# Reproducibility of sphero-cylindrical prescriptions*

Graeme E. MacKenzie

Vision for the Developing World, Nuffield Laboratory of Ophthalmology, University of Oxford, Walton Street, Oxford, UK

**Abstract**

*Purpose:* To investigate the reproducibility of the sphero-cylindrical prescriptions provided by 40 optometrists.

*Methods:* Forty registered optometrists were randomly selected from the cities of Oxford and Westminster in the UK to perform a sphero-cylindrical refraction on an asymptomatic 29-year-old male subject. The 95% limits of reproducibility for each component of refraction were assessed and are presented together with scatterplots, distribution ellipsoids and polar profiles of dioptric power.

*Results:* The mean stigmatic (spherical equivalent) refraction for the right eye was −0.83 D (S.D. = 0.28 D) with 95% limits of agreement −1.38 and −0.28 D. The 95% reproducibility limits for stigmatic data $[1.96(\sqrt{2})(\text{S.D.})]$ was 0.78 D. The average inter-ocular difference in the stigmatic components of refraction was −0.044 D (S.D. = 0.20 D) but estimates ranged from −0.50 to +0.50 D. Mean ortho- and oblique antistigmatic refractions were −0.23 D (S.D. = 0.084 D) and −0.14 D (S.D. = 0.086 D) respectively.

*Conclusions:* The findings of this paper suggest that refractions performed by multiple optometrists on a single eye will differ in their stigmatic component by over 0.78 D on average not more than once in 20 refractions. The reproducibility of refractions reported here, approximately twice as variable as those reported under repeatability conditions, has profound implications for the analysis of refractive data collected by multiple optometrists over the course of replication, longitudinal and epidemiological studies.

**Keywords:** dioptric power, refraction, refractive error, reproducibility

## Introduction

The performance of any new refraction device is typically judged on the extent to which it agrees with the outcome of the sphero-cylindrical subjective refraction (Bannon, 1977; Zadnik *et al*., 1992; Johnson *et al*., 1996; Elliott *et al*., 1997; Bullimore *et al*., 1998; Salchow *et al*., 1999; Walline *et al*., 1999; Chat and Edwards, 2001; Mallen *et al*., 2001; Dave and Fukuma, 2004; Farook *et al*., 2005; Choong *et al*., 2006). Implicit in comparative studies of this sort is the assumption that the subjective refraction provides an estimate of refractive state that is at once accurate and precise. Indeed, those familiar with the literature on the subject are likely to cite the study by Rosenfield and Chiu (1995) who reported the 95% limits of agreement for subjective refraction to be ±0.29 D. Their findings suggest that the subjective refraction is accurate to about a quarter dioptre and that a change in the magnitude of a prescription of 0.50 D or more should be viewed as clinically significant. Unfortunately, a patient is likely to experience this sort of precision only if they are in the habit of consulting the same eye-care professional year after year: a practice that allows one to benefit from predictably low intra-examiner variability. What would happen if this same person were to consult another equally skilled eye-care provider? To what extent will a

*Received*: 17 August 2007
*Revised form*: 30 October 2007, 18 December 2007
*Accepted*: 2 January 2008

*Correspondence and reprint requests to*: Dr Graeme E. MacKenzie.
Tel.: +44 1865 248996; Fax: +44 1865 794508.
E-mail address: graeme.mackenzie@eye.ox.ac.uk

*Aspects of this work were presented at The Inaugural World Congress on Refractive Error and Service Development on 15 March 2007 in Durban, South Africa.

 doi: 10.1111/j.1475-1313.2008.00549.x

different eye-care professional's refraction agree with previous findings? While this knowledge may prove interesting to a person wishing to gain deeper insight into the reliability of the refraction in general, the issues raised by these questions are central to the analysis and interpretation of refractive data collected by multiple examiners during replication, longitudinal and epidemiological studies.

To fully appreciate these issues one must consider in greater detail the concept of precision; that is, the closeness of agreement between independent test results, and the manner in which it manifests itself in studies on intra- and inter-examiner variability. An experiment designed to evaluate intra-examiner variability is typically conducted under what the International Organisation for Standardisation (ISO) refers to as repeatability conditions: 'where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time' (ISO 1994). Evaluation of inter-examiner variability, however, requires data collected under reproducibility conditions: 'where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment' (ISO 1994). Repeatability and reproducibility thus represent two extremes of precision describing the minimum and maximum expected variability in test results respectively. The key difference between these two methodologies is that repeatability studies require all measurements to be made by one examiner, whereas reproducibility studies require more than one examiner to collect data. Unsurprisingly, the validity and statistical power of the findings from either type of study is a function of the number of measurements taken and examiners used.

Much of the literature on the variability of refractive data comes from studies specifically designed to evaluate the repeatability of auto-refractors (Rubin, 1993; McKendrick and Brennan, 1995; Rosenfield and Chiu, 1995; Elliott *et al*., 1997; Harvey *et al*., 1997; Bullimore *et al*., 1998; Salchow *et al*., 1999; Walline *et al*., 1999; Chat and Edwards, 2001; MacKenzie *et al*., 2001; Mallen *et al*., 2001; Raasch *et al*., 2001; Dave and Fukuma, 2004; Pesudovs and Weisinger, 2004; Sheedy *et al*., 2004; Farook *et al*., 2005; Choong *et al*., 2006). While a number of these studies (Rosenfield and Chiu, 1995; Elliott *et al*., 1997; Walline *et al*., 1999; Raasch *et al*., 2001; Sheedy *et al*., 2004) have provided valuable insights into the repeatability of the subjective refraction itself, it is worth noting that the findings of these studies are based on only two (Elliott *et al*., 1997; Walline *et al*., 1999; Raasch *et al*., 2001; Sheedy *et al*., 2004) or, at the very most, five (Rosenfield and Chiu, 1995) distinct measures of subjective refraction per subject. Similarly, while there are a number of studies that could be classified as reproducibility studies (Sloan *et al*., 1954; Perrigin *et al*., 1982; Zadnik *et al*., 1992; Johnson *et al*., 1996; Bullimore *et al*., 1998; Chat and Edwards, 2001; Sheedy *et al*., 2004; Leinonen *et al*., 2006), they are based on subjective refraction data collected from only two (Zadnik *et al*., 1992; Johnson *et al*., 1996; Bullimore *et al*., 1998; Chat and Edwards, 2001; Sheedy *et al*., 2004; Leinonen *et al*., 2006) or, at most, three (Sloan *et al*., 1954; Perrigin *et al*., 1982) examiners (coincidentally, Bullimore *et al*. (1998) did find a significant difference in the outcomes of two refractionists despite similarities in their education, training and certification). In addition to the limitations associated with using so few measures of subjective refraction from so few examiners, the analyses in these studies are based on the assumption that variation in refractive data is uniform across all ages and ametropias. The validity of many of these studies is further undermined by the fact that the clinicians were not masked to the results of previous subjective refractions or spectacle prescriptions. Another limitation in many of the earlier studies was the manner in which astigmatism was analysed: it was either completely ignored or erroneously analysed in terms of cylinder power and axis. Fortunately, the methods that allow one to study the important interactions between the astigmatic and non-astigmatic components of refraction have been developed (Fick, 1972, 1973; Long, 1976; Keating, 1982, 1986; Harris, 1991a, 1999, 2001; Thibos and Horner, 2001) and are now widely used by clinical scientists (Rubin, 1993; McKendrick and Brennan, 1995; Elliott *et al*., 1997; Harvey *et al*., 1997; Chat and Edwards, 2001; Koch, 2001; MacKenzie *et al*., 2001; Mallen *et al*., 2001; Naeser and Hjortdal, 2001; Raasch *et al*., 2001; Dave and Fukuma, 2004; Pesudovs and Weisinger, 2004; Sheedy *et al*., 2004; Farook *et al*., 2005; Gillan, 2006; Leinonen *et al*., 2006). The purpose of this paper is to evaluate the reproducibility of the sphero-cylindrical prescriptions using data collected from 40 independent and fully qualified optometrists.

## Methods

### *Sampling*

Simple random sampling was used to select 40 optometry practices from a pool of 90 located in the cities of Oxford and Westminster in the UK. Optometrists within these practices were then randomly selected from the National Health Service and General Optical Council registers so as to ensure that only one optometrist per practice was included in the study. The number of practices sampled was determined using the method of sample size estimation for reproducibility studies recommended by the ISO (International Organisation for Standardisation, 1994). The object is to determine the range $[-P,P]$ within which an estimate [standard deviations (S.D.)] of the true S.D. ($\sigma$) of a set of independent measurements lies with a specified level of probability (arbitrarily set at 95%) such that

$$P\left(-A_{\mathrm{R}} < \frac{S.D. - \sigma}{\sigma} < A_{\mathrm{R}}\right) = P.$$

The estimated S.D. is thus expected to lie within $A_{\mathrm{R}}$ percent of the true S.D. with a certainty of 95%. $A_{\mathrm{R}}$ is calculated using

$$A_{\mathrm{R}} = 1.96\sqrt{\frac{p[1 + n(\gamma^2 - 1)]^2 + (n - 1)(p - 1)}{2\gamma^4 n^2 p(p - 1)}},$$

where $n$ is the number of independent measurements taken at each practice ($n = 1$), $p$ is the number of practices sampled and $\gamma$ is an estimate of the ratio of the reproducibility to repeatability S.D.s (ISO 1994). It was assumed that the reproducibility S.D. would be approximately twice as large as the repeatability S.D. published elsewhere (Jennings and Charman, 1973; Kratz and Flom, 1977; Rosenfield and Chiu, 1995). Sampling from 40 practices would thus yield an estimated S.D. that would lie within 22.19% of the true S.D. with a certainty of 95% (doubling the number of practices sampled reduces the uncertainty of the estimated S.D. by only 6.5%).

### *Data collection*

An asymptomatic 29-year-old male subject was examined by 40 randomly selected optometrists, each of whom was masked to the nature of the study (Grant *et al.*, 2002). The subject was of excellent health, non-diabetic, free of systemic and ocular disease and had bilateral unaided and best corrected visual acuities of 6/7.5 and 6/4.5 respectively. British optometrists are licensed professionals trained to the high standards set by the UK optical regulatory body, the General Optical Council. The subject underwent a full ocular examination and submitted to any and all procedures deemed necessary by the optometrist. All but two of the optometrists consulted during the course of this study determined the subject's refractive state by means of trial frame refraction. The trial frame refraction is the preferred method of refraction in the UK and it offers a degree of subject-masking not achievable by phoropter-based refractions. Whereas it is conceivable that a particularly observant subject could monitor the order and power of the lenses presented during the course of a phoropter-based refraction, the potential for introducing this sort of bias during trial frame refraction, where the order of lens presentation occurs in a non-predictable manner, is considerably limited. Two optometrists used automated phoropters to determine the subject's refractive state. In all cases, the final prescription was determined by means of a trial frame and loose trial lenses. Refractive data were transcribed directly from a certified copy of the sphero-cylindrical prescription each optometrist is required by law (Office of Public Sector Information, 1989) to provide. The examinations took place within a 3-week period and all examinations were paid for in full at the time of the consultation. In an attempt to prevent inter-assessor bias and ensure measurement independence, optometrists were not given details of previously determined refractive outcomes.

### *Data analysis*

Variability of refractive data should necessarily be studied in a way that fully accounts for both the astigmatic and non-astigmatic components of refraction. To this end all sphero-cylindrical refractive data for the subject's right eye was transformed (Fick, 1972, 1973; Long, 1976; Keating, 1982, 1986; Harris, 1991a, 1999) into dioptric power format (Fick, 1972, 1973; Harris, 1991a) and analysed in symmetric dioptric power space. This representation of power allows one to study all sphero-cylindrical prescriptions in terms of three independent powers: a stigmatic power and two Jacksonian powers. The stigmatic power referred to here is identical to the 'spherical equivalent' and the Jacksonian powers, one with its power meridians along 90° and 180°, the other with its power meridians along 45° and 135°, are known as the ortho- and oblique antistigmatic powers. These two powers, sometimes referred to as $J_0$ and $J_{45}$, respectively, completely characterise the astigmatic nature of each refractive outcome. Not only is it possible to perform meaningful statistical analyses on each of these components of power, the dioptric power format has the added advantage of allowing one to evaluate the covariance between component powers. Formal hypothesis testing on the variance–covariances and means was performed using multivariate methods published elsewhere (Harris, 1990, 1991a,b). For a more detailed description of the representation and analysis of astigmatic data in symmetric dioptric power space the reader is referred to the work of Harris (1991b, 1999, 2005). The distribution of refractive data for the left eye (not presented here) was similar to that of the right.

Residuals were calculated by subtracting from each refractive outcome the sample mean. The residuals and their S.D. were used to calculate the 95% limits of agreement [±1.96(S.D.); Altman and Bland, 1983; Bland and Altman, 1986] for the stigmatic and antistigmatic components of refraction. The absolute difference in the component powers of the refractions reported here are expected to lie within the boundaries of the region spanned by these limits with a probability of 95%. The 95% reproducibility limit for stigmatic data was calculated by multiplying the absolute value of the 95% limit of agreement by the square root of two $[1.96(\sqrt{2})(\text{S.D.})]$. The reproducibility limit is the value less than or equal to which the absolute difference between any two test results obtained under reproducibility conditions may be expected to be, with a probability of 95%. Put differently, the reproducibility limit may be interpreted as the maximum expected difference in measures of refractive state collected by any two optometrists. Data analyses were performed using Matlab® version 7.3 (The MathWorks, Inc., Natick, MA, USA) and SPSS® version 14.0 (SPSS Inc., Chicago IL, USA).

## Results

The mean stigmatic refraction for the right eye was −0.83 D (S.D. = 0.28 D) with 95% confidence interval −0.92 to −0.74 D. The maximum and minimum estimates of the stigmatic component of refraction differed by a dioptre and all outcomes lay in the range −1.25 to −0.25 D. The residuals of these data over the mean stigmatic refraction are shown in *Figure 1*. The region spanned by the 95% limits of agreement for these residuals extended from −0.55 to +0.55 D and the 95% confidence intervals for these lower and upper limits were −0.71 to −0.40 D and +0.40 to +0.71 D respectively. The 95% reproducibility limit for the stigmatic data is 0.78 D.

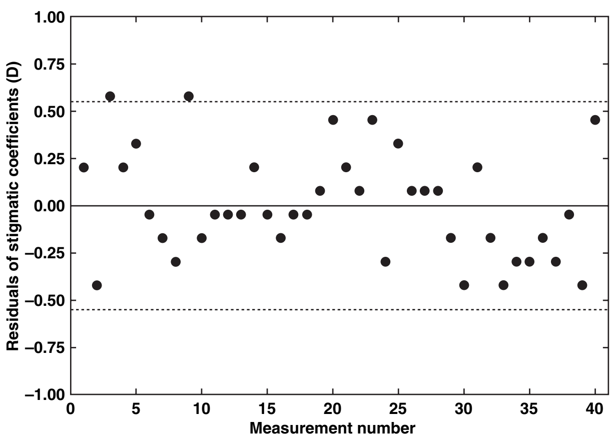


**Figure 1.** The 95% limits of agreement for the stigmatic components of refraction for a single eye as determined by 40 optometrists. The residual of each measurement over the mean stigmatic refraction ($F_I$ = −0.8281 D) is plotted against the vertical axis. The solid line represents the mean residual and broken lines represent the upper and lower limits of agreement at the 95% level of confidence (±0.55 D).

The mean ortho-antistigmatic refraction was −0.23 D (S.D. = 0.084 D, 95% confidence interval −0.26 to −0.20 D) and individual measures ranged from −0.38 to −0.08 D. The 95% limits of agreement for the residuals of the ortho-antistigmatic data were ±0.17 D and the 95% confidence intervals for these upper and lower limits were +0.13 to +0.21 D and −0.21 to −0.13 D respectively. The 95% reproducibility limit for the ortho-antistigmatic data was 0.24 D. The mean oblique antistigmatic refraction was −0.14 D (S.D. = 0.086 D, 95% confidence interval −0.17 D to −0.11 D). These data extended over the range −0.32 to +0.035 D and the 95% limits of agreement for the residuals of the oblique antistigmatic data were ±0.17 D (95% confidence intervals: +0.12 to +0.22 D and −0.22 to −0.12 D respectively). Despite the fairly substantial variability in the refractive results, all forty prescriptions provided monocular and binocular visual acuities of 6/6 or better as measured at the examination.

*Figure 2* shows a stereo-pair scatterplot of the astigmatic refractive data in symmetric dioptric power space. Each point in the figure represents a single optometrist determined refractive outcome ($n$ = 40) for the subject's right eye. The origin of the figure represents the null power (0.00 D) and data are plotted relative to three orthogonal axes representing the stigmatic (vertical axis **I**), ortho-antistigmatic (horizontal axis **J**) and oblique antistigmatic (horizontal axis **K**) powers. In this figure a point located at 0.50**J** D represents a 1.00 D Interval

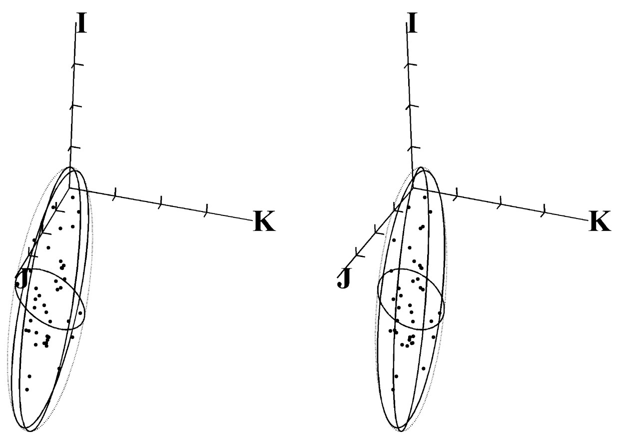


**Figure 2.** Stereo-pair scatterplots in symmetric dioptic power space for 40 measurements of refractive error on a single eye. The origin represents the null power (0.00 D) and the three mutually orthogonal axes labelled **I**, **J** and **K** represent the stigmatic, ortho-antistigmatic and oblique antistigmatic powers respectively. All points not on the vertical axis represent astigmatic powers. Axis length is 1.00 D and the tick intervals are 0.25 D. To view each plot, one should allow one's eyes to drift apart as if looking behind the plane of the page until a three-dimensional percept is obtained. The principal diameter of the ellipsoid is tilted relative to the stigmatic axis, suggesting that the stigmatic and oblique antistigmatic components of refraction are not entirely independent.

**Table 1.** Summary of statistics for the refractive data set of the right eye

Mean dioptric power matrix

$$\mathrm{F} = \begin{pmatrix} -1.0618 & -0.1449 \\ -0.1449 & -0.5945 \end{pmatrix} \mathrm{D}$$

Coefficient of powers: $F_I = -0.8281$ D; $F_J = -0.2339$ D; $F_K = -0.1438$ D

Conventional notation: −0.55/−0.55 × 106°

Norm of mean = 0.8724 D

Variance–covariance matrix ($n = 40$)

$$\mathrm{S} = \begin{pmatrix} 0.0791 & -0.0008 & 0.0092 \\ -0.0008 & 0.0071 & 0.0027 \\ 0.0092 & 0.0027 & 0.0073 \end{pmatrix} \mathrm{D}^2$$

Mean refractive state is shown in dioptric power format, component power format and conventional sphero-cylindrical power format.

of Sturm bounded by vertical and horizontal line foci given by the prescription +0.50/−1.00 × 180°. The boundary of the ellipsoid describes the volume of space within which an estimated 95% of the population of refractive outcomes lie and the centre of the distribution ellipsoid represents the sample mean (−0.55/−0.55 × 106°). *Table 1* details the mean refractive outcome, its norm, and the variance–covariance matrix of the data set for the right eye. The variances of the stigmatic, ortho-antistigmatic and oblique antistigmatic coefficients of power are given by elements $s_{11}$, $s_{22}$ and $s_{33}$ respectively. Elements $s_{12}$, $s_{23}$ and $s_{13}$ represent the stigmatic-ortho-antistigmatic, ortho-oblique antistigmatic and stigmatic-oblique antistigmatic covariances respectively. The stigmatic-oblique antistigmatic covariance is somewhat greater than the other covariances. The effect of this is manifest in the slight tilt observed in the distribution ellipsoid shown in *Figure 2*.

The myopic refractive state of the eye is demonstrated by the fact that all points lie below the origin of the axes in *Figure 2*. The points closest to and furthest from the origin represent the prescriptions 0.00/−0.50 × 90° and −0.75/−1.00 × 110° respectively. The variability of the stigmatic components of power is approximately three times greater than that of the antistigmatic components. Careful inspection of the distribution ellipsoid reveals that its major axis is tilted such that the lower end of the ellipsoid is situated farther from the stigmatic axis than the upper end. This tilt, attributable to the relatively high stigmatic-oblique antistigmatic covariance (see element $s_{13}$ or $s_{31}$ of the variance–covariance matrix given in *Table 1*), demonstrates that the stigmatic and oblique antistigmatic components of power are not strictly independent. There was no statistically significant difference in the refractive data determined by the optometrists in Westminster and Oxford (*Table 2*). All hypothesis tests were performed at the 95% level of significance. It should be noted that the validity of the outcome of the test on the means is conditional on the

**Table 2.** Test statistics and critical values for the hypothesis tests to evaluate differences in the distributions of refractive data determined by the optometrists in Westminster and Oxford

| Condition | Test | Test statistic | Critical value |
|---|---|---|---|
| Oxford/ Westminster | Variance–covariances | $u = 8.13$ | $\chi^2_{0.05,6} = 12.59$ |
| | Means | $w = 0.56$ | $F_{0.05,3,37} = 2.86$ |

equality of the variance–covariances of the respective data sets. For each test the null hypothesis is rejected only if the test statistic exceeds the critical value. In both instances, that is, for both the variance–covariance and mean tests, this was found not to be the case. It follows, therefore, that in both instances there was insufficient evidence to reject the null hypotheses at the 5% level of significance.

The left hand polar profile in *Figure 3* shows the variation of curvital power ($f_{11}$) across different reference meridians from 0° to 180° for each of the refractive outcomes determined by the optometrists. If all the prescriptions were purely stigmatic then each of the profiles would form a perfect semi-circle. The greatest and least variation in meridional power occurred along the 64° and 154° reference meridians respectively and the figure clearly shows a range in power of over a dioptre across all reference meridians. The right hand polar profile in *Figure 3* shows the variation of the orthogonal ($f_{12}$) component of the refractions across all reference meridians.

The inter-ocular difference in the stigmatic components of refraction was calculated by subtracting the stigmatic power of the left eye from that of the right. A histogram of the frequency of these inter-ocular differences is shown in *Figure 4*. The mean inter-ocular

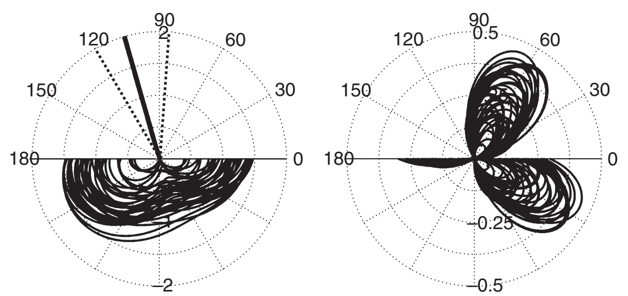


**Figure 3.** Polar profiles of meridional and orthogonal components. The profiles on the left demonstrate the variation of the meridional ($f_{11}$) components of each refraction. Those on the right demonstrate variation of the orthogonal ($f_{12}$) components. The scale for the meridional components is four times that of the orthogonal components. Negative values are represented below the horizontal straight line. Three lines have been added to the left hand figure. The two dashed lines indicate the minimum and maximum cylinder axis orientation (86° and 120° respectively) in the data set. The single solid line indicates the mean cylinder axis orientation (106°). Note that the lines are not intended to convey the magnitude of said cylinders, merely their orientation.

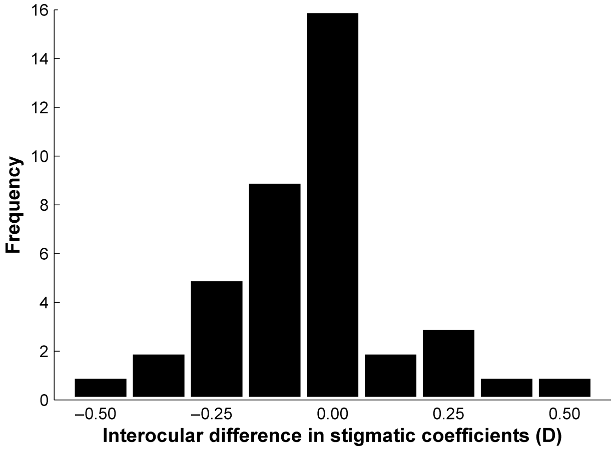


**Figure 4.** A histogram of the differences in the stigmatic components of refraction measured by each optometrist. This inter-ocular difference in stigmatic power was determined by subtracting the stigmatic power of the right eye from that of the left for each measurement.

difference in stigmatic powers was −0.044 D (S.D. = 0.20 D, 95% confidence interval: −0.11 to +0.019 D) but individual estimates of this binocular endpoint of refraction ranged from −0.50 D (the right eye being 0.50 D more myopic than the left) to +0.50 D (the left eye being 0.50 D more myopic than the right). The 95% limits of agreement for the residuals of these findings are ±0.39 D.

## Discussion

The subjective refraction is likely to remain the benchmark against which all refraction devices are measured for quite some time. So entrenched is the notion of the precision of the subjective refraction that it is often considered a measure of *the* refractive state of the eye. And why not? That optometry has flourished is in no small part due to the fact that the subjective refraction 'works'. One may be surprised, however, to learn that very few studies have met the design requirements necessary to investigate the precision of the subjective refraction in a rigorous manner.

Repeatability studies require multiple measures of subjective refraction obtained on a single eye by one examiner. The study that comes closest to satisfying this condition was based on only five independent measures of subjective refraction per eye. Similarly, reproducibility studies require multiple measures of refractive state obtained on a single eye by many independent examiners. Yet all previous studies examining the reproducibility of the subjective refraction were based on data collected by only two or, at most, three examiners. Moreover, the refractionists in a number of these studies were not masked to the outcomes of previous refractions or spectacle prescriptions. This study is novel in that it evaluates the reproducibility of the sphero-cylindrical refraction using data collected from forty qualified optometrists. While this study is unique in that it is designed to estimate the reproducibility limit of the sphero-cylindrical prescription for a particular subject, it is limited by the fact that this subject is not representative of the population in general. While we have no assurance that this subject's visual system is representative even of age- and ametropia-matched individuals, it is interesting to note that the findings reported by Bullimore *et al.* (1998) and Sheedy *et al.* (2004) are consistent with those presented in this paper.

The data presented here suggest that refractions performed by multiple optometrists on a single eye will differ in their stigmatic components by >0.78 D on average not more than once in 20 refractions. Similarly, optometrists will differ in their estimation of the antistigmatic components of refraction by no more than 0.24 D (approximately 0.50 D cylinder) in 95% of repeated measures. The reproducibility of the refraction findings reported here are thus approximately twice as variable as those reported under repeatability conditions (Jennings and Charman, 1973; Kratz and Flom, 1977; Rosenfield and Chiu, 1995). While this finding is consistent with those reported by Bullimore *et al.* (1998) and Sheedy *et al.* (2004), it is somewhat lower than that reported by Zadnik *et al.* (1992) who estimated the 95% limits of agreement for non-cyclopegic subjective refraction to be ±0.63 D (it should be noted, however, that their findings were based exclusively on an analysis of powers measured along the vertical meridian of each eye and are thus not comparable with the results presented here). The close agreement of these findings with those of Bullimore *et al.* (1998) and Sheedy *et al.* (2004) may imply that the results reported here offer a reasonable estimate of the precision of subjective refraction findings on age- and ametropia-matched subjects.

The variability of the stigmatic component of refraction reported here is approximately three times greater than that of the antistigmatic components. The seemingly high variability in the stigmatic components of refraction may be attributed to differences in each optometrist's philosophy toward prescription and the fact that the subject is pre-presbyopic and thus still able to accommodate. Specifically, while some optometrists chose to employ the traditional 'maximum plus to best visual acuity' endpoint, others chose to overcompensate somewhat and prescribe a lens that stimulates accommodation. Many practitioners view this sort of stigmatic overcompensation as acceptable on condition that it does not result in a reduction in visual performance.

Optometrists demonstrated relatively poor agreement in their estimation of the binocular endpoint of refraction. The data reported here suggest that optometrists will, on average, differ in their estimation of the binocular stigmatic endpoint by >0.55 D in 5% of repeated measures. This high variability in results may be attributable to differences in the optometrists' approach to determining the binocular endpoint: a number ($n = 15$) relied exclusively on the outcome of the monocular subjective refraction whereas others made use of the results of one of several 'spherical balancing' techniques ($n = 25$). These spherical balancing techniques, intended to ensure precise equalisation of the accommodative demand on each eye, are considered by many to be of great clinical importance. The balancing techniques used by the optometrists consulted in this study include the Humphriss Immediate Contrast test ($n = 13$), equalisation by alternate occlusion ($n = 2$) and the dissociated bichrome (duochrome) balance ($n = 10$) (Benjamin, 1998). While it is interesting to speculate that the somewhat leptokurtic distribution of data shown in *Figure 4* may represent an inherent difference in the variability of the various balancing techniques (one can imagine a superposition of distributions from each technique, some with high variance and the other with low variance), casual inspection revealed no systematic differences in the outcomes of the various methods of binocular endpoint determination. Unfortunately, the design of the study did not allow for a comparison of these procedures and there is insufficient statistical power to justify a formal *post hoc* analysis of the observed results.

This study is further limited by the fact that the reproducibility of refraction findings is a function of both age and refractive state. One is unlikely to find the sort of stigmatic variation reported here amongst presbyopes who, due to a lack of sufficient accommodative amplitude, will tend to tolerate stigmatic overcompensation poorly. Amongst children the variability in all components of refraction may well be substantially higher than that reported here; due in part to difficulties in examiner-patient communication and, more importantly, the role of an active accommodative system (thus cycloplegic refraction remains the gold standard for measuring refractive status in children). Furthermore, amongst those with very high stigmatic ametropia the role of amblyopia, blur detection threshold and lens vertex distance will all influence the reproducibility of refraction findings.

In spite of these limitations the findings presented in this paper have important implications for the collection and analysis of refractive data. Whereas a single optometrist may be able to perform a refraction with precision of ±0.25 D, refractions performed by different optometrists on age- and ametropia-matched subjects may differ in their stigmatic component by 0.75 D or more. Variability of this magnitude will have a profound impact on the analysis of refractive data collected by multiple optometrists over the course of replication, longitudinal and epidemiological studies. The design of these studies should thus allow for a formal evaluation of the feasibility of treating refraction data from different refractionists as though it were recorded by a single examiner. Ideally, this would involve examining the refractive outcomes of each refractionist under repeatability conditions and formally testing for the equality of variance–covariances and means for each distribution (Harris, 1990, 1991a,b).

## Acknowledgements

Scatterplots were generated using software courtesy of W.F. Harris, D.J. Malan and A. Rubin of the Optometric Science Research Group, Department of Optometry, University of Johannesburg, Johannesburg, South Africa. This work was supported by a research grant from ADLENS Ltd.

## References

Altman, D. G. and Bland, J. M. (1983) Measurement in medicine: the analysis of method comparison studies. *Statistician* **32,** 307–317.

Bannon, R. E. (1977) A new automated subjective optometer. *Am. J. Optom. Physiol. Opt*. **54,** 433–438.

Benjamin, W. J. (1998) Monocular and binocular subjective refraction. In: *Borish*'s Clinical Refraction, 1st edn (eds I. M. Borish and W. J. Benjamin), WB Saunders Company, Philadelphia, PA, pp. 670.

Bland, J. M. and Altman, D. G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **i,** 307–310.

Bullimore, M. A., Fusaro, R. E. and Adams, C. W. (1998) The repeatability of automated and clinical refraction. *Optom. Vis. Sci*. **75,** 617–622.

Chat, S. W. S. and Edwards, M. H. (2001) Clinical evaluation of the Shin-Nippon SQR-5000 autorefractor in children. *Ophthal Physiol. Opt*. **21,** 87–100.

Choong, Y. F., Chen, A. H. and Goh, P. P. (2006) A comparison of autorefraction and subjective refraction with and without cycloplegia in primary school children. *Am. J. Ophthalmol*. **142,** 68–74.

Dave, T. and Fukuma, Y. (2004) Clinical evaluation of the Topcon BV-1000 automated subjective refraction system. *Optom. Vis. Sci*. **81,** 323–333.

Elliott, M., Simpson, T., Richter, D. and Fonn, D. (1997) Repeatability and accuracy of automated refraction: a comparison of the Nikon NRK-8000, the Nidek AR-1000, and subjective refraction. *Optom. Vis. Sci*. **74,** 434–438.

Farook, M., Venkatramani, J., Gazzard, G., Cheng, A., Tan, D. and Saw, S. (2005) Comparisons of the handheld autorefractor, table-mounted autorefractor, and subjective refraction in Singapore adults. *Optom. Vis. Sci*. **82,** 1066–1070.

Fick, H. H. (1972) Fortschrittliche Rechnungsarten in der Augenoptik. Folge 11. *Der Augenoptiker*. **12,** 60–63.

Fick, H. H. (1973) Fortschrittliche Rechnungsarten in der Augenoptik. Folge 12. *Der Augenoptiker*. **2,** 45–49.

Gillan, W. D. H. (2006) Multiple measures of refractive state of a test eye making use of a Nidek ARK-10000 OPD-scan. *S. Afr. Optom* **65,** 144–149.

Grant, C., Nicholas, R., Moore, L. and Salisbury, C. (2002) An observational study comparing quality of care in walk-in centres with general practice and NHS Direct using standardised patients. *BMJ* **324,** 1556–1559.

Harris, W. F. (1990) Statistical inference on mean dioptric power. Hypothesis testing and confidence regions. *Ophthalmic Physiol. Opt*. **10,** 363–372.

Harris, W. F. (1991a) Representation of dioptric power in Euclidean 3-space. *Ophthalmic Physiol. Opt*. **11,** 121–129.

Harris, W. F. (1991b) Statistical inference on mean dioptric power, asymmetric powers and singular covariance. *Ophthalmic Physiol. Opt*. **11,** 263–269.

Harris, W. F. (1999) A unified paraxial approach to astigmatic optics. *Optom. Vis. Sci.* **76,** 480–499.

Harris, W. F. (2001) Analysis of astigmatism in anterior segment surgery. *J. Cataract Refract. Surg*. **27,** 107–128.

Harris, W. F. (2005) Reduction of artefact in scatter plots of spherocylindrical data. *Ophthalmic Physiol. Opt*. **25,** 13–17.

Harvey, E. M., Miller, J. M., Wagner, L. K. and Dobson, V. (1997) Reproducibility and accuracy of measurements with a hand held autorefractor in children. *Br. J. Ophthalmol.* **81,** 941–948.

International Organisation for Standardisation (ISO) (1994) *ISO 5725-1 Accuracy (Trueness and Precision) of Measurement Methods and Results. Part 1: General Principles and Definitions*. BSI, London, pp. 1–10.

Jennings, J. A. M. and Charman, W. N. (1973) A comparison of errors in some methods of subjective refraction. *Ophthal. Opt* **13,** 8–18.

Johnson, B. L., Edwards, J. S., Goss, D. A., Penisten, D. K. and Fulk, G. W. (1996) A comparison of three subjective tests for astigmatism and their interexaminer reliabilities. *J. Am. Optom. Assoc*. **67,** 590–598.

Keating, M. P. (1982) Advantages of a block matrix formulation for an astigmatic system. *Am. J. Optom. Physiol. Opt* **59,** 851–857.

Keating, M. P. (1986) Dioptric power in an off-axis meridian: the torsional component. *Am. J. Optom. Physiol. Opt* **63,** 830–838.

Koch, D. D. (2001) How should we analyze astigmatic data? *J. Cataract Refract. Surg*. **27,** 1–3 (editorial).

Kratz, L. D. and Flom, M. C. (1977) The Humphrey Vision Analyzer: reliability and validity of refractive error measures. *Am. J. Optom. Physiol. Opt* **54,** 653–659.

Leinonen, J., Laakkonen, E. and Laatikainen, L. (2006) Repeatability (test–retest variability) of refractive error measurement in clinical settings. *Acta Ophthalmol. Scand.* **84,** 532–536.

Long, W. F. (1976) A matrix formalism for decentration problems. *Am. J. Optom. Physiol. Opt*. **53,** 27–33.

MacKenzie, G. E., Badenhorst, A., Brand, C., van Niekerk, F., Burger, S. and Hall, G. (2001) Variance in daily autokeratometric means across the menstrual cycle. *S. Afr. Optom* **60,** 89–96.

Mallen, E. A., Wolffsohn, J. S., Gilmartin, B. and Tsujimura, S. (2001) Clinical evaluation of the Shin-Nippon SRW-5000 autorefractor in adults. *Ophthalmic Physiol. Opt*. **21,** 101–107.

McKendrick, A. M. and Brennan, N. A. (1995) Clinical evaluation of refractive techniques. *J. Am. Optom. Assoc*. **66,** 758–765.

Naeser, K. and Hjortdal, J. (2001) Polar value analysis of refractive data. *J. Cataract Refract. Surg*. **27,** 86–94.

Office of Public Sector Information (1989). *The Opticians Act 1989*. Published by HMSO, London. Available at: http://www.opsi.gov.uk/acts/acts1989/Ukpga_19890044_en_1.htm (accessed 15 August 2007).

Perrigin, J., Perrigin, D. and Grosvenor, T. (1982) A comparison of clinical refractive data obtained by three examiners. *Am. J. Optom. Physiol. Opt*. **59,** 515–519.

Pesudovs, K. and Weisinger, H. S. (2004) A comparison of autorefractor performance. *Optom. Vis. Sci*. **81,** 554–558.

Raasch, T. W., Schechtman, K. B., Davis, L. J., Zadnik, K. and the CLEK Study Group. (2001) Repeatability of subjective refraction in myopic and keratoconic subjects: results of vector analysis. *Ophthalmic Physiol. Opt*. **21,** 376–383.

Rosenfield, M. and Chiu, N. N. (1995) Repeatability of subjective and objective refraction. *Optom. Vis. Sci*. **72,** 577–579.

Rubin, A.. (1993) *Short-Term Variation of Refractive Behavior in Human Eyes*. Masters Thesis, Rand Afrikaans University, Johannesburg, South Africa.

Salchow, D. J., Zirm, M. E., Steildorf, C. and Parisi, A. (1999) Comparison of objective and subjective refraction before and after laser *in situ* keratomileusis. *J. Cataract Refract. Surg*. **25,** 827–835.

Sheedy, J., Schanz, P. and Bullimore, M. (2004) Evaluation of an automated subjective refractor. *Optom. Vis. Sci*. **81,** 334–340.

Sloan, A. E., Dunphy, E. B., Emmons, W. V. and Gallagher, J. R. (1954) A comparison of refraction results on the same individuals. *Am. J. Ophthalmol*. **37,** 696–699.

Thibos, L. N. and Horner, D. (2001) Power vector analysis of the optical outcome of refractive surgery. *J. Cataract Refract. Surg*. **27,** 80–85.

Walline, J. J., Kinney, K. A., Zadnik, K. and Mutti, D. O. (1999) Repeatability and validity of astigmatism measurements. *J. Refract. Surg*. **15,** 23–31.

Zadnik, K., Mutti, D. O. and Adams, A. J. (1992) The repeatability of measurement of the ocular components. *Invest. Ophthalmol. Vis. Sci*. **33,** 2325–2333.