Measurement Validation: Considerations in Premarket Review

Gene Hilmantel, O.D., M.S.
Clinical reviewer
DONED/ODE/CDRH/FDA

Significant contributions made by
R. Lee Kramm, MD, former FDA reviewer

The author has no financial interest in the subject matter of this presentation.

Overview

- General Considerations
- Bench Testing
- Agreement Studies
- Precision Studies
- Labeling

Regulatory Route for OCT Devices

- 510(k)
  - A 510(k) is a premarket submission made to FDA to demonstrate that the device to be marketed is substantially equivalent (SE) to a legally marketed device (21 CFR 807.92(a)(3))

http://www.fda.gov/medicaldevices/deviceregulationandguidance/howtomarketyourdevice/premarketsubmissions/premarketnotification510k/default.htm

Measurement Validation Studies

- Bench testing
- "Segmentation agreement": Segmentation algorithm validation using clinical images
- Agreement Studies and Precision Studies
  - » These are the clinical studies comparing new device to the predicate
  - » Above clinical studies include measurement of each parameter of interest, analyzed separately in:

    (a) "Normal" eyes, and

    (b) Eyes with pathology of interest (e.g., glaucoma)

Bench Testing

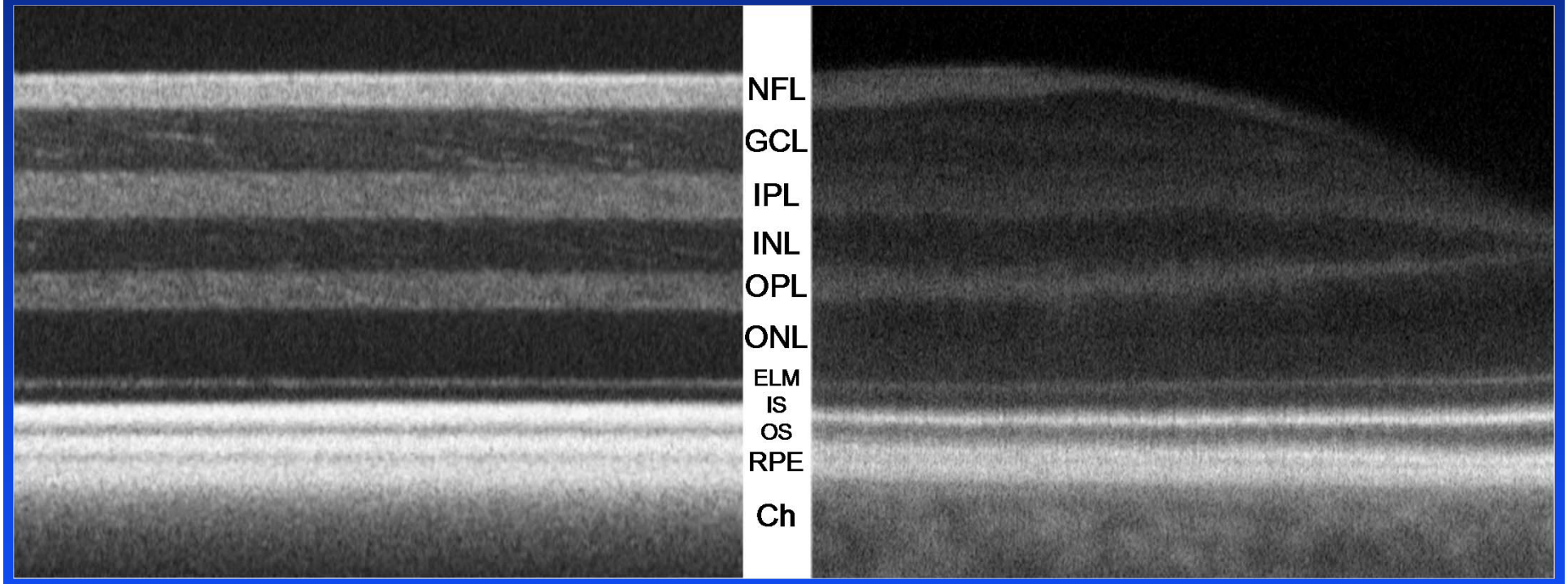
- Resolution testing (axial and lateral)
- Validation of thickness measurements
- Validation of lateral measurements

Development of Retinal Phantom Model for Diagnostic Device Bench Testing

- Would enable rapid and rigorous assessment of OCT image quality and measurement validity, because:
  - The true thickness of each layer is known with high precision
  - Permits evaluation of a range of tissue/layer thicknesses
  - Permits simulation of topography (fovea and ONH), pathology (edema, drusen, lesions, retinal holes, ONH cupping, etc.), and media opacity
- To advance scientific evaluation of diagnostic devices, an anatomically accurate Retina Phantom is being developed in FDA's Optical Diagnostic Devices Lab of the Office of Science and Engineering Laboratories (OSEL) *

* Agrawal et al., "Characterizing the point spread function of retinal OCT devices with a model eye-based phantom" Biomedical Optics Express, 2012.

Comparison: Phantom to Human Retina



Retina Phantom
OSEL/FDA

Normal Human Retina

Courtesy of Jigesh Baxi, Anant Agrawal, and Josh Pfefer, Optical Diagnostic Devices Lab

Segmentation Agreement

- Demonstrates correct identification of anatomical layer boundaries by automated detection algorithm
- In this study:
  - » Images from multiple disease-free eyes are used
  - » Trained experts hand-draw layer boundaries on several individual B-scans from each image
  - » Layer thicknesses (μm) are compared with those from the segmentations generated by the device
    - – Error = absolute value of difference
    - – Errors are presented by measurement zones (mean, median, mode, distribution)

Agreement Studies and Precision Studies

Subject Selection Considerations

- Depends upon what the device measures and claims
- Goal of recruiting subjects representative of the intended range of measurements (e.g., thin to thick) and pathologies
  - » For all parameters: "Normal Eyes" - multiple eyes free of retinal disease or glaucoma
  - » For retinal parameters, e.g., total retinal thickness: "Retinal Disease Eyes" - additional non-glaucomatous eyes with variety of pathologies, such as, diabetic macular edema, wet age-related macular degeneration (AMD), dry AMD, cystoid macular edema, epiretinal membrane, macular hole, etc.
  - » For glaucoma-related parameters, e.g., retinal nerve fiber layer thickness: "Glaucomatous Eyes" - additional eyes spanning range of disease severity

Agreement Studies

- Characterize the difference between measurements from the new device and a comparator (“predicate”) device in a clinical study

Agreement Study
Example of Clinical Study Design

- One eye per subject enrolled
- One measurement per eye
  - » Measurement on both new device and predicate
- Order of testing is randomized
- Subjects recruited from multiple clinical sites
- Study performed for each patient group (normals, retinal disease, glaucoma)

Agreement Studies
Example of Analyses

Analyses performed separately for normal eyes, retinal pathology eyes, glaucomatous eyes:

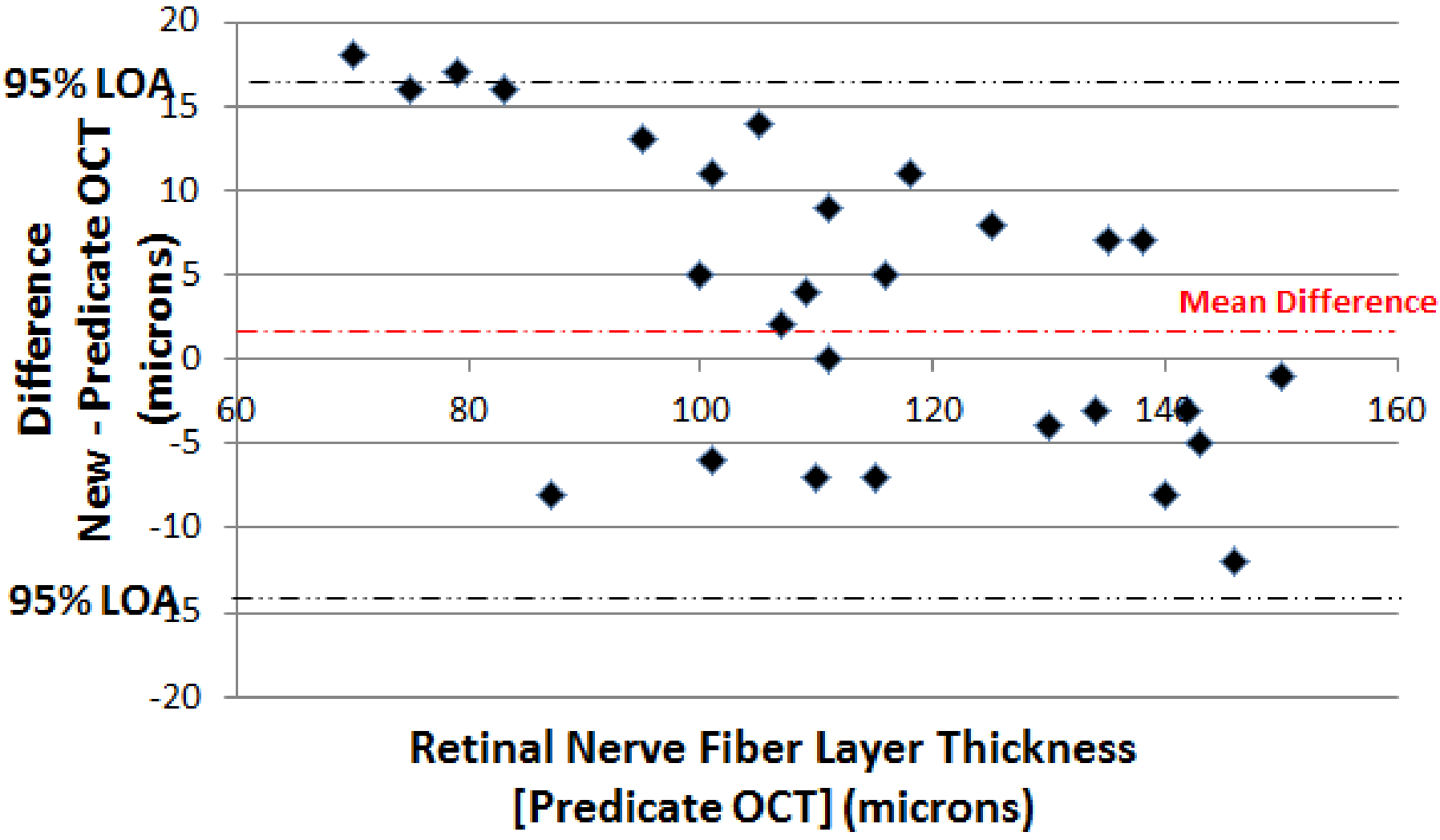
- Number of eyes
- Difference between new device and predicate device
  - mean, minimum & maximum, the standard deviation
- Absolute difference
  - median, min & max, interquartile range
- % Difference [(difference ÷ predicate measurement)×100]:
  - median, min & max, interquartile range

Agreement Studies

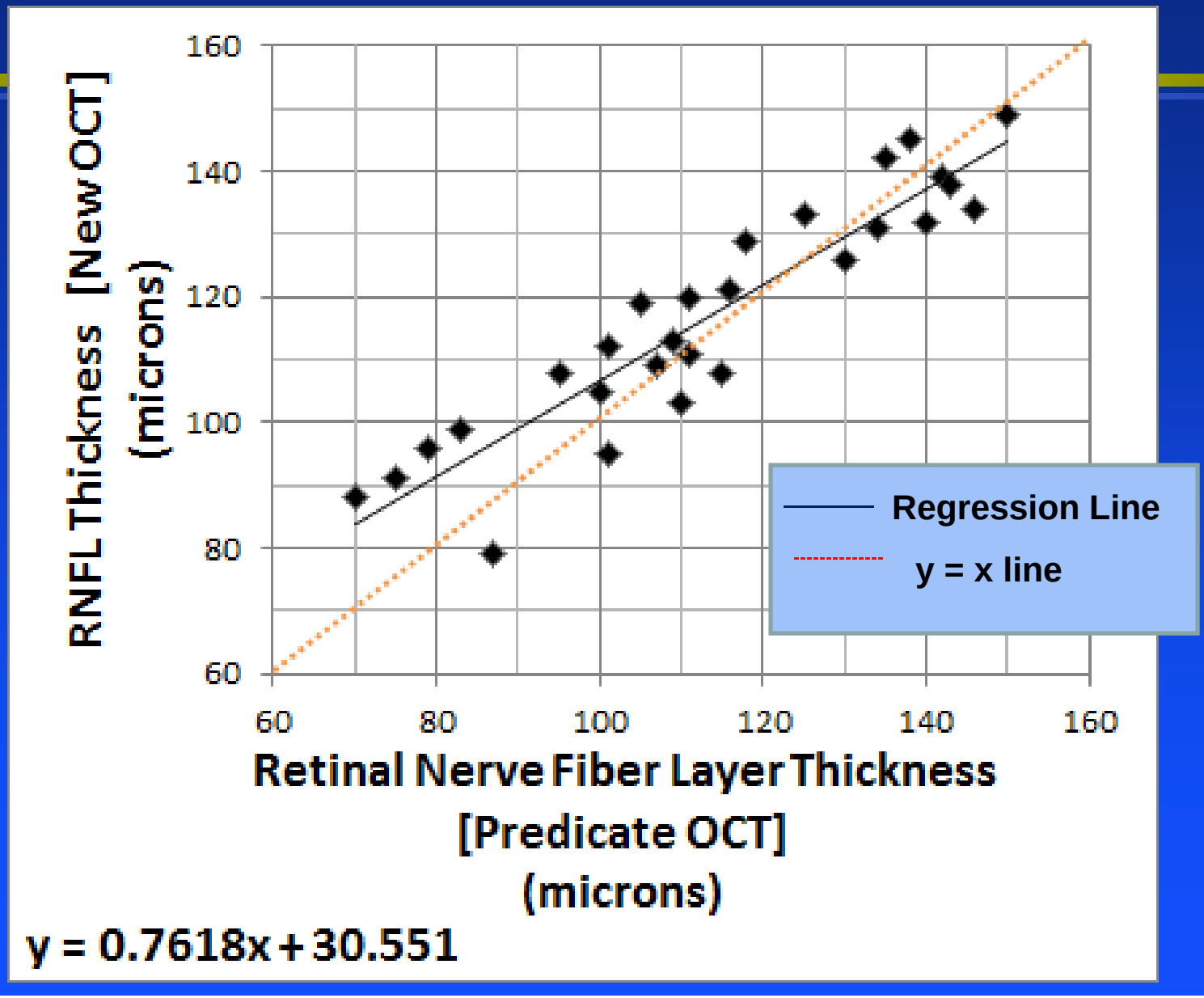
Example of Analyses

- Difference plots (difference vs. predicate)
  - » mean and 95% limits of agreement
- Scatter plots (new device vs. predicate)
  - » regression line and y=x reference line
- Regression
  - » measurement error regression
  - » estimated slope, intercept and 95% confidence intervals

Difference plots
(difference vs. predicate measurement)



Scatter plots (new device vs. predicate)



Example of Agreement Results in the printed labeling

(for each patient category: "normal," retinal disease, glaucoma)

Parameters	New Device Mean (µm)	Predicate Mean (µm)	Mean of Differences (SD)	95% confidence interval for mean difference	95% Limits of Agreement* for differences
RNFL Parameters	number of eyes	number of eyes	number of eyes	number of eyes	number of eyes
Average RNFL Thickness (µm)					
Temporal (µm)					
Superior (µm)					
…					
ONH Parameters					
Cup Disc Ratio					
…					

Example of Agreement Results in the printed labeling *(continued)*

- ***Limits of Agreement:**
  - » for each eye, calculate the difference between the new device result and the predicate device result
  - » Limits are Mean of Differences ± 2 x (SD of the Differences). ~95% of Differences lie between the upper and lower limits.

Questions for Discussion

- Are there more clinically useful ways to describe the agreement results in the User's Manual, e.g., percent differences or differences in microns?
- What difference in agreement, if any, between new device and predicate would cause clinical concern (e.g., if they differed by more than 10%)?

Precision Studies

- Characterize the precision (repeatability and reproducibility) of the new device in a **clinical study**
  - » look for constant variability across range (e.g. thin to thick RNFL measurements)
- Compare precision of new device precision to predicate device in support of substantial equivalence
  - » precision of predicate obtained in same clinical study
  - » side-by-side comparison and ratio of SDs (new divided by old)

Precision Study: Example of Design

- Study uses >2 devices (of a given model) and >2 operators
  - » each device can be used by a single operator
- Each eye is measured on each device
- Each eye is measured >2 times on each device
  - » Patient removed from headrest and repositioned between repeated measurements

Precision Study: Example of a Study Design

Subject *	Operator 1 & Device 1	Operator 2 & Device 2	Operator & Device …
Subject 1	>2 replicates with repositioning	>2 replicates with repositioning	>2 replicates with repositioning
Subject 2	>2 replicates with repositioning	>2 replicates with repositioning	>2 replicates with repositioning
…..			

- This study performed for each patient category in the study (normal, retinal disease, glaucoma)
- Entire above study design is done for new OCT and repeated for predicate OCT

Sources of Variability (Precision Study Design Issues)

- Operator
- Device
  - » Cannot have separate variance estimates for these two factors in this study design
- Other
  - » Day-to-Day Variations?
  - » IOP?
  - » ???

Questions for Discussion

Considering the increased burden due to the greater number of measurements that would be needed:

a. Are there additional sources of variability that should be evaluated in Precision Studies, e.g., day-to-day variability in the glaucoma related measurements?

b. Should studies be designed to allow separate estimates of variability due to the device and due to the operator, or is it acceptable for these sources of variability to be confounded (intermixed), i.e., the study only allows for estimates of device-operator variability?

Precision Studies
Example of Analysis

- For each device, within-subject SD versus within-subject mean plotted to assess constant variability across the measuring range and devices
- Analysis of Variance (Random Effects Model)
  - » analysis of mean differences
  - » estimates of individual variance components
- Precision Analyses – specific calculations:
  - » Repeatability SD = estimated as square root of the residual variance component
  - » Reproducibility SD = estimated as the square root of [the sum of repeatability variance and the variance components of device/operator]
  - » %Coefficient of variation = 100 × SD ÷ Mean
- Separate analysis for each patient group (normal, retinal disease, glaucoma)

Current Presentation of Precision Results

(for each patient category: normal, retinal disease, glaucoma)

Parameters	Mean	Repeatability (SD)	%CV	Reproducibility (SD)	%CV
RNFL Parameters	number of eyes	number of eyes		number of eyes	
Average RNFL Thickness (µm)					
Temporal (µm)					
Superior (µm)					
…					
ONH Parameters					
Cup Disc Ratio					
Disc Area					
…					

Question for Discussion

- As advances in technology continue to be made, we expect the precision to improve. What magnitude of difference (described as ratio of standard deviations, SDs) between the precision of a new device and the predicate would cause clinical concern? For example, would it be acceptable if the precision of the new device were worse than that of the predicate by a factor of 2, 3, or more?

Question for Discussion

1. Are there more clinically useful ways to describe the agreement results in the User's Manual, e.g., percent differences or differences in microns?
2. What difference in agreement, if any, between a new device and its predicate would cause clinical concern (e.g., if they differed by more than 10%)?

Question for Discussion (continued)

3. Considering the increased burden due to the greater number of measurements that would be needed:
    a. Are there additional sources of variability that should be evaluated in Precision Studies, e.g., day-to-day variability in the glaucoma related measurements?
    b. Should studies be designed to allow separate estimates of variability due to the device and due to the operator, or is it acceptable for these sources of variability to be confounded (intermixed), i.e., the study only allows for estimates of device-operator variability?

Question for Discussion (continued)

4. As advances in technology continue to be made, we expect the precision to improve. What magnitude of difference (described as ratio of standard deviations, SDs) between the precision of a new device and the predicate would cause clinical concern? For example, would it be acceptable if the precision of the new device were worse than that of the predicate by a factor of 2, 3, or more?

- Thank you for your attention.