Measurement of Ultrasound Biomicroscopy Images: Intraobserver and Interobserver Reliability

Celso Tello,* Jeffrey Liebmann,* Seth D. Potash,* Henry Cohen,† and Robert Ritch*

Purpose. To evaluate intraobserver and interobserver reproducibility of measurement of images obtained during ultrasound biomicroscopy.

Methods. Four anterior segment images of four normal patients were obtained by a single examiner. The measurements of three independent observers were compared to assess interobserver reproducibility in quantifying the images. Thirteen different anterior segment parameters were measured by each observer on each image. Intraobserver and interobserver reproducibility of measurement were assessed by calculating the coefficient of variation for each individual observer and by using the F test to detect a difference among observers.

Results. Intraobserver reproducibility was high. Interobserver reproducibility for the measured parameters varied considerably and was affected by subjective interpretation of visualized anatomic landmarks.

Conclusions. The optimal parameters for quantitative ultrasound biomicroscopy require refinement. Measurements of alterable parameters are best measured presently by a single observer. Ultrasound biomicroscopy has the potential to elucidate anatomic relationships underlying much anterior segment disease, but caution in interpreting quantitative differences is warranted. Invest Ophthalmol Vis Sci. 1994;35:3549–3552.

High frequency ultrasound biomicroscopy (UBM) is capable of imaging anterior segment structures in vivo at resolutions approaching 50 μm.1–6 The organs surrounding the posterior chamber, previously hidden from clinical observation, can be imaged, and abnormal anatomic configurations or relationships contributing to the etiologies of various anterior segment pathologic entities can be evaluated. To analyze pathologic changes quantitatively, normal relationships and parameters must be determined. For a parameter to be quantitatively useful, its measurement must be reproducible.

The configurations and relative proportions of structures in images obtained by scanning depend on the plane of section, any degree of tilt from perpendicular in the scanning probe, and the distance from the center of the anterior chamber. Thus, there is significant potential for an artifact to confound the interpretation of results. Pavlin et al have described and quantified a series of anterior segment parameters for investigations with the UBM.3 We evaluated three independent observers to assess intraobserver and interobserver reproducibility during measurement of UBM images.

METHODS. Equipment. The original ultrasound biomicroscope developed by Pavlin, Sherar, and Foster is based on 50- to 100-MHz transducers incorporated into a B-mode clinical scanner.1 Different transducer frequencies are used, depending on the region to be imaged. In general, higher frequency transducers are used for fine resolution of more superficial structures and lower frequency transducers are used when increased depth of penetration is necessary. In the present study, a prototype scanner (Zeiss/Humphrey, San Leandro, CA) based on the original biomicroscope was used. This system operates at 50 MHz and provides maximum resolution of approximately 50 μm. A 20-mm eyecup holds the methylcellulose coupling medium. The probe is suspended from an articulated arm to diminish motion artifacts. Lateral distortion is minimized by a linear scan format. Tissue penetration is approximately 4 mm. The scanner produces a 5 × 5 mm field with 256 image lines at a scan rate of 8 frames per second. In this series of patients, a soft contact lens was used to prevent potential corneal injury.7

Patient Examination. Written, informed consent was obtained from all patients using a consent form approved by the Institutional Review Board of the New York Eye and Ear Infirmary. The procedures used in this study were in conformity with the Declaration of Helsinki. Eyes were imaged using a fixation target for the fellow eye to position the eye and to maintain a fixed amount of accommodation. Room lighting was held constant for all scans.

Measurements of 13 different anterior segment parameters defined by Pavlin3 were made using a cali-
FIGURE 1. (above) Yellow line = Corneal thickness; red line = anterior chamber depth.

FIGURE 2. (below) Pink line = Scleral thickness; long yellow line = trabecular meshwork-ciliary process distance, iris thickness at position 1, and iridociliary process distance; short yellow lines = iris thickness at positions 2 and 3; aqua line = iris-zonular distance; purple line = iris-lens contact distance; green lines = angle opening distances at 250 and 500 μm; red intersecting lines = anterior chamber angle.
per provided in the computer software and manipulated by the observer:

1. Corneal thickness (CT): measured from the inner surface of the corneal endothelium to the outer epithelial surface (Fig. 1, yellow line).
2. Anterior chamber depth (AC): measured from the central corneal endothelium to the anterior lens capsule (Fig. 1, red line).
3. Scleral thickness (ST): measured as a perpendicular from the scleral spur to the episcleral surface (Fig. 2, pink line).
4. Trabecular meshwork–ciliary process distance (TCPD): measured as a line extending from a point 500 μm anterior to the scleral spur along the corneal endothelium and dropped perpendicularly through the iris to the most anterior ciliary process seen during scanning in that meridian (Fig. 2, long yellow line).
5. Iridociliary process distance (ICPD): measured from iris pigment epithelium to the ciliary process along the same line as TCPD (Fig. 2).
6. Iris thickness in three zones (IT):
   a. IT-1: along the same line as TCPD (Fig. 2).
   b. IT-2: 2 mm centrally from the iris root (Fig. 2, short yellow line).
   c. IT-3: thickest area near the pupillary margin (Fig. 2, short yellow line).
7. Iris-zonular distance (IZD): measured from the iris pigment epithelium to the zonule at a point just clearing the ciliary process (Fig. 2, aqua line).
8. Iris-lens contact distance (ILCD): measured along the iris pigment epithelium from the pupillary border to the point where the iris physically leaves the anterior lens capsule (Fig. 2, purple line).
9. Angle opening distance (AOD 250 and 500): measured on a line perpendicular to the trabecular meshwork, 250 μm and 500 μm from the scleral spur to the iris stromal surface (Fig. 2, green lines).
10. Anterior chamber angle (Angle): measured with the apex in the iris recess and the arms of the angle passing through a point on the trabecular meshwork 500 μm from the scleral spur and a point on the iris perpendicularly opposite (Fig. 2, red intersecting lines).

**Statistical Analysis.** Reproducibility for each parameter was measured by evaluation of the proportional relationship of the standard deviation of the repeated measures to the mean of those measures (i.e., coefficient of variation (CV)). A CV < 10% was considered indicative of good reproducibility.8

Four different images (four eyes of four patients) were evaluated by each operator. For each image, 15 parameters were measured by each observer on five separate occasions. A measure of intraobserver reproducibility was then obtained by calculating the CV of the results for each of three operators.

Interobserver reproducibility, obtained by evaluation of the differences of the means between observers, was evaluated by an F test in the context of a repeated measures analysis of variance design. P < 0.05 suggests significant differences between observer means. P > 0.05 could represent either interobserver reproducibility or an inability to detect a difference between observers because of small sample size.

**RESULTS.** Intraobserver reproducibility was high (CV ≤ 10%) for all measured parameters with the exception of angle measurement in degrees by observer 1 (Table 1). Interobserver reproducibility, however, was low, demonstrating significant differences (P ≤ 0.05) between examiners for five parameters. (Table 1).

**DISCUSSION.** Assessment of reproducibility is necessary before the quantitative usefulness of ultrasound biomicroscopy can be confirmed. This preliminary study is the first to evaluate the method of parameter measurement of UBM images.

Reliability or reproducibility is a measure of how closely a series of observations match each other.9 Values in Table 1 are the coefficient of variation (%) obtained for each observer. The difference in the values for these parameters was analyzed using a three-way repeated measures analysis of variance design. The intraobserver reproducibility was calculated for each observer and each parameter tested. The differences between the means for each parameter were evaluated using Tukey's Honestly Significant Difference post hoc test and significance was set at P < 0.05.

### Table 1. Intraobserver and Interobserver Reproducibility

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Intraobserver</th>
<th>Interobserver</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT</td>
<td>1.9</td>
<td>0</td>
<td>0.03</td>
</tr>
<tr>
<td>AC</td>
<td>0.5</td>
<td>0</td>
<td>0.3</td>
</tr>
<tr>
<td>ST</td>
<td>2.2</td>
<td>1.8</td>
<td>1.1</td>
</tr>
<tr>
<td>Angle</td>
<td>11.1</td>
<td>5.3</td>
<td>4.5</td>
</tr>
<tr>
<td>AOD-250</td>
<td>9.9</td>
<td>7.5</td>
<td>5.1</td>
</tr>
<tr>
<td>AOD-500</td>
<td>4.7</td>
<td>1.8</td>
<td>2.2</td>
</tr>
<tr>
<td>TCPD</td>
<td>5.6</td>
<td>3.7</td>
<td>6.7</td>
</tr>
<tr>
<td>ICPD</td>
<td>8.3</td>
<td>5.4</td>
<td>3.7</td>
</tr>
<tr>
<td>IT-1</td>
<td>2.8</td>
<td>6.5</td>
<td>2.3</td>
</tr>
<tr>
<td>IT-2</td>
<td>4.4</td>
<td>2.5</td>
<td>2.9</td>
</tr>
<tr>
<td>IT-3</td>
<td>7.1</td>
<td>2.6</td>
<td>2.8</td>
</tr>
<tr>
<td>ILCD</td>
<td>3.3</td>
<td>2.9</td>
<td>3.2</td>
</tr>
</tbody>
</table>

CT = Corneal thickness; AC = anterior chamber depth; ST = scleral thickness; AOD-250 = angle opening distance 250 μm from the scleral spur; AOD 500 = angle opening distance 500 μm from the scleral spur; TCPD = trabecular meshwork ciliary process distance; ICPD = iridociliary process distance; IT-1 = iris thickness at position 1; IT-2 = iris thickness at position 2; IT-3 = iris thickness at position 3; IZD = iris–zonular distance; ILCD = iris–lens contact distance.

Coefficients of variation (%) are given for each observer and each parameter tested.
lidity or accuracy, on the other hand, is the degree to which the obtained measurement reflects the true measurement. Inadequate reproducibility can arise from systematic differences between observers or instruments of measurement or physiological changes in the parameter measured. By using single images, we were able to limit the present study to an evaluation of the error in measurement induced by human interpretation of the images. Our results were not affected by possible measurable changes in ocular parameters related to room illumination or accommodation, scanning technique, or machine-related error.

In this study, individual examiners appear to be internally reliable when obtaining repeated measurements from the same scan on multiple occasions, indicating that intraobserver reproducibility is high. When several examiners are compared to each other, however, differences in measurement appear. This suggests that although one observer may consistently use the same points for caliper placement on repeated examinations, a different observer may choose different reference points when performing a particular measurement. This was particularly true for measurements of iris thickness at positions 2 and 3, iridozonular distance, and iridociliary process distance, all of which require the observer to follow a fairly complicated formula involving at least some subjective interpretation of anatomic landmarks. A difference was also detected between the observers' mean measurements of corneal thickness, despite the low coefficient of variation for each observer.

Potential sources of error exist in this study. Parameters with $P > 0.05$ could represent either interobserver reproducibility of measurement or an inability to detect a difference between observers because of small sample size. In addition, we may have been too strict in our criteria in choosing a CV < 10% as the measure of acceptable reproducibility. Increasing observer experience with this technique may also improve observer-related measurement error. Finally, we did not attempt to determine the validity of UBM measurement.

Despite our findings that problems exist with measurements of the previously described parameters, this does not imply that comparison of information gleaned from two groups of patients cannot be compared. Although the CV may be >10% for any given measurement, statistical analysis may still yield differences between the two, despite a wide distribution around the mean, provided the difference between means is great enough.

In summary, the optimal parameters for quantita-