Autorefraction as an outcome measure of laser in situ keratomileusis

Konrad Pesudovs, PhD

Purpose: To determine the limits of agreement between subjective refraction and autorefraction before and after laser in situ keratomileusis (LASIK) to assess whether autorefraction is a valid refractive outcome measure of refractive surgery.

Setting: Ultralase, Leeds, United Kingdom.

Method: The prospective study involved consecutive preoperative normal patients and post-LASIK patients who had autorefraction using the Nidek ARK 700A autorefractor and careful subjective refraction (masked to autorefraction). Inclusion criteria were age greater than 18 years and healthy eyes with a visual acuity better than 0.1 logMAR (6/7.5) with or without previous LASIK. Refractions were compared by spherical equivalent (SE) using Bland-Altman limits of agreement and astigmatic vector difference using the median and the 95th percentile. The effect of time after treatment and treatment strength were explored.

Results: Data were collected from 208 preoperative patients and 237 post-LASIK patients. Preoperatively, the agreement between subjective refraction and autorefraction for the SE was $0.10 \pm 0.35$ (SD) and the median difference for the astigmatic vector was $0.28 \pm 0.39$ D. Post-LASIK, the SE agreement was similar, $0.09 \pm 0.39$ D, but the astigmatic vector agreement decreased slightly with a median of 0.31 D and a 95th percentile of 1.02 D. This decrease reflected poorer agreement in patients whose pre-LASIK refractive error was greater than $4.00 \pm 0.50$ D. Removing this group brought the median astigmatic difference post-LASIK to $0.27 \pm 0.87$ D, similar to that in the preoperative normals. The percentage within $0.50 \pm 1.00$ D of the attempted correction was 56.1% and 78.5%, respectively, with subjective refraction and 51.2% and 78.1%, respectively, with autorefraction.

Conclusions: Autorefraction showed excellent agreement with subjective refraction and was unaffected by refractive surgery except after LASIK for high hyperopia. Most outcomes were correctly classified in the standard categories ($0.50 \pm 1.00$ D), illustrating that autorefraction is a valid outcome measure of refractive surgery.

AUTOREFRACTION AFTER LASIK

does not necessarily require a clinician, this study looked at whether autorefraction could be substituted for subjective refraction as an outcome measure of refractive surgery.

Previous studies suggest that autorefraction is less reliable after refractive surgery than before. However, all the studies have at least 1 methodological flaw. Several studies treat sphere and cylinder refraction values as independent variables. This is not appropriate as the 2 are not independent. Two studies report more negative sphere and more positive cylinder on autorefraction after photorefractive keratectomy (PRK) and laser in situ keratomileusis (LASIK). Although the differences were significant separately, if combined, they would have partly canceled each other. To account for this potential partial compensation of sphere by cylinder, data should be analyzed as spherical equivalent (SE). However, astigmatism data also have to be reported. Several studies treat the cylinder axis as independent of cylinder power or do not report the axis of astigmatism. Astigmatic differences between 2 refractions have to be calculated by vector analysis to give a true difference in power and direction, as for surgical outcomes. Only 1 study of autorefraction after refractive surgery (PRK) uses SE and astigmatic vector differences and reports that the autorefractor measured increased myopia compared with the spherical refraction. However, there was no comparison of postoperative to preoperative variance as this was not the purpose of the study and the autorefractor used is no longer commercially available.

Previous studies also show evidence of collection bias, particularly in sphere or SE data. This problem is evident when the postoperative subjective refraction data are not distributed normally. Data that are highly kurtotic at zero error suggest that subjective refraction was not carefully carried out postoperatively, or at least “plano” was recorded in many cases in which there were small errors (probably within ±0.50 diopter [D]). Therefore, the discrepancy between the 2 measures may indicate that the subjective refraction was less accurate than the autorefraction. This misuse of a gold standard is a major issue in research. A standard must be carefully selected, if used at all, to ensure the study’s results are valid. Previous studies also use methods for assessing agreement between refractions, such as correlation and regression, that are suboptimal (emphasize the mean difference rather than the standard deviation of the differences). The agreement between the 2 measurements would be better compared using the Bland-Altman limits of agreement (LoA) method for parametric data and the median and the 95th percentile for non-Gaussian distributions.

Given these issues, further investigation of the agreement between subjective refraction and autorefraction before and after LASIK was warranted. This is a prospective study with careful subjective refraction (to avoid collection bias); autorefraction using the Nidek ARK 700A, a more recent version of autorefractor than reported in older studies; and statistical analysis of Bland-Altman LoA for SE, median, and 95th percentiles of astigmatic vector differences. Consideration was also given to whether agreement was influenced by other factors including preoperative refractive error and the time since surgery at which postoperative data were collected. This should allow us to determine whether autorefraction is sufficiently accurate to substitute for subjective refraction as a measure of refractive outcome after LASIK.

Patients and Methods

Patients were prospectively recruited from those presenting consecutively for initial preoperative assessment for refractive surgery and for postoperative follow-up at Ultralase, Leeds. Informed consent was obtained from all patients after the nature of the study had been fully explained. The tenets of the Declaration of Helsinki were followed, and the Leeds Regional Ethical Committee approved the study. The study was cross-sectional in design with 2 independent subject groups: preoperative normals and postoperative (LASIK) patients. Inclusion criteria were age 18 years or older, normal healthy eyes with a visual acuity better than 0.1 logMAR (6/7.5) in the preoperative normals, and in the postoperative group, previous LASIK refractive surgery. Exclusion criteria were ocular pathology (including any condition known to interfere with autorefractor performance; eg, asteroid hyalosis) or abnormality including amblyopia and strabismus, previous ocular surgery (other than LASIK in the postoperative group), neurological problem, systemic disease, use of medication that might affect vision, and inability to speak English sufficiently or insufficient mental ability to comply with subjective refraction with confidence.

The preoperative and postoperative examinations included corneal topography, manifest refraction, automated refraction, intraocular pressure measurement, pupillometry
(Colvard), corneal pachymetry, and a complete ophthalmoscopie examination.

Manifest refraction was determined using subjective refraction only. One clinician (K.P.) conducted all subjective refraction and autorefraction. Subjective refraction was performed before autorefraction to maintain masking. However, the clinician was aware of the prescription of the previous spectacles in preoperative cases and previous postoperative refraction in postoperative cases. Subjective refraction was performed using a trial frame, in which loose lenses could be inserted so the lens with the highest refraction was next to the eye at a vertex distance of 12.0 mm. Careful subjective refraction was undertaken by determining the best-vision sphere and the Jackson’s cross-cylinder technique. Changes in cylinder power were compensated for by adjustment of sphere power, but all such compensations were double-checked subjectively. The final cylinder power was defined as the highest cylinder power for which an increase was requested. Each eye was refracted monocularly followed by binocular balancing.

The final spherical power was defined as the highest plus value or the lowest minus value that gave the best visual acuity. All refractive measurements were done without cycloplegia. Manifest refraction was recorded to the nearest 0.25 D sphere, 0.25 D cylinder, and 2.5 degrees.

The autorefractor was measured with the Nidek ARK-700A autorefractor according to the manufacturer’s instructions (Nidek Auto Ref/Keratometer Model ARK-700A Operators Manual, Nidek Co. Ltd., 1999). This device has an autofogging mechanism to relax accommodation. The patient was instructed not to blink and to “stare down the road into the distance at the balloon. The picture will blur in and out of focus. Don’t worry, this is expected; please keep staring into the distance.” The autorefractor has a measurement range from −18.0 to +23.0 D in sphere and up to ±8.0 D in cylinder. Measurements were taken using the auto-tracking and auto-shoot functions, with accuracy set to 0.12 D for power and 1 degree for axis. Five measurements of the eye were taken and the values automatically averaged.

Laser in situ keratomileusis was performed under topical anesthesia using the Technolas® 217 (V2 9997) excimer laser (Bausch & Lomb Surgical) and the Hansatome® microkeratome (Bausch & Lomb Surgical). In all eyes, the corneal flap was 160 μm thick with a 9.5 mm diameter where possible or an 8.5 mm diameter; the optical zone was at least 6.0 mm and increased to 0.5 mm greater than the scotopic pupil for pupils over 5.5 mm.

Subjective and autorefraction data were stored in a spreadsheet and converted to SEs (sphere + 1/2 cylinder) for calculation of spherical differences. The differences in SE are shown as positive if the subjective refraction was more hyperopic (less myopic) than the autorefraction and negative if the subjective refraction was more myopic (less hyperopic) than the autorefraction. Astigmatic differences were determined by vector analysis, which considers the magnitude and the direction of 2 cylinders when calculating their difference. This method for calculating the difference between 2 measurements of astigmatism, in this case subjective refraction and autorefraction, is no different than calculating the difference between preoperative and postoperative astigmatism as in a refractive surgery outcome study, eg, surgically induced astigmatism, or between postoperative and desired astigmatism, eg, difference vector. It is also equivalent to the technique of calculating the vectorial difference between topographical and refractive astigmatism (ocular residual astigmatism [ORA]). The vector difference in astigmatism reported here is mathematically identical to the “difference of the cylindrical corrections” and “astigmatic difference” described by Schimitzek and Wesemann.

### Statistical Analysis

Calculated differences between subjective refraction and autorefraction for SE and astigmatic vectors were exported to the SPSS statistical analysis program (SPSS software, version 10.1). Bland-Altman LoA (mean difference between the 2 methods ± 1.96 SD of the differences) were calculated for the SE. Since astigmatic data are by definition skewed toward zero, the LoA method is not appropriate. The median vector differences and 95th percentile are presented as an alternative. To look at the variance of agreement across a range of measures, the mean of the 2 measures was used. The effect of various cofactors including preoperative and postoperative refractive error, age, and time since surgery was also investigated using the graphic method of Bland and Altman. The percentage of cases that achieved within ±0.50 D and ±1.00 D of the attempted correction was calculated for subjective refraction and autorefraction. Taking subjective refraction as the gold standard, the sensitivity, specificity, and positive and negative predictive values of autorefraction were calculated. Success of matching groups was assessed with an analysis of variance (ANOVA).

### Results

The preoperative normal group included 208 eyes of 208 patients; 55.4% were women (mean age 40.4 years ± 11.8 [SD]). The post-LASIK group included 237 eyes of 237 patients; 60.6% were women (mean age 42.0 ± 10.0 years). In the preoperative group, the mean SE was −2.24 ± 3.68 D (range −12.13 to +8.88 D) and the mean cylindrical power, −0.77 ± 0.90 D (range 0 to −4.25 D). In the post-LASIK group, the mean preoperative SE was −1.48 ± 4.43 D (range −11.00 to +7.63 D) and the mean preoperative cylindrical power, −0.85 ± 0.95 D (range 0 to −7.00 D).
groups were matched for age (ANOVA $F_{1,444} = 2.64, P>.05$), preoperative SE (ANOVA $F_{1,444} = 3.60, P>.05$), and preoperative astigmatism (ANOVA $F_{1,444} = 0.79, P>.05$). The post-LASIK group had less SE, $-0.16 \pm 0.89$ D (ANOVA $F_{1,444} = 93.84, P<.001$), and less astigmatism, $-0.61 \pm 0.69$ D (ANOVA $F_{1,444} = 5.51, P<.05$), than the preoperative group.

In the preoperative normal group, the mean difference between the SE subjective refraction and autorefraction was $-0.10 \pm 0.35$ D (LoA $-0.79$ to $+0.59$ D) (Figure 1). In the post-LASIK group, the mean difference was $-0.09 \pm 0.39$ D (LoA $-0.85$ to $+0.67$ D) (Figure 2). The agreement within the post-LASIK group varied according to the refractive error treated (Figure 3, left). The agreement appeared constant over the myopic treatment range ($n = 139$, $+0.03 \pm 0.32$ D, LoA $-0.60$ to $+0.66$ D) but was worse in those treated for hyperopia ($n = 90$, $-0.30 \pm 0.44$ D, LoA $-1.16$ to $+0.56$ D). The mean difference shift from $+0.03$ D for myopia to $-0.30$ D for hyperopia indicates a tendency for autorefraction to overestimate hyperopia. Further analysis of the hyperopic data revealed good agreement in those treated for hyperopia less than 4.00 D ($n = 38$, $-0.30 \pm 0.26$ D, LoA $-0.81$ to $+0.21$ D) but poor agreement in those treated for hyperopia greater than 4.00 D ($n = 36$, $-0.30 \pm 0.57$ D, LoA $-1.42$ to $+0.82$ D) (Figure 3, right). There was also a small difference in agreement over time. The standard deviation decreased from 0.39 D during the first 30 weeks to 0.30 D thereafter. As seen in Figure 4, most outliers occurred within 30 weeks of surgery, but 86% of the data fall within this time frame. No other factor (age, astigmatism power present, astigmatism axis present, astigmatism power treated, and astigmatism axis treated) had a significant effect on agreement.

In the preoperative normal group, the median astigmatic vector difference between subjective refraction and autorefraction was 0.28 D with a 95th percentile of 0.72 D (Figure 5). In the post-LASIK group, the difference was 0.31 D with a 95th percentile of 1.02 D (Figure 6). Reanalysis after removing the post-LASIK patients who had pretreatment hyperopia greater than 4.00 D yielded a decreased median astigmatic vector difference of 0.27 D, 95th percentile 0.87 D, similar to that in the preoperative normal group. No other factor (age, astigmatism power present, astigmatism axis present, astigmatism power treated, astigmatism axis treated, or weeks since surgery) had a significant effect on agreement. The direction of astigmatic vector difference between subjective refraction and autorefraction was unbiased before (Figure 7) and after (Figure 8) treatment. There were no specific trends in the direction of vector difference in astigmatism in the post-LASIK group.

Figure 1. (Pesudovs) Scatterplot demonstrating the agreement between subjective refraction and autorefraction SE (D) in the preoperative normal group. The linear regression line and 95% individual confidence interval are shown.

Figure 2. (Pesudovs) Scatterplot demonstrating the agreement between subjective and autorefraction SE (D) in the post-LASIK group. The linear regression line and 95% individual confidence interval are shown.
Figure 3. (Pseudovs.) Left: The agreement between subjective and autorefraction in the post-LASIK group mapped as a function of the SE refractive error treated. The lines indicate mean agreement (−0.09 D) and the 95% LoA (−0.85 D to 0.67 D). The agreement appeared constant over the myopic treatment range but was worse for those treated for high hyperopia. The mean difference also shifts from +0.03 D for myopia to −0.30 D for hyperopia, indicating a tendency for autorefraction to overestimate hyperopia. Right: The agreement in those treated for hyperopia only. The solid lines represent the mean (−0.30 D) and 95% LoA (−0.81 D to +0.21 D) in 38 cases treated for hyperopia less than +4.00 D. The dashed lines represent the LoA (−1.42 D to +0.82 D) in 36 cases treated for hyperopia greater than +4.00 D.

Based on the subjective refraction data, the achieved correction was within ±0.50 D and ±1.00 D of the attempted correction in 56.1% and 78.5% of eyes, respectively. Based on the autorefraction data, the proportions were 51.2% and 78.1%, respectively. There were some misclassifications in these data. Using subjective refraction as the gold standard, autorefraction had a sensitivity of 0.80, a specificity of 0.86, a positive predictive value of 0.86, and a negative predictive value of 0.82 for achieved correction within ±0.50 D and

Figure 4. (Pseudovs.) The agreement of subjective and autorefraction SE in the post-LASIK group mapped as a function of time since surgery. The lines indicate mean agreement (−0.09 D) and the 95% LoA (−0.85 D to +0.67 D). There was little difference in agreement over time. While most of the outliers occurred within 30 weeks of surgery, 86% of the data falls within this time frame.

Figure 5. (Pseudovs.) The agreement in astigmatism of subjective and autorefraction in the preoperative normal group mapped as a function of the SE refractive error. The lines indicate median agreement (0.28 D) and the 95th percentile (0.72 D). There was no difference in agreement across the range of refractive error tested.
AUTOREFRACTION AFTER LASIK

This suggests autorefraction is as clinically valid as subjective refraction for the measurement of refractive error. In LASIK patients, the agreement between subjective refraction and autorefraction was almost identical (−0.09 ± 0.39 D, LoA −0.85 to +0.67 D). This is much better than the agreement in the only valid study of autorefraction (Nikon NRK-8000) after refractive surgery (−0.78 ± 0.91 D, LoA −2.56 to +1.00 D). Therefore, autorefraction with the Nidek ARK-700A is highly accurate at predicting subjective refraction SE before and after LASIK. However, after LASIK for myopia, the mean difference between subjective refraction and autorefraction was +0.03 D for myopic treatments and −0.30 D for hyperopic treatments. Therefore, the accuracy of autorefraction after hyperopic LASIK could be improved by adjusting the autorefraction result by 0.25 D. The postoperative group also showed poorer agreement in those treated for high hyperopia (LoA −1.16 to +0.56 D). The difficulty in measuring autorefraction after hyperopic LASIK may be due to the reduced effective optical zone size and the associated increased corneal asphericity. This causes refractive power to vary across the pupil area, so a difference in sampling location between subjective refraction and autorefraction would lead to a difference in results. This is of little clinical significance as many surgeons limit the level of hyperopia they are willing to treat or do not treat hyperopia. Time since surgery also has a small effect on the agreement. However, the improved agreement seen after 30 weeks (SD 0.30 D)
may be a spurious finding based on a small sample (n = 34, 14% of the data) since the agreement before 30 weeks was the same as in the normal group (SD 0.39 D). No other factors that significantly influenced agreement between subjective refraction and autorefracton were identified.

Autorefraction measurement of astigmatism was clinically significantly different than subjective refraction in the preoperative normal (median 0.28 D) and the post-LASIK (0.31 D) groups. There was also a decrease in agreement between subjective refraction and autorefracton for astigmatic power after LASIK (95th percentile 1.02 D compared to 0.72 D for preoperative normals). Since the patients treated for hyperopia greater than +4.0 D had more variable SE data and astigmatic data (Figure 6), they were removed and the astigmatism data reexamined. The 95th percentile (and the median) in the post-LASIK group improved to almost match the preoperative normals. This poorer agreement in the high hyperopic subgroup is again likely due to sampling difference problems caused by the smaller effective optic zone size and increased asphericity after high hyperopic ablations. There was no bias in the direction of vector difference in astigmatism between subjective refraction and autorefracton from the preoperative normals to the post-LASIK group. No other factors that significantly influenced astigmatic agreement between subjective refraction and autorefracton were identified. Again, this was better than the results in the only previous valid study of autorefracton after refractive surgery, which found −0.66 ± 0.92 D.9

These findings that LASIK has little deleterious effect on the agreement between subjective refraction and autorefracton is at odds with findings in previous studies, which suggest there is an unacceptable disagreement between autorefracton and subjective refraction after PRK and LASIK. However, these conclusions are not reliable because the sphere and cylinder data were analyzed, inappropriately, as independent variables rather than as SEs. It is not known whether the disagreement would have been as large if the studies had reported SEs, but the opposite signs of the sphere and cylinder differences suggest not. Similarly, previous studies reportsing unacceptable disagreement of astigmatism results are unreliable because cylinder power and axis were treated as independent variables. It is not known whether the astigmatic differences would be significant if they were calculated as vector differences. Previous subjective refraction and autorefraction disagreements should also be discounted due to evidence of collection bias of the data. This study corrects the previous flawed reports that autorefraction is not reliable after refractive surgery.

The key question addressed by this study was whether autorefraction is a valid outcome measure for refractive surgery. The tight LoA between subjective refraction and autorefraction provide empirical support for this. However, using subjective refraction as the gold standard, the classification of outcomes by autorefraction into percentages that achieved within ±0.50 D and ±1.00 D of the attempted correction was tested. The sensitivity and specificity of autorefraction as a substitute for subjective refraction was very high, but there were some false positives and false negatives, especially within ±0.50 D. Because the subjective refraction and autorefraction SE data were normally distributed, the misclassifications as false negative and false positive were roughly equivalent. The percentage achieving within ±0.50 D and ±1.00 D of attempted correction were virtually identical whether taken from subjective refraction or autorefraction data. This suggests autorefraction is an adequate surrogate for subjective refraction as an outcome measure of refractive surgery. It is worth noting that the percentages achieved of attempted correction reported in this study were relatively low. This can be explained by the composition of the patient cohort, which included one-third treated for hyperopia and half of those for greater than +4.0 D.

These data do not demonstrate the superiority of autorefraction over subjective refraction. Therefore, subjective refraction should remain the standard method for determining the refractive outcome of refractive surgery. However, in nonideal circumstances, when subjective refraction data may be incomplete or for other reasons, autorefraction could be substituted with confidence as the refractive outcome measure. It is widely accepted that autorefraction is not suitable to substitute for subjective refraction for prescribing spectacles and tends to be used as a screening test to provide a starting point for subjective refraction. Human testing has advantages over autorefraction in that additional procedures such as binocular balancing and measurement of oculomotor coordination improve on refraction toward information necessary for prescribing. Refraction and
prescribing are different concepts; the latter involves a thought process that considers the previous prescription, the likelihood of the new prescription being tolerated, and the needs of the patient. For these reasons, refraction alone, whether subjective or automated, cannot substitute for prescribing. However, autorefraction can substitute for subjective refraction as an outcome measure of refractive surgery.

References

From Ultralase, Leeds, United Kingdom.


Supported by Ultralase, Leeds, United Kingdom. Dr. Pesudovs is supported by NHMRC Sir Neil Hamilton Fairley Fellowship 0061, Canberra, Australia.

The author has no proprietary or financial interest in any product mentioned.