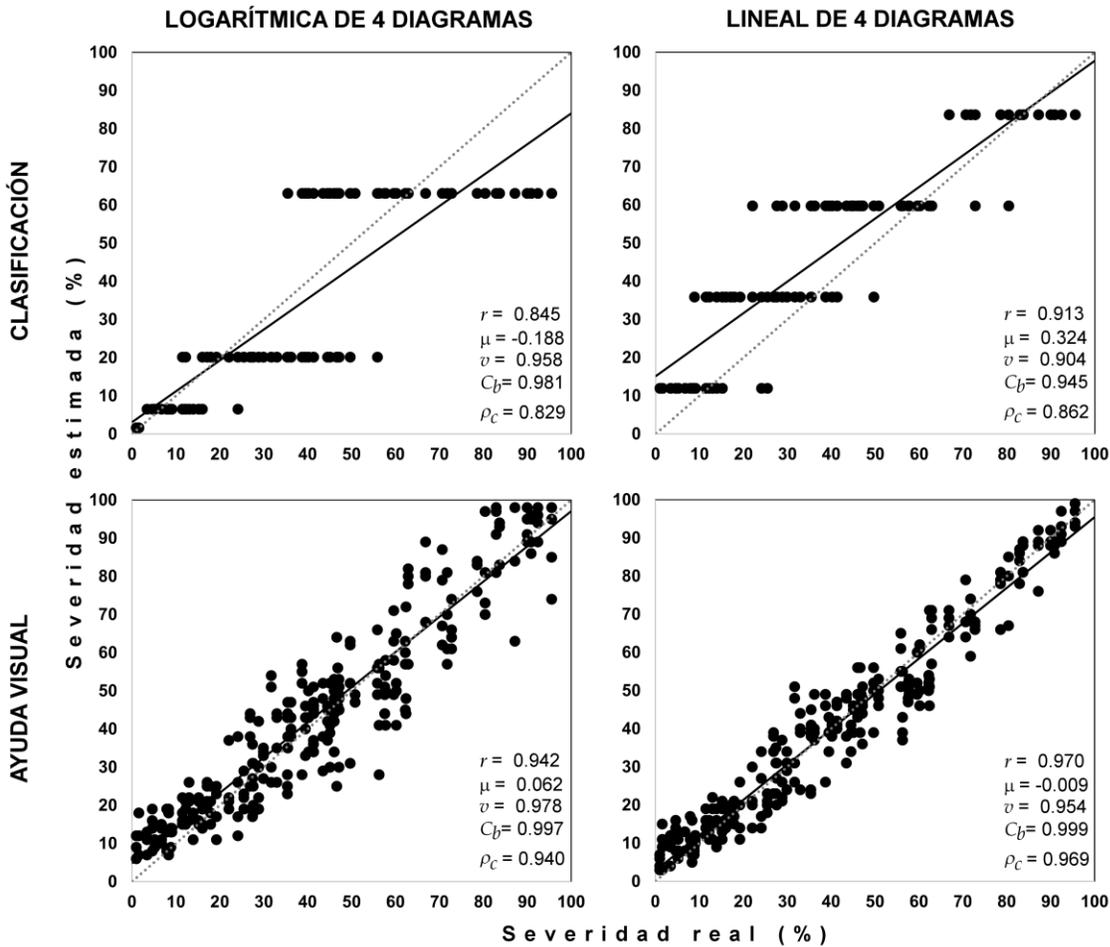


Figura 2.12. Análisis de concordancia para escalas de 4 diagramas. Comparativo del modo de uso entre escalas logarítmicas y lineales. Estimadores calculados con base en 2 000 muestras bootstrap.

En evaluación visual se ha detectado una tendencia a evaluar usando puntos preferidos o "nudos" (5, 10, 20, 30%, etc.) (Bock *et al.*, 2010); esta práctica es inconsciente en la mayoría de los casos, y se puede alentar cuando se diseñan SADs con diagramas cercanos a los nudos. Nutter y Esker (2006) indican que la evaluación con nudos puede estar relacionada con factores psicológicos; y aunque este fenómeno no se ha estudiado completamente, es claro que hay una fuente de error adicional, relacionada con la

elección del nudo, situación similar a lo que ocurre con las escalas categóricas. Schwanck y Del Ponte (2014) observaron una sobreestimación consistente cuando los evaluadores prefirieron los nudos convencionales (5, 10, 20, 30%, etc.) o crearon sus propios nudos en torno a los valores de los diagramas.

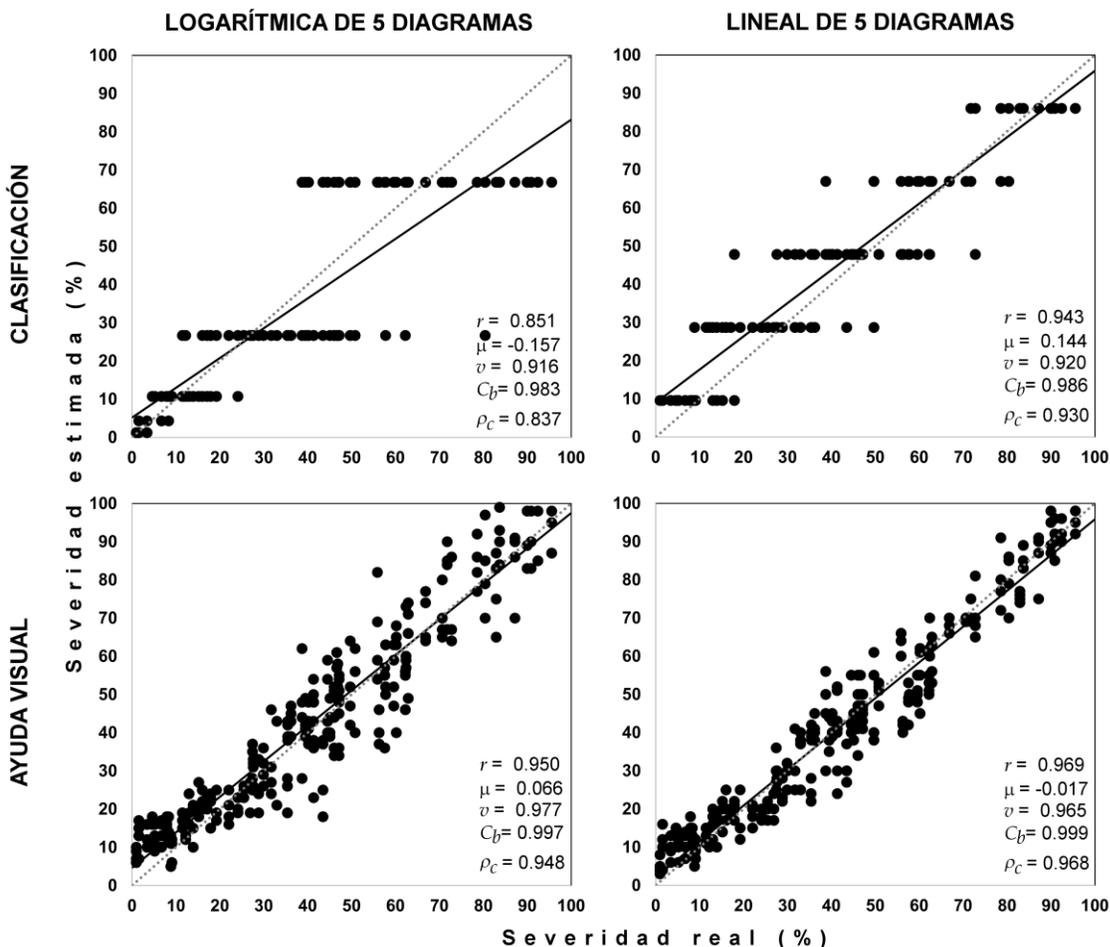


Figura 2.13. Análisis de concordancia para escalas de 5 diagramas. Comparativo del modo de uso entre escalas logarítmicas y lineales. Estimadores calculados con base en 2 000 muestras bootstrap.

En esta investigación, los incrementos de los SADs se definieron de manera independiente a los nudos. Para las escalas de proporción, los participantes fueron instruidos para que libremente, mediante interpolación o extrapolación, asignaran un número entero entre 0 y 100%. Las gráficas de AV no presentan observaciones en línea horizontal, por lo que se infiere que no predominó la evaluación en nudos.

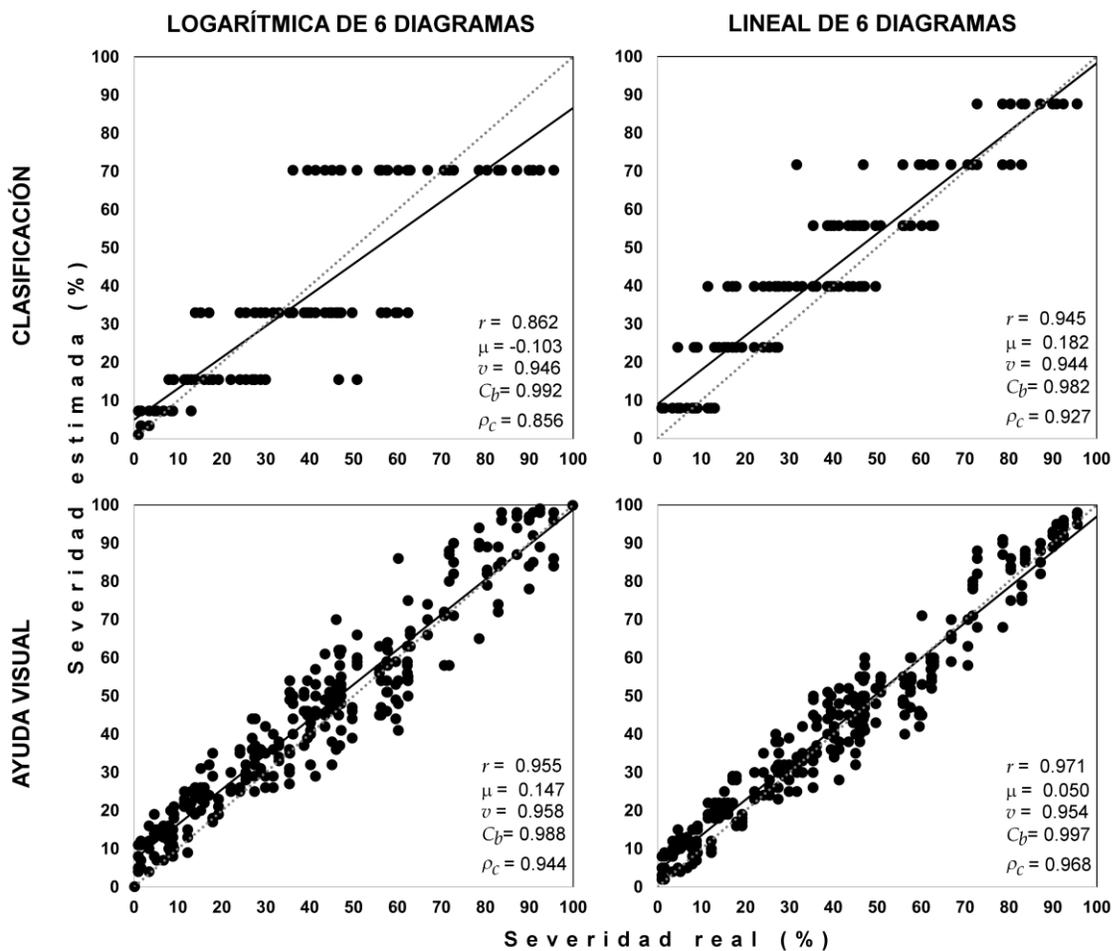


Figura 2.14. Análisis de concordancia para escalas de 6 diagramas. Comparativo del modo de uso entre escalas logarítmicas y lineales. Estimadores calculados con base en 2 000 muestras bootstrap.

De manera específica, el uso de escalas de proporción con SADs incrementó la confiabilidad inter-evaluadores, medida a través del coeficiente de correlación intra-clase ($0.88 < \rho < 0.96$), este resultado coincide con estudios previos (Bardsley y Ngugi, 2013; Nita *et al.*, 2003; Yadav *et al.*, 2013). En contraste, las escalas categóricas con SADs no evidenciaron algún incremento en la confiabilidad inter-evaluadores ($0.83 < \rho < 93$), en comparación con la EVD ($\rho = 93$); este resultado es relevante porque, en teoría, los métodos de evaluación estándar deberían producir estimaciones más consistentes entre los evaluadores (Schwanck y Del Ponte, 2014).

Algunos autores proponen el uso alternado de escalas; intervalos logarítmicos cuando la severidad real tenga niveles bajos (<20 %) e intervalos lineales cuando la severidad en campo alcance niveles altos (>60 %) (Schwanck y Del Ponte, 2014). En Figura 2.15 se observa que las escalas lineales proporcionan resultados más concordantes que las escalas logarítmicas en el intervalo 60-100%; además, las escalas logarítmicas originan estimaciones más cercanas a la línea de concordancia en el intervalo 0-20%, pero este resultado es notable únicamente con las escalas categóricas. Con base en los resultados de la presente investigación, se recomienda utilizar SADs a intervalos lineales, añadiendo diagramas adicionales a bajos niveles de severidad (0 a 10%) para incrementar la concordancia en ese intervalo; recomendaciones similares fueron derivadas del estudio realizado por Nita *et al.* (2003), aunque en su estudio no desarrollaron SADs.

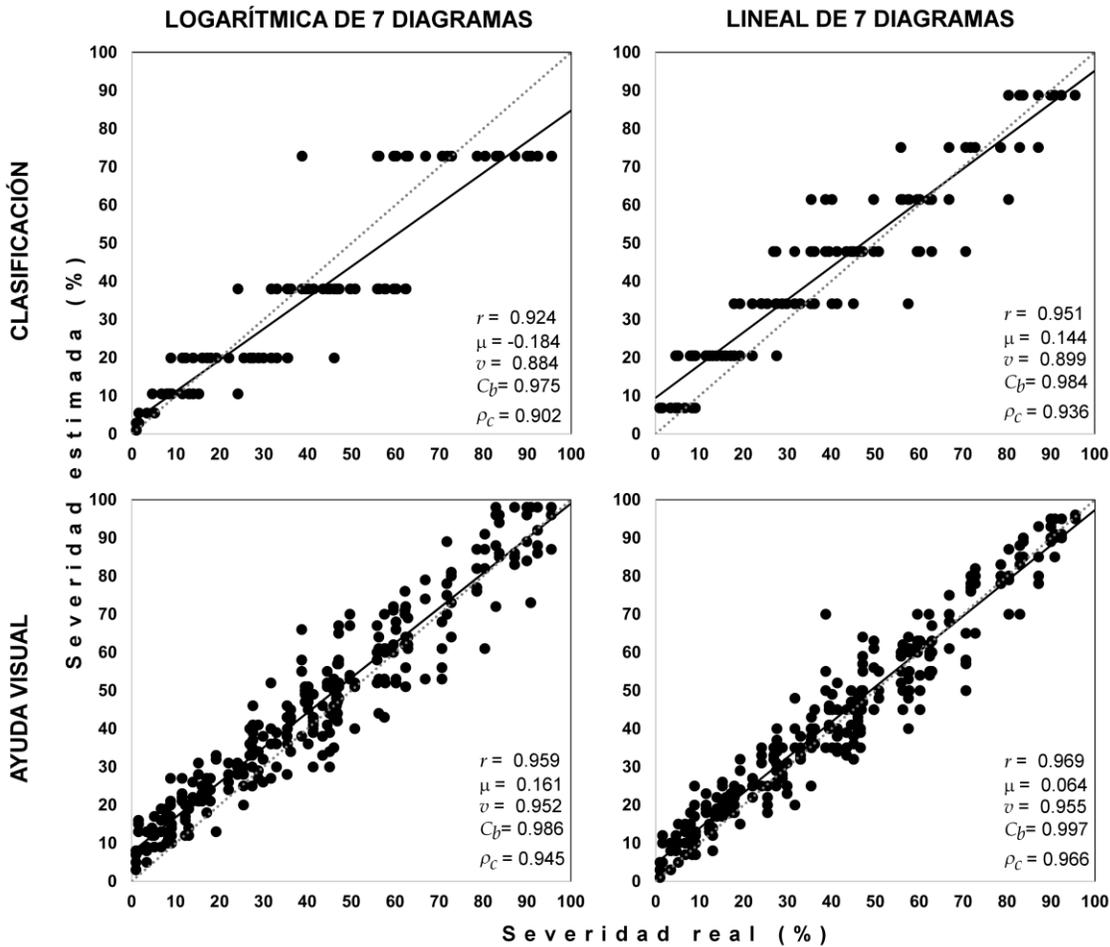


Figura 2.15. Análisis de concordancia para escalas de 7 diagramas. Comparativo del modo de uso entre escalas logarítmicas y lineales. Estimadores calculados con base en 2 000 muestras bootstrap.

El análisis de concordancia evidencia que los SADs con intervalos lineales proporcionan más exactitud, precisión y confiabilidad que los SADs con intervalos logarítmicos. Además, las escalas de proporción son significativamente mejores que las escalas categóricas. En este estudio, los instrumentos de evaluación más precisos, exactos y confiables fueron las escalas de proporción (AV) con SADs a intervalos lineales; para este tipo de instrumento, el número de diagramas puede variar entre 3 y 7 ya que sus

niveles de concordancia son similares; este resultado coincide con los de Bock *et al.* (2015).

En estudios posteriores, en donde se analice el tiempo de evaluación, se podría identificar el número de diagramas que produce una evaluación más eficiente. Para fines prácticos, es recomendable utilizar escalas de proporción (AV) con SADs a intervalos lineales con 3 a 5 diagramas, además de colocar 1 o 2 diagramas adicionales a bajos niveles de severidad.

2.6. LITERATURA CITADA

- Bardsley, S. J., and Ngugi, H. K. 2013. Reliability and accuracy of visual methods to quantify severity of foliar bacterial spot symptoms on peach and nectarine. *Plant Pathology*. 62:460–474
- Bock, C. H., Chiang, K. S. and Del Ponte, E. M. 2016. Accuracy of plant specimen disease severity estimates: Concepts, history, methods, ramifications and challenges for the future. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources* 11, 039, pp 1-13.
- Bock, C. H., El Jarroudi, M., Kouadio, L. A., Mackels, C., Chiang, K. S., and Delfosse, P. 2015. Disease Severity Estimates—Effects of Rater Accuracy and Assessment Methods for Comparing Treatments. *Plant Disease*. 99:1104–1112
- Bock, C. H., Gottwald, T. R., Parker, P. E., Cook, A. Z., Ferrandino, F., Parnell, S., *et al.* 2009. The Horsfall-Barratt scale and severity estimates of citrus canker. *European Journal of Plant Pathology*. 125:23–38
- Bock, C. H., Gottwald, T. R., Parker, P. E., Ferrandino, F., Welham, S., Bosch, F. van den, *et al.* 2010. Some Consequences of Using the Horsfall-Barratt Scale for Hypothesis Testing. *Phytopathology*. 100:1030–1041
- Bock, C. H., Hotchkiss, M. W., and Wood, B. W. 2015. Assessing disease severity: accuracy and reliability of rater estimates in relation to number of diagrams in a standard area diagram set. *Plant Pathology*.

- Bock, C. H., Parker, P. E., Cook, A. Z., and Gottwald, T. R. 2008. Visual rating and the use of image analysis for assessing different symptoms of citrus canker on grapefruit leaves. *Plant Disease*. 92:530–541
- Bock, C. H., Poole, G. H., Parker, P. E., and 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., Den Bosch, F. Van, Parnell, S., and 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
- Chiang, K. S., Liu, S. H., Bock, C. H., and Gottwald, T. R. 2014. What interval characteristics make a good categorical disease assessment scale? *Phytopathology*. 104:575–85
- Cobb, N. 1892. Contribution to an economic knowledge of the Australian rusts (Uredinae). *Agricultural Gazette (NSW)* 3:60.
- Costa Lage, D. A., Marouelli, W. A., Duarte, H. S., and Café-Filho, A. C. 2015. Standard area diagrams for assessment of powdery mildew severity on tomato leaves and leaflets. *Crop Protection*. 67:26–34
- Del Ponte, E. M., Pethybridge, S. J., Bock, C. H., Michereff, S. J., Machado, F. J., Spolti, P., 2017. Standard Area Diagrams for Aiding Severity Estimation: Scientometrics, Pathosystems and Methodological Trends in the last 25 years. *Phytopathology*
- El Jarroudi, M., Kouadio, A. L., Mackels, C., Tychon, B., Delfosse, P., and Bock, C. H. 2015. A comparison between visual estimates and image analysis

measurements to determine septoria leaf blotch severity in winter wheat. *Plant Pathology*. 64:355–364

- Godoy, C. V., Koga, L. J., y Canteri, M. G., 2006. Diagrammatic scale for assessment of soybean rust severity. *Fitopatol Bras* 31:63–68.
- Horsfall, J. C., y Barrat, R. W., 1945. An improved grading system for measuring plant diseases. *Phytopathology* 35:665.
- Large, E. C. 1966. Measuring plant disease. *Ann. Rev. Phytopath.* 4: 9–26.
- Lin, L. I. 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 45:255–268
- Madden, L. V., Hughes, G., and Den Bosch, F. Van. 2007. *The Study of Plant Disease Epidemics*. St. Paul, MN.: American Phytopathological Society.
- Nita, M., Ellis, M. A., and Madden, L. V. 2003. Reliability and accuracy of visual estimation of phomopsis leaf blight of strawberry. *Phytopathology*. 93:995–1005
- Nutter, F. W., and Esker, P. D. 2006. The Role of Psychophysics in Phytopathology: The Weber–Fechner Law Revisited. *European Journal of Plant Pathology*. 114:199–213
- Nutter Jr, F. W., Teng, P. S., and Shokes, F. M. 1991. Disease Assessment Terms and Concepts- Special Tipoc-. *Plant Disease*. 75:1187–1188
- Nutter Jr., F. W.; Worawitlikit, O. 1989. Disease Pro: a computer program for evaluating and improving a person ability to assess disease proportion. *Phytopatology* 79: 1135.

- Osada, V. H. K., Mora, A. G. 1997. 2LOG. Programa para desarrollar escalas de severidad por el método de Horsfall y Barratt. Manual del Usuario. Montecillo, estado de México, Colegio de Postgraduados.
- Subrahmanyam, P., McDonald, D., Gibbons, R. W., Nigam, S. N., and Nevill, D. J. 1982. Resistance to rust and late leafspot diseases in some genotypes of *Arachis hypogea*. *Peanut Sci.* 9:6-10.
- Schwanck, A. A., and Ponte, E. M. Del. 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
- Yadav, N. V., Vos, S. M. de, Bock, C. H., and 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