To the Editor:

We read with great interest the important study by Epstein et al. describing the significant variability in corneal curvature measurements in patients with ectatic corneas. Even when using a modern, advanced imager (Pentacam HR; Oculus Optikgeräte GmbH, Wetzlar, Germany), single measurements with 95% confidence interval of a true change in $\Delta K_{\text{max}}$ were 1.51 diopters (D). Repeated measurements in the study by Epstein et al. reduced that to 0.68 D. The authors proposed criteria involving multiple measurements in future studies analyzing the results of collagen cross-linking. We commend Epstein et al. on their work, which serves the dual purposes of establishing measurement criteria for gauging the magnitude of effect delivered by medical and surgical treatments of ectatic corneas and initiating a discussion on the relationship between severity of corneal ectasia and our ability to accurately and consistently measure these corneas.

The study by Epstein et al. analyzes a cumulative data set of patients with keratoconus and postoperative LASIK ectasia and stratifies the variability of measurement accuracy observed based on disease severity. It has also been our experience that the more advanced the ectasia (keratoconus or otherwise), the more variable the metrics produced by corneal topographic and tomographic measurements. Put in statistical terms, the standard error increases with the severity of the disease. Figure 1 is an example of three Pentacam measurements taken of the same eye within minutes of each other. The variation in these three measurements totals 2.4 D. As evident from the image, this patient has rather advanced keratoconus.

Some comments on the statistical methods of Epstein et al.’s study are also in order. The formula for calculating the $R_{\text{min}}$ COMBO appears to be a harmonic mean over standard deviations, but the authors provide no justification for this weighting. Because this formula is not standard in the statistical literature, justification should be provided.

The table appears to confuse the probability of an event with a confidence interval. The authors probably want a tolerance interval for the next observation: that is, constants $a,b$ such that $\text{Prob}(a \leq Y \leq b) = 0.95$ for the next observation $Y$. Furthermore, the table appears to be generated by a repeated measures analysis based on unbalanced group sizes, and therefore the authors need to justify computing tail probabilities under these circumstances. Finally, the table appears to test group mean differences, and not individual differences (ie, within-person variation).

Given the growing importance of corneal collagen cross-linking and other treatment modalities for keratoconus, continued efforts to optimize measurement accuracy and establish reliable metrics for describing abnormally shaped corneas are valuable. We look forward to more research and debate in this area with the hope and expectation that more universally accepted standard measurement criteria will emerge.

REFERENCES


Roy S. Rubinfeld, MD
Roxanne Littner, MS
William B. Trattler, MD
Karoline Rocha, MD
Jonathan Talamo, MD
Richard L. Lindstrom, MD
Waltham, Massachusetts

Drs. Rubinfeld, Littner, Trattler, Rocha, Talamo, and Lindstrom are consultants for and Drs. Rubinfeld, Littner, Trattler, Talamo, and Lindstrom are stockholders in CXL Ophthalmics, Encinitas, California.

Reply

We appreciate the praise of our work from these leaders in ophthalmology, and hope that the use of the conclusions of the article may help lead to more rapid diagnosis of the worsening of keratoconus disease and comparison of alternative treatments. The Pentacam HR (Oculus Optikgeräte GmbH, Wetzlar, Germany) and other Scheimpflug devices allow for analysis of both anterior and posterior corneal curvature. Taking five Pentacam HR measurements rather than one and comparing maximal keratometry reading ($K_{\text{max}}$) averages between office visits is a reasonable, current clinical
approach. The difference in Kmax average can then be computed manually and the clinician can then use the probability table to determine the likelihood that Kmax actually changed. Determining the likelihood that the cornea has changed using both anterior and posterior measurements is more accurate but less practical in a busy office, but the process is enhanced by software that will hopefully become incorporated directly into the imaging device. Our data revealed that the standard error in Pentacam HR-measured Kmax was uncorrelated with measured Kmax magnitude below measured Kmax of 70 diopters (D). But measurement variability sharply rose at average Kmax of 70 D and these cases were excluded from our statistical study. Whether the rise in Kmax variability at 70 D is machine-related or a problem due to poor patient fixation ability in these worst keratoconus cases is yet to be discovered.

The formula for calculating the Rmin COMBO is a weighted average, a linear combination, not a harmonic mean. An alternative expression would be

$$\frac{aR_{\text{min}_F} + bR_{\text{min}_B}}{a + b}$$

where the weight for Rmin_F is $a = 1/STD_F$ and the weight for Rmin_B is $b = 1/STD_B$.

In practice, an unknown true standard deviation (parameter) was estimated by the sample standard deviation. According to Slutsky’s theorem: If $X_n \rightarrow X$ in distribution and $C_n \rightarrow c$ (a constant) in probability, then $C_n X_n \rightarrow cX$ in distribution. This also holds true for the linear combination $C_n X_n + D_n Y_n \rightarrow cX + dY$. Notice that each of the estimated weights using sample standard deviations is a consistent estimator. The authors believe the application of Slutsky’s theorem was standard in the statistical literature.

The table title clearly used the term “probability”; there was no confusion with a “confidence interval” on the authors’ part. A confidence interval was not the intended expression, because a confidence interval is normally established by a give or take number surrounding the sample statistic, which is not a fixed parameter. For the calculation of the probability in the study, if there was no change in the curvature, the difference would truly be fixed at zero. By correctly expressing the value as probability in the table without using the term “confidence interval,” there was no confusion between the probability with a confidence interval.

The table was indeed generated by a double repeated measures analysis. The “group size” referred to in the question is the number of repeated measures performed within each individual per visit (old and new). The actual trial dataset contained $k_{old} = k_{new} = 5$ repeated measures per individual; however, the precision could be derived for numbers other than 5 for each of the two visits, to help physicians understand the loss of precision, from the convenience gained by taking fewer repeated measures ($k_{old}$ or $k_{new} < 5$). Using the within-subject variance estimate $\hat{\sigma}^2_r$ from the double-repeated measures analysis, the formula for the precision calculation (standard error) is:

$$\hat{\sigma}_r \left( \frac{1}{k_{old}} + \frac{1}{k_{new}} \right)$$

To further illustrate additional scenarios with additional values for $k_{old} + k_{new}$ from 1 + 1 to 5 + 5, see Figure 1, which did not appear in the original article.

For the study, the tail probabilities were then generated using the normal approximation. The practice above was also considered standard by the authors in the statistical literature for repeated measures analysis. As explained above, the differences were “within-person” because k repeated measures were performed within an individual per visit.

Robert L. Epstein, MD  
Yi-Lin Chiu, PhD  
Greg Epstein, BS  
McHenry, Illinois

The authors have no financial or proprietary interest in the materials presented herein.

doi:10.3928/1081597X-20130919-02