Focus on Data

Assessing Intereye Symmetry and Its Implications for Study Design

Maureen G. Maguire

Department of Ophthalmology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States

Correspondence: Maureen G. Maguire, Department of Ophthalmology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; maguirem@pennmedicine.upenn.edu.

Received: March 19, 2020
Accepted: March 31, 2020
Published: June 12, 2020

Citation: Maguire MG. Assessing intereye symmetry and its implications for study design. *Invest Ophthalmol Vis Sci.* 2020;61(6):27. https://doi.org/10.1167/iovs.61.6.27

Keywords: intereye symmetry, statistical methods, data analyses

Many morphologic and functional characteristics are very similar between right and left eyes of humans and animals. In disease-free individuals, iris color, axial length, refractive error, visual acuity, and cup-to-disc ratio are among the features that are nearly the same for the two eyes. In fact, deviations from symmetry often indicate the presence of disease or pathology. For example, asymmetry in cup-to-disc ratio is a sign of glaucoma, and asymmetry in refractive error is a sign of amblyopia. When one eye has a disease, similar findings in the contralateral eye may be evidence that genetic or systemic factors are the cause. The symmetry in measurements between eyes also has implications for the design of research studies, including whether to include one or both eyes in observational studies, and whether to treat one eye and use the other as a control in treatment trials.

The objective of this article is to provide data analytic approaches to assessing the degree of symmetry between eyes and to provide some guidance on how to use knowledge about intereye symmetry in designing research studies. Many of the statistical methods used to assess symmetry fall under the general category of methods to assess agreement between two sets of measurements, in this case, right eye measurements and left eye measurements. JMP (SAS Institute, Cary, NC, USA), Stata (StataCorp, College Station, TX, USA), SAS (SAS Institute), R (R Foundation for Statistical Computing, Vienna, Austria), and many other statistical software packages have the capability of performing all or most of the analyses described later. The names of the recommended computational methods are provided so that the best software options can be selected. Results of computations are provided so they can be checked against results when using specific software.

Descriptive Statistics for Assessing Binary (Presence or Absence) Measures

A first step in assessing symmetry between eyes is to examine the proportion of right and left eyes with the feature of interest. If presence of the feature were perfectly symmetric (i.e., all subjects either have two eyes affected or zero eyes affected), then the proportions of affected left and right eyes would be the same. Ying et al.[1] investigated the symmetry of the retinopathy of prematurity (ROP) between eyes of 1180 premature infants and found that 353 (29.9%) right eyes and 364 (30.8%) left eyes had referral-warranted ROP (RW-ROP). These proportions are nearly equal, consistent with having symmetry between eyes; however, the proportions could be equal even if the agreement between right and left eyes was no more than expected by chance; that is, having equal proportions is a necessary but not sufficient requirement for perfect symmetry.

To assess the agreement in disease status between eyes, the presence of RW-ROP needs to be cross-classified between the two eyes, as in Table 1A. The percentage of infants for whom the disease state is the same in both eyes is 91.6%. Because 30.4% of all eyes of the infants in the study had RW-ROP, we expect that some infants will have disease in both eyes even if the likelihood of one eye having RW-ROP is unaffected by (independent of) whether the contralateral eye has RW-ROP. For the hypothetical data in Table 1B, the percentages of affected right and left eyes are the same as in Table 1A; however, the percentage of infants with the same disease status in both eyes is less, 57.7%. The percentage of infants in Table 1B with RW-ROP in both eyes is 9.2%, equal to the expected percentage if the status of the eyes are independent (29.9% * 30.8% = 9.2%); that is, the probability

Table 1. Cross-Classification of Right and Left Eyes for RW-ROP

	Right Eye RW-ROP		
A. Data from table 7 of the publication by Ying et al.[1]			
Left Eye RW-ROP	No	Yes	Total
No	772 (65.4%)	44 (3.7%)	816 (69.1%)
Yes	55 (4.7%)	309 (26.2%	364 (30.8%)
Total	827 (70.1%)	353 (29.9%)	1180 (100%)
B. Data if right and left eyes are independently affected by RW-ROP			
No	572 (48.5%)	244 (20.7%)	816 (69.1%)
Yes	255 (21.6%)	109 (9.2%)	364 (30.8%)
Total	827 (70.1%)	353 (29.9%)	1180 (100%)

Difference in percentage of right and left eyes with RW-ROP = 0.9%.

95% CI (–0.7%, 2.6%), calculated using the Wilson score method.

Test for equality of proportions:

Large sample test: $\chi_1^2 = 1.22$, $P = 0.27$.

Exact binomial test, $P = 0.31$.

Percent agreement is 65.4% + 26.2% = 91.6%.

95% CI is (89.9%, 93.1%), calculated using the exact method.

95% CI is (90.0%, 93.2%), calculated using the normal approximation method.

95% CI is (89.9%, 93.1%), calculated using the Wilson score statistic method, no continuity adjustment.

κ is (91.6%–57.7%)/(1–57.7%) = 0.80.

95% CI is (0.76, 0.84).

Percent agreement is 48.5% + 9.2% = 57.7%.

of both eyes having RW-ROP by chance is 9.2%. The higher proportion, 26.2%, in the observed data in Table 1A indicates that the RW-ROP is bilateral (symmetric) more often than expected by chance. The κ statistic (κ) is a measure of overall agreement corrected for chance agreement is calculated by $(p_o\text{-}p_e)/(1\text{-}p_e)$, where p_o is the observed proportion of agreement and p_e is the proportion of agreement expected by chance. κ can range from +1 to –1, in which a value of 1 indicates perfect agreement, a value of 0 indicates no agreement, and a value of –1 indicated that there is perfect disagreement between eyes; that is, the disease status is never the same in right and left eyes. For the ROP data in Table 1A, κ is 0.80 ([0.916–0.577]/(1–0.577)). Agreement between the two eyes is high, even after correcting for agreement by chance.

Statistical Tests and Confidence Intervals (CIs) for Summary Statistics for Binary Measures

The description of the case in the earlier text provides summary measures of symmetry between eyes for presence and absence of a condition but does not address whether the measures are statistically significant. Because outcome measures in a person's two eyes usually are related to each other, commonly used statistical tests, such as a Z-test or a χ^2 test, of the equality of two proportions cannot be used because they require the assumption that the outcomes are independent.

To test the equality of the proportions affected between right and left eyes, the McNemar test is used. When available, the exact McNemar test (also known as the binomial sign test) should be used to calculate the P value. When the sum of the counts of people with a right eye status that differs from the left eye status (discordant) is greater than 40, the P value may be approximated by a χ^2 test. Newcombe[2] describes several methods for calculating a 95% CI on the difference in proportions for right and left eyes; one or more of these methods are available in most statistical software packages. The Wilson score method (sometimes referred to as the Newcombe score method) and likelihood profile methods are generally preferred to the Wald method. In Table 1A, the P value associated with the difference in proportions with RW-ROP between right and left eyes is 0.27 using the large sample approximation χ^2 test, and is 0.27 using the exact binomial test, with a 95% CI of (–0.7%, 2.6%), indicating no significance difference.

There is more than one good choice for calculating accurately a 95% CI for the percent of agreement (p). Two recommended methods that are available in several statistical software packages are the Wilson score method and the Clopper-Pearson method.[3] The simple asymptotic formula (also known as the Normal or Gaussian approximation or the Wald method) of $p \pm 1.96*\sqrt{((p*(1\text{-}p))}$, should be reserved for instances when the sample size is large and the estimated proportion in not close to 0 or 1. Because of the large sample size in Table 1A, all of the methods used for the 95% CIs yielded the interval (89.9%, 93.1%) or very similar values. Estimators of the SD of κ estimates, needed for testing whether agreement between eyes exceeds chance agreement and for forming CIs, are available in closed form for large samples and via bootstrapping methods for small samples.[4,5] The CI for κ based on the large sample method is (0.76, 0.84). The agreement between left and right eyes is strong.

Table 2. Cross-Classification of Right and Left Eyes for Stage of ROP.

	Right Eye Stage			
Left Eye Stage	**0**	**1 or 2**	**≥3**	**Total**
0	584 (49.0%)	30 (2.5%)	6 (0.6%)	620 (52.1%)
1 or 2	65 (5.5%)	110 (9.2%)	41 (3.4%)	216 (18.1%)
≥3	9 (0.8%)	52 (4.4%)	294 (24.7%)	355 (29.8%)
Total	658 (55.2%)	192 (16.1%)	341 (28.6%)	1191 (100%)

Percent exact agreement is 49.0% + 9.2% + 24.7% = 83.0%.

κ is (83.0%–40.2%)/(1–40.2%) = 0.71; 95% CI is (0.68, 0.75).

$\kappa_w = 0.86$. 95% CI of (0.85, 0.88), calculated using squared error weights and the large sample variance from Fleiss et al.[4]

Data from table 1 of the publication by Ying et al.[1]

Assessing Symmetry for Ordered Categorical Measures

Features in eyes may be classified into levels of severity or size that are pertinent in assessing symmetry. Ying et al.[1] also classified eyes by stage of ROP, in which stage 0 indicated no ROP (Table 2). Again, a first step is to compare the percentages in each category between right and left eyes (right eye percent, left eye percent). The percentages with each stage are similar and consistent with symmetry between eyes. Agreement between eyes on ROP stage is best summarized by the cross-classification of stage in right and left eyes (Table 2). The percentage of exact agreement is 83.0%. As in Table 1, the percentage of eyes expected in each cell of the table can be calculated under the assumption that the stage of ROP in one eye is independent of the stage in the contralateral eye. For example, the percentage of infants with stage 0 in both eyes is 52.1%*55.2% = 28.8%. The sum of the expected percentages for the three table cells corresponding to exact agreement is 40.2%. The κ statistic for

Investigative Ophthalmology & Visual Science

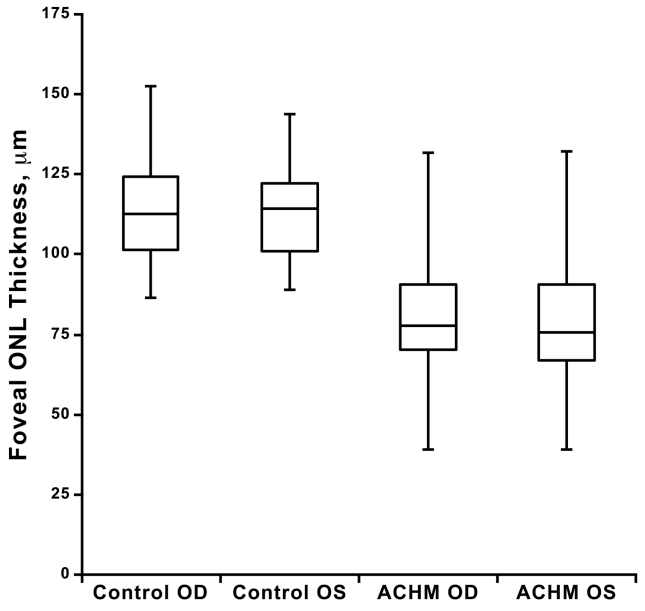

Figure 1. Distributions of foveal ONL thickness for control (N = 42) and achromatopsia (ACHM) groups (N = 76), by right (OD) and left (OS) eye.[8] Available for reuse under a CC BY-NC-ND license.

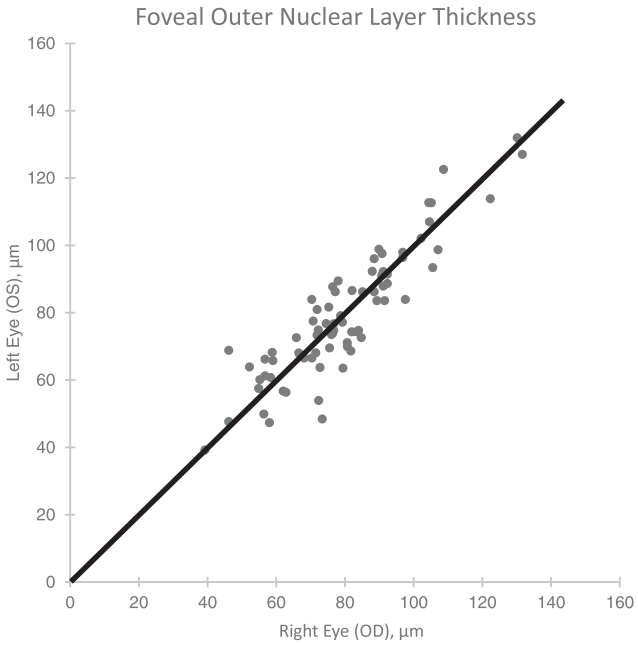

Figure 2. Foveal ONL thickness for right (OD) and left (OS) eyes of individuals with achromatopsia (N = 76). *Solid line* is the line of equality.

exact agreement has the same formula as given in the previous section, (83.0%–40.2%)/(1%–40.2%) = 0.71. Measuring exact agreement treats all disagreements the same; for example, a (right eye, left eye) disagreement of (stage 0, stage 1 or 2) is considered the same as a disagreement of (stage 0, stage 3, and higher). Combinations of exact agreement between eyes are scored or weighted with 1 and all disagreements are weighted with 0. Weighted κ, κ_w, assigns less severe penalties for less severe disagreements; combinations of exact agreement are weighted 1, combinations of the worst agreement are weighted 0, but combinations of intermediate agreement are weighted with values between 1 and 0 and decrease with the severity of the disagreement.[6] Researchers can choose among weighting options based on their perception of the severity of the disagreement. Linear weights and squared-error weights (also known as quadratic weights) are commonly used weighting systems; for 3 ordered categories they would be (1, 0.5, 0) and (1, 0.75, 0), respectively, for the two eyes being 0, 1, or 2 categories apart.[7] Applying the squared-error weights to Table 2 yields $\kappa_w = 0.86$. Using the large sample standard error formulas given by Fleiss et al.[4] provides a 95% CI of (0.85, 0.88). The ROP stage between right and left eyes show a high degree of symmetry.

Assessing Symmetry in Continuous Measures

Often the feature of interest is present in both eyes of all individuals, and the indicator of symmetry is the equality of a continuous measure between right and left eyes. Mastey et al.[8] examined the symmetry between eyes of the outer nuclear layer (ONL) thickness of the fovea measured with optical coherence tomography among patients with achromatopsia. A dataset with the raw values with some key calculations is in Supplementary Table S1. Similar to the analysis described earlier for features measured on a categorical scale, a first step involved comparing the distributions of values in the right and left eyes. Figure 1, extracted from the Mastey et al.[8] publication, shows side-by-side box plots for right and left eyes, both for control individuals and for individuals with achromatopsia. The horizontal bars in these plots show the 75th, 50th (median), and 25th percentiles of the thickness values, with bars extending to the minimum and maximum values in each group. The distributions of the achromatopsia values show that, on average, they are lower than the control values. The five values (maximum, 75th percentile, median, 25th percentile, and minimum) are very similar for the right (OD) and left (OS) eyes in the achromatopsia group. A paired t-test can evaluate whether the mean value in right eyes differs from the mean value in left eyes unless the data are highly skewed.[9] In the Mastey et al.[8] data, the mean (SD) for the right eye was 79.7 µm (18.3) and for the left eye was 79.2 µm (18.7), with a mean difference of 0.5 µm (P value from paired t-test = 0.64). The similarity of the thickness distributions in right and left eyes is consistent with the symmetry between the eyes, but as when evaluating categorical data, cross-classification of the values for the two eyes is required.

A plot of the right versus left eye values for the achromatopsia patients is displayed in Figure 2. If all the right and left eyes had exactly the same thickness, all the points would fall on the 45° line of equality. A Pearson correlation coefficient is a measure of how well the points fit a straight line, and in this case $r_{Pearson} = 0.90$ (95% CI, 0.85–0.94). However, the measure of agreement is the intraclass correlation coefficient (ICC, r_{ICC}) that measures how well the points fit the line of equality. When the difference in means between right and left eyes is zero, $r_{Pearson} = r_{ICC}$. In Mastey et al.[8] data, $r_{ICC} = 0.89$, is very close to 0.90 because the mean difference (0.5 µm) is very small relative to the range of the thickness measurements.

A Bland-Altman graph provides additional information on the symmetry of right and left eyes. The difference between eyes is plotted against the mean of the two eyes for each person.[10] Figure 3 displays a graph from the Mastey et al.[8] publication. The dotted horizontal line extends from

Investigative Ophthalmology & Visual Science

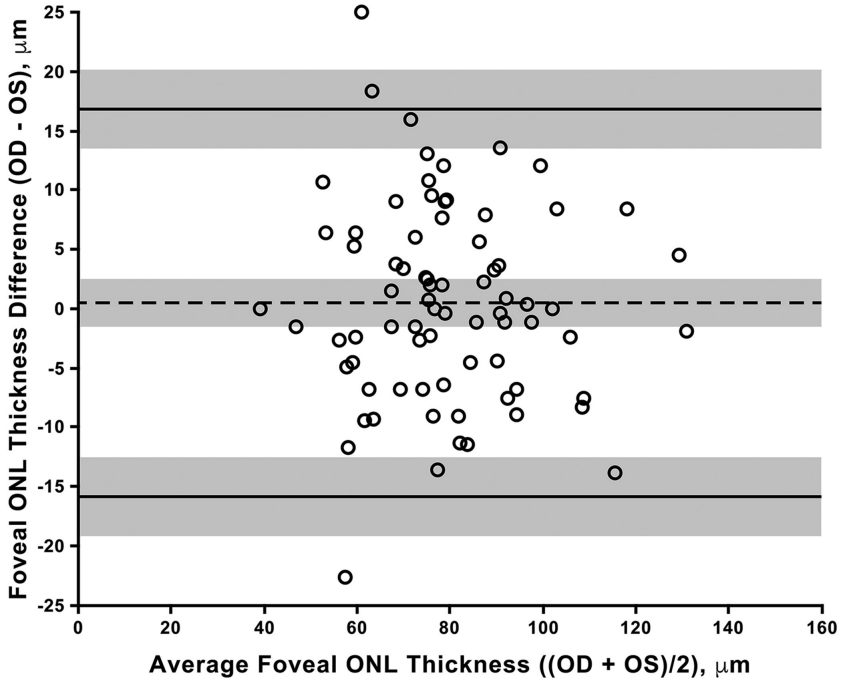

Figure 3. Bland-Altman plot for foveal ONL thickness for right (OD) and left (OS) eyes of individuals with achromatopsia.[8] Interocular symmetry of foveal ONL thickness. Available for reuse under a CC BY-NC-ND license.

the *y*-axis at the value of the mean difference (right eye minus left eye), calculated earlier to be 0.5 µm. The plot is reviewed to assess whether there is a pattern to the differences in thickness between eyes. In the Mastey et al.[8] data, the points appear to be symmetric around the dotted horizontal line, and the spread of the points is similar over the entire range of mean thickness values (*x*-axis). This is typical for ocular data; however, if there are patterns in the difference between eyes, such as increasing or decreasing values with higher values of the mean or increasing differences in both directions with higher values of the mean, transformations of the data may be warranted, as explained by Ludbrook.[11] The Bland-Altman plot also provides quantitative information on the expected magnitude of differences between eyes. The upper and lower horizontal lines in Figure 3 are the 95% limits of agreement. These lines are extended from the *y*-axis at the values corresponding to (the mean of the differences) $\pm$ 2 * SD_{diff}, where SD_{diff} is the SD of the distribution of differences between right and left eyes.[10] They provide estimates of the bounds for the range for 95% of right-left differences. Similar to other statistical estimates, the limits of agreement are subject to sampling variation, and the shaded areas around the limits denote the 95% CI for the limits.[10] These CIs can be determined for large samples by adding and subtracting $t_{(n-1)}(0.975)$ * 1.71* SD_{diff} /$\sqrt{n}$ from the bounds for the limits of agreement, where *n* is the number of individuals and $t_{(n-1)}(0.975)$ is the value corresponding to the 97.5th percentile of a *t* distribution with *n*-1 degrees of freedom. When *n* is small, more complex formulas are required.[10,12] For the Mastey et al.[8] data, the limits of agreement are –15.9 and 16.8 µm. Because of uncertainty in the estimates of the limits of agreement, we may consider the outer bounds of the 95% CIs as the upper bounds for the limits of agreement, extending them to –19.1 and 20.0 µm. In summary, the ICC for the ONL thickness between right and left eyes was (0.89), indicating a high degree of symmetry between eyes. The small mean difference (0.5 µm) and nonsignificant paired *t*-test indicate that neither the left nor the right eye is thicker on average. These data indicate that approximately 95% of differences between eyes will be within approximately 20 µm.

Implications of Symmetry for Study Design

The high intereye symmetry presented in the two articles used as examples is common for ocular characteristics in individuals without ocular disease and in individuals with a wide variety of ocular diseases. When both eyes of an individual are in the *same* comparison group, the additional information provided by the second eye, once the value of the first eye is known, is decreased. In the extreme case when right and left eyes always have equal values ($r_{ICC} = 1$), there is no gain in information from measuring the second eye. Alternatively, when right and left eyes are in *different* groups, the precision of the comparison is increased relative to having the same number of eyes from two different groups of people because all the person-specific characteristics that may affect the measurement are perfectly balanced between groups and removed as sources of variability.

This correlation in measurements from the two eyes of an individual must be considered when calculating sample size for a study. To illustrate this point, we explore the impact of different experimental designs for a study to evaluate a gene therapy for achromatopsia with ONL thickness as the outcome. We assume an α level of 0.05 ($Z_{\alpha/2} = 1.96$), power of 0.80 ($Z_{\beta} = 0.84$), and based on the Mastey et al.[8] study, an SD (denoted by s) of 19 µm among treated and control eyes, a mean of 80 µm in control eyes, and $r_{ICC} = 0.89$. The smallest difference (denoted by d) that we believe is clinically

Table 3. Sample Sizes for a Gene Therapy Trial for Achromatopsia under 3 Different Study Designs

Design	Eyes Per Group*	People Per Group	Total Eyes	Total People
1. 1 eye per person	57	57	114	114
2. 2 eyes per person, both in same group	108	54	216	108
3. 2 eyes per person, eyes in different groups†	7	7	14	7

* Rounded up to the next whole number.
† From the sample size formula for large n for illustrative purposes; exact calculations needed for applications.

meaningful is 10 µm. We use sample size formulas provided by Gauderman and Barlow[13] for large samples.

- Design 1: One eye per patient, the number of eyes per group is $n(1) = [2s^2 (Z_{\alpha/2} + Z_{\beta})^2]/d^2$.
- Design 2: Two eyes per patient, both in the same treatment group, $n(2) = n(1)*(1+ r_{ICC})$.
- Design 3: Two eyes per patient, one in the gene therapy group, the other in the control group, $n(3) = n(1) * (1- r_{ICC})$.

The results of the calculations are displayed in Table 3. In this case with a very high degree of symmetry between eyes, there is little decrease in the number of patients needed for design 2, even though the number of eyes needing evaluation is almost double. For design 3, the number of patients and number of eyes are decreased dramatically. Design 3 is very efficient; however, it can be used only for local (vs. systemic treatments that affect both eyes) treatments and cannot be used when the effects of treatment may crossover to the contralateral eye. These three designs should be evaluated for any study comparing groups of eyes. Particularly when the symmetry is not as strong, as in the case of achromatopsia, and the cost of finding and following an individual is high relative to the marginal cost of evaluating a second eye, design 2 may be less costly than design 1, even though more eyes need to be evaluated.

Conclusions

The methods described in this article help researchers to fully understand the degree of symmetry of features in the eyes of individuals. The degree of symmetry can have a strong effect on the design of experiments and in the choice of statistical methods for data analyses.

Acknowledgments

Supported by a grant from to the Department of Ophthalmology from Research to Prevent Blindness.

Disclosure: **M.G. Maguire**, None

References

1. Ying G-S, Pan W, Quinn GE, et al. Inter-eye agreement of retinopathy of prematurity from image evaluation in the Telemedicine Approaches to Evaluating Acute-Phase ROP (e-ROP) Study. *Ophthalmol Retina*. 2017;1:347–354.
2. Newcombe RG. Improved confidence intervals for the difference between binomial proportions based on paired data. *Stat Med*. 1998;17:2635–2650.
3. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med*. 1998;17:857–872.
4. Fleiss JL, Cohen J, Everitt BS. Large-sample standard errors of Kappa and weighted Kappa. *Psychol Bull*. 1969;72:323–327.
5. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Stat Sci*. 1986;1:54–77.
6. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *EPM*. 1973;33:613–619.
7. Cicchetti DV. Assessing inter-rater reliability for rating scales: resolving some basic issues. *Brit J Psychiatry*. 1976;129:452–456.
8. Mastey RR, Gaffney M, Litts KM, et al. Assessing the interocular symmetry of foveal outer nuclear layer thickness in achromatopsia. *Trans Vis Sci Technol*. 2019;8(5):21.
9. Ganju J. Comment on diagnostics for assumptions in moderate to large simple clinical trials: do they really help? *Stat Med*. 2006;25:1799–1800.
10. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8:135–160.
11. Ludbrook J. Confidence in Altman-Bland plots: a critical review of the method of differences. *Clin Exp Pharmacol*. 2010;37:143–149.
12. Carkeet A, Goh YT. Confidence and coverage for Bland-Altman limits of agreement and their approximate confidence intervals. *Stat Methods Med Res*. 2018;27:1559–1574.
13. Gauderman WJ, Barlow WE. Sample size calculations for ophthalmologic studies. *Arch Ophthalmol*. 1992;110:690–692.

Investigative Ophthalmology & Visual Science