Agreement and Repeatability of Infrared Pupillometry and the Comparison Method

Brian S. Boxer Wachler, MD,[1] Ronald R. Krueger, MD, MSE[2]

Objective: To evaluate the accuracy and repeatability of the widely used comparison method of measuring pupil size.

Design: Cross-sectional study.

Participants: Fourteen eyes of seven healthy myopic subjects were examined.

Intervention: Two examiners made two repeated measures of pupil diameters of 14 eyes using Rosenbaum card comparison pupillometry and infrared pupillometry. Subjects fixated on a distant visual acuity chart, and pupils were measured under three luminance conditions. The agreement and inter-rater repeatability of both methods were determined.

Main Outcome Measures: Outcomes were pupil diameters, limits of agreement, and coefficient of repeatability of two examiners.

Results: The mean difference between the two techniques ranged from 0.3 to 0.5 mm. The limits of agreement within two standard deviations ranged from 2.4 to 2.8 mm. Coefficient of repeatability ranged from 0.6 to 1.4 mm for infrared pupillometry and 1.0 to 1.2 mm for Rosenbaum pupillometry. Inter-rater repeatability of Rosenbaum pupillometry was consistently pupil diameter biased. Pupil diameters measured with the Rosenbaum method were consistently larger than diameters measured with the infrared technique for both examiners under all luminance conditions.

Conclusions: Results indicate that although the mean difference in techniques was small, the range of the agreement between the Rosenbaum and the infrared techniques was large. The Rosenbaum method consistently overestimated pupil diameters and was subject to inter-rater repeatability bias. Rosenbaum pupillometry may not be appropriate when accurate pupil measurements are required. The results have implications for many clinical trials in ophthalmology, including those evaluating refractive surgery that use Rosenbaum comparison pupillometry. *Ophthalmology 1999;106:319–323*

Interest in comparing pupil diameters dates back to Archimedes (287–212 B.C.), although Galileo (1564–1642) was likely the first to calculate pupil diameters using an entopic method.[1] Today, sophisticated pupillometry video equipment is available, which is based on early infrared technology,[2,3] that can determine diameters to hundredths of a millimeter. Such devices often are used in neuro-ophthalmic evaluation.[4] However, these devices can be expensive, which may explain in part why clinicians use the comparison method of pupillometry. Because the comparison technique is relatively subjective, it is assumed to be less accurate than infrared pupillometry. Despite the prevalent use of the comparison method in ophthalmology, there are only two reports that investigated its accuracy and repeatability[5,6] but without rigorous scientific evaluation.

Originally received: February 9, 1998.
Revision accepted: August 28, 1998. Manuscript no. 98061.

[1] Department of Ophthalmology, Jules Stein Eye Institute, UCLA School of Medicine, Los Angeles, California.

[2] Department of Ophthalmology, Saint Louis University Eye Institute, Saint Louis University, St. Louis, Missouri.

Address correspondence to Ronald R. Krueger, MD, MSE, Cleveland Clinic Foundation, Division of Ophthalmology, 9500 Euclid Ave/A-31, Cleveland, OH 63104.

In refractive surgery, pupil diameters can directly affect visual performance and patient satisfaction. In radial keratotomy, visual acuity and glare were influenced by changes in pupil diameter.[7,8] In photorefractive keratectomy, there is an important relationship between the corneal ablation diameter and the pupil diameter. When the pupil dilates larger than zone of treatment, myopic blur is expected, which could reduce visual function.[9] This mathematical model is clinically supported by the high incidence of halo and glare complaints with small ablation diameters.[10] Even when the ablation zone equals the pupil diameter, off-axis light, such as a street light, may still cause glare.[11] The "Rosenbaum" card is a commonly used device for the comparison method in many ongoing U.S. Food and Drug Administration (FDA) clinical trials in refractive surgery. Because the importance of pupillometry in refractive surgery is becoming more widely recognized, we performed this study to better understand the accuracy and the inter-rater repeatability of the comparison method using a Rosenbaum card.

Materials and Methods

Comparison pupillometry was performed using the Rosenbaum card (Cleveland, OH) using a technique often used in FDA refractive surgery clinical trials. A series of increasing half-circle diam-

Table 1. Pupil Measurements of Each Examiner: Mean (SD)

	Examiner 1		Examiner 2	
Luminance Condition	*Infrared (mm)*	*Rosenbaum (mm)*	*Infrared (mm)*	*Rosenbaum (mm)*
Dim	4.95 (1.08)	5.4 (1.1)	4.82 (0.92)	5.4 (0.7)
Bright	4.09 (1.16)	4.4 (1.0)	4.10 (1.01)	4.7 (0.7)
Very bright	3.22 (0.41)	3.5 (0.6)	2.99 (0.45)	3.7 (0.5)
All conditions	4.09 (1.17)	4.4 (1.2)	4.02 (1.11)	4.6 (1.0)

SD = standard deviation.

eters were denoted by 1.0-mm intervals printed on one edge of the card that were verified for accuracy. The card was held temporal to the patient's eye in the corneal plane. The examiner matched the horizontal pupil diameter with the appropriate half circle. Pupil measurements were made in 1.0-mm steps unless the pupil appeared between two half circles on the card, and a 0.5-mm measurement interval then was made. Infrared pupillometry was performed with the Iowa Pupillometer (Henry Louis, Inc., Iowa City, IA). The device consists of a charge coupled device (CCD) camera tuned for infrared detection and two infrared side lamps for pupil illumination. Baseout prisms are placed over the camera lens so both eyes are displayed on the video monitor. The monitor was hidden so as to not illuminate the subject's eyes during testing. Using two vertical lines on the video monitor, calibration of the pupillometer was performed for each subject before pupil measurements were made. The center of the monitor was used for measurement because distortions in the peripheral monitor may induce measurement error. This device measured the horizontal pupils to 0.01-mm increments, and analysis was performed rounding to the nearest 0.1 mm.

Both eyes of seven healthy subjects, five men and two women ranging in age from 28 to 42 years, were examined. Pupillometry was performed by two examiners using both pupillometry techniques under three luminance conditions. Patients fixated on the visual acuity chart displayed in the VectorVision light box (Dayton, OH). Ambient luminance at the chin rest was 0.42 cd/m^2 (dim). Bright and very bright conditions were created by adjusting a direct light source housed on the pupillometer without room lights, 25.55 cd/m^2, and with room lights, 344.33 cd/m^2. Each examiner made one measurement under each of the three luminance conditions for each eye. The examiners were blinded to each other's measurements. As hippus may have influenced pupil measurements,[12] examiners attempted to measure the largest pupil diameter in the hippus cycle for both infrared and comparison techniques. To approximate conditions of pupil measurements in FDA refractive surgery clinical trials, the examiner made only one measurement per condition in contradistinction to averaging multiple measurements.

Research was conducted in accordance with the tenets of the Declaration of Helsinki and was approved by the institutional review board of Saint Louis University Hospital. Informed consent was obtained from all subjects before participation in the study.

The limits of agreement between both infrared and Rosenbaum pupillometry and the inter-rater reliability of each method were determined using the techniques described by Bland and Altman.[13] The limits of agreement were defined as the mean difference in measurements of two techniques $\pm$ 2 standard deviations. For each pupillometry technique, the inter-rater repeatability was defined as twice the standard deviation of the differences between both examiners' measurements. Plotting the measurement difference between the two methods against the mean pupil diameter can demonstrate any relationship (bias) between the measurement error and the mean measurement. This graphical method is a more appropriate method of determining agreement and repeatability than a correlation coefficient and graph.[13]

Results

Pupil Diameters

Table 1 provides descriptive statistics of the pupil measurements for each examiner. There is a consistent trend of larger pupils measured with the Rosenbaum method.

Agreement

Table 2 shows the limits of agreement for the three luminance conditions and combined conditions. The mean difference in techniques is of small magnitude. The range of error within 2 standard deviations (SD) of the mean is considerably larger. Figure 1 shows the difference in measurements obtained by comparison and infrared pupillometry plotted against the mean measurements for each luminance condition and combined conditions. The random scatter of points on the plots indicates lack of bias in the analysis.

Repeatability

Table 3 lists the coefficients of inter-rater repeatability for each pupillometry technique for the three luminance conditions and for combined conditions. The lower numbers represent better repeatability. For most conditions, the infrared technique was more repeatable than the Rosenbaum method. Figure 2 shows the inter-rater differences plotted against the mean measurements using the infrared pupillometer. Under the very bright condition, examiner 1 tended to overestimate the pupil diameter; otherwise, there was no consistent bias. Figure 3 shows the inter-rater differences plotted against the mean using the Rosenbaum card. The Rosenbaum method appeared to be biased with respect to the pupil diameter being measured.

Table 2. Limits of Agreement* between Infrared and Rosenbaum Pupil Measurements

Luminance Condition	Mean Difference (SD) (mm)	Limits of Agreement (mm)	Range (mm)
Dim	−0.5 (0.6)	−1.7,0.7	2.4
Bright	−0.3 (0.7)	−1.7,1.1	2.8
Very bright	−0.3 (0.6)	−1.5,0.9	2.4
All conditions	−0.4 (0.6)	−1.6,0.8	2.4

* Defined as mean ± 2 SD of the difference between infrared measurements and Rosenbaum measurements.

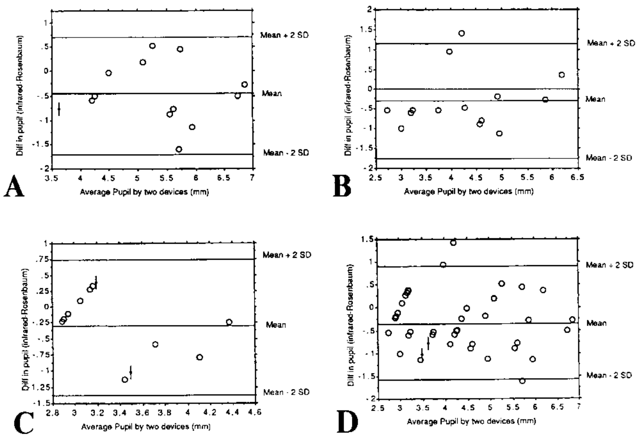

Figure 1. Agreement of infrared and Rosenbaum pupillometry. (**A**) Dim luminance. (**B**) Bright luminance. (**C**) Very bright luminance. (**D**) All luminance conditions. Note the wide range of differences between the two techniques.

Discussion

In this study we have shown that, although the mean difference between Rosenbaum comparison and infrared pupillometry is small, the limits of agreement are large. A range of error from 2.4 to 2.8 mm is considerable. Rosenbaum pupillometry is a form of the comparison method that was first described in 1863 by Follin.[1] It is more subjective than using infrared technology, and when highly accurate pupil measurements are needed, the comparison method will not be appropriate.

This study has shown that the comparison method consistently overestimated the pupil diameter. This result may be explained by the position the Rosenbaum card was held in relation to the subject's eye. If the Rosenbaum card was held posterior to the corneal plane, then the angular minification of the half circles of the card would account for the overestimation. It was assumed the examiners made measurements at the cornea as they were not specifically instructed to do so. This may explain our observation.

These results have implications in the field of refractive surgery. Using physiologic optics, pupils larger than the ablation diameter of photorefractive keratectomy can cause nocturnal glare and halos based on a mathematical model.[11] In one study, 65.6% (21 of 32) of patients treated with a

Table 3. Coefficients of Inter-rater Repeatability* for Infrared and Rosenbaum Pupil Measurements (mm)

Luminance Condition	Infrared Pupillometry (mm)	Rosenbaum Pupillometry (mm)
Dim	1.4	1.0
Bright	1.2	1.2
Very bright	0.6	1.2
All conditions	1.0	1.2

* Defined as 2 × SD of mean difference between examiner 1 and examiner 2. Lower numbers indicate higher inter-rater reliability.

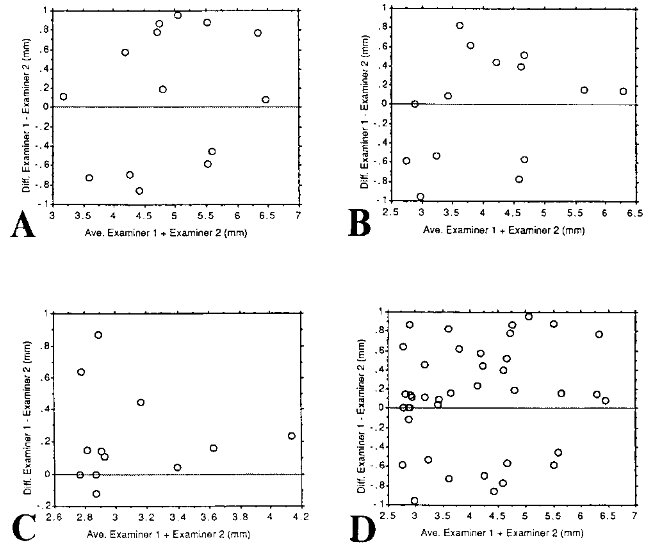

Figure 2. Inter-rater repeatability for infrared pupillometry. (**A**) Dim luminance. (**B**) Bright luminance. (**C**) Very bright luminance. (**D**) All luminance conditions. Under the very bright luminance condition, examiner 1 consistently overestimated the pupil size compared to examiner 2. The random scatter illustrated in all plots indicates the lack of inter-rater bias based on the pupil diameter being measured.

5.0-mm ablation zone reported night halos and glare.[14] As glare has been shown clinically to be related to overall patient satisfaction with photorefractive keratectomy,[15] attention to the patient's pupil diameter is important. We agree that the ablation diameter should be no smaller than the nocturnal pupil diameter to minimize glare and halos.

Since the Rosenbaum card will likely remain prevalent in clinical practice because it is convenient and inexpensive compared to infrared devices, we are recommending a sim-

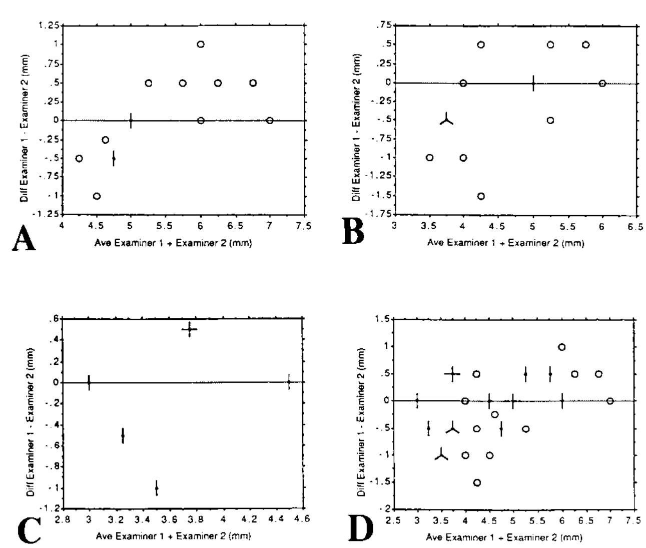

Figure 3. Inter-rater repeatability for Rosenbaum pupillometry. (**A**) Dim luminance. (**B**) Bright luminance. (**C**) Very bright luminance. (**D**) All luminance conditions. Examiner 1 consistently overestimated the pupil size for large pupils while examiner 2 consistently overestimated the diameter for small pupils.

ple conversion factor that accounts for the overestimated pupil size by the Rosenbaum technique. For refractive surgeons, pupils under dim conditions are most important. In Table 2, the comparison under the dim condition indicated the mean Rosenbaum overestimation was 0.5 mm, with a standard deviation of 0.6 mm. The Z table indicates that approximately 79% of patients have overestimated pupil diameters. Subtracting 0.5 mm from the Rosenbaum measurement will decrease the overestimation to 50% of the group. Having equal proportions of overestimation and underestimation seems an appropriate compromise. Therefore, we recommend a 0.5-mm adjustment when using the Rosenbaum card for pupillometry.

By using a theoretical model, a pupil that is larger than the ablation zone may decrease visual function in addition to causing glare and halos.[9] In an unpublished study, we have shown that there is a direct relationship between contrast sensitivity and ablation diameter in photorefractive keratectomy. Even if the ablation diameter is larger than the pupil, a 20% loss of retinal image contrast may still result based on a mathematical model.[16] The results of this study show that if pupillometry is to be used to guide surgical planning, the Rosenbaum method can lead to a significant number of underestimates and overestimates of pupil diameters. Underestimation will lead to halos and glare and reduced contrast sensitivity while overcorrection will result in unnecessary removal of corneal tissue. In laser in situ keratomileusis, the laser ablation is done under a raised corneal flap. Unnecessarily, large ablation diameters will be closer to the endothelium. There also is an increased risk of ablating the hinge of the flap with large ablation diameters, which can result in irregular astigmatism.

One might assume that using a 6.0-mm ablation diameter in photorefractive keratectomy is sufficient to reduce the anulus of blur that results when the pupil dilates greater than the ablation zone. For the population aged 30 to 34 years of age, the mean pupil diameter is approximately 6.0 mm.[17] Younger patients with larger pupils will tend to have greater blur at night from a 6.0-mm ablation. More than 34% (11 of 32) of patients (mean age, 37.6 years; range, 24–52 years) reported disturbances in night vision when treated with a 6.0-mm ablation zone.[14] There appears to be a need for accurately determining the pupil diameter in a dark environment as patients with pupils larger than 6.0 mm should be counseled about the high risk of glare and halos or a larger ablation diameter may be selected.

Infrared pupillometry is ideal to measure the pupil in a darkness, although it is subject to measurement error due to iris physiology. Hippus of the iris and the emotional status of the patient may affect the pupil diameter and thus prevent obtaining the largest natural pupil diameter.[12] However, the optics of the anterior chamber may help minimize this effect. The entrance pupil is 14% larger than the real pupil.[18] If the ablation diameter is matched to the measured pupil, there is inherent latitude that may counter underestimation due to physiologic factors.

We found the inter-rater reliability of the Rosenbaum comparison method was subject to bias of the pupil diameter being measured. Examiner 1 tended to overestimate larger pupils while examiner 2 overestimated the smaller pupils. Infrared pupillometry did not experience this bias. The infrared device provided a magnified image of the pupil, which was measured on a video monitor, and likely allowed for more objective assessment between examiners. Inter-rater bias of Rosenbaum pupillometry has implications for multicenter studies that use this method. As the FDA conducts numerous clinical trials in refractive surgery, the use of the Rosenbaum card likely limits the reliability of pupil measurements obtained from different sites.

This study is not without flaws. The sample size is relatively small, but our analysis was descriptive in nature as we used analytic methods described by Bland and Altman,[13] in which sample size calculations are not appropriate. Although 14 eyes were used, 3 luminance conditions created 42 measurements in each group that were parametrically distributed. We compared the two techniques based on the dynamic pupil that is subject to hippus. Using a static artificial pupil or a pharmacologically dilated one would have eliminated the variability of pupillary hippus; however, such a technique would differ from that used in most refractive surgery trials. Additionally, there may be a potential error in comparison of the two techniques since the increments of the Rosenbaum card were 0.5 mm and the infrared device was 0.01 mm. However, we thought it was important to compare the devices as they are clinically used. There are techniques different than ours that can be used to obtain measurements with the Rosenbaum card; comparative outcomes may be different with other Rosenbaum techniques. Using more than two examiners may have added increased depth to the inter-rater analysis. Although we verified the accuracy of the half circles on our Rosenbaum card, these cards may originate from different sources, and clinicians should verify the accuracy of the half circles before using these cards because they are not printed with standard guidelines.[19]

It is clear why it is not advised to use ablation diameters smaller than the patient's nocturnal pupil diameter. However, it is unclear how much larger ablations should be to maintain visual function and prevent off-axis lights from causing halos. Further study is needed to address these issues.

References

1. Loewenfeld IE. The pupil: anatomy, physiology, and clinical applications, 1st ed. Ames: Iowa State University Press, 1993; v. 1, chap. 15.
2. Lowenstein O, Loewenfeld IE. Electronic pupillography. A new instrument and some clinical applications. Arch Ophthalmol 1958;59:352–63.
3. Loewenfeld IE, Rosskothen HD. Infrared pupil camera. A new method for mass screening and clinical use. Am J Ophthalmol 1974;78:304–13.
4. Bloom PA, Papakostopoulos D, Gogolitsyn Y, et al. Clinical and infrared pupillometry in central retinal vein occlusion. Br J Ophthalmol 1993;77:75–80.
5. Wilson SF, Amling JK, Floyd SD, McNair ND. Determining interrater reliability of nurses' assessments of pupillary size and reaction. J Neurosci Nurs 1988;20:189–92.

6. Lord–Feroli K, Maguire–McGinty M. Toward a more objective approach to pupil assessment. J Neurosurg Nurs 1985;17:309–12.
7. Holladay JT, Lynn MJ, Waring GO III, et al. The relationship of visual acuity, refractive error, and pupil size after radial keratotomy. Arch Ophthalmol 1991;109:70–6.
8. Applegate RA, Gansel KA. The importance of pupil size in optical quality measurements following radial keratotomy. Refract Corneal Surg 1990;6:47–54.
9. Baron WAS, Munnerlyn C. Predicting visual performance following excimer photorefractive keratectomy. Refract Corneal Surg 1992;8:355–62.
10. Seiler T, Wollensak J. Myopic photorefractive keratectomy with the excimer laser. One-year follow-up. Ophthalmology 1991;98:1156–63.
11. Roberts CW, Koester CJ. Optical zone diameters for photorefractive corneal surgery. Invest Ophthalmol Vis Sci 1993;34:2275–81.
12. Miller NR. Walsh and Hoyt's Clinical Neuro-Ophthalmology, 4th ed. Vol 2. Baltimore: Williams & Wilkins, 1985;414–41.
13. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.
14. O'Brart DP, Corbett MC, Lohmann CP, et al. The effects of ablation diameter on the outcome of excimer laser photorefractive keratectomy. A prospective, randomized, double-blind study. Arch Ophthalmol 1995;113:438–43.
15. Hersh PS, Schwartz–Goldstein BH. Corneal topography of phase III excimer laser photorefractive keratectomy. Characterization and clinical effects. Summit Photorefractive Keratectomy Topography Study Group. Ophthalmology 1995;102:963–78.
16. Ludwig K, Schaffer P, Gross H, et al. Mathematical stimulation of retinal image contrast after photorefractive keratectomy with a diaphragm mask. J Refract Surg 1996;12:248–53.
17. Loewenfeld IE. "Simple central" anisocoria: a common condition, seldom recognized. Trans Am Acad Ophthalmol Otolaryngol 1977;83:832–9.
18. Uozato H, Guyton DL. Centering corneal surgical procedures. Am J Ophthalmol 1987;103(3 Pt 1):264–75.
19. Horton JC, Jones MR. Warning on inaccurate Rosenbaum cards for testing near vision. Surv Ophthalmol 1997;42:169–74.