Repeatability of In Vivo Corneal Confocal Microscopy to Quantify Corneal Nerve Morphology

Ioannis N. Petropoulos, MSc,* † Tauseef Manzoor, MBChB,* Philip Morgan, PhD,‡ Hassan Fadavi, MD,* † Omar Asghar, MBChB,* † Uzaman Alam, MBChB, MPH,* † Georgios Ponirakis, MPhil,* † Mohammad A. Dabbah, PhD,§ Xin Chen, PhD,§ James Graham, PhD,§ Mitra Tavakoli, PhD,* † and Rayaz A. Malik, MBChB, FRCP, PhD* ‡

**Purpose:** To establish intraobserver and interobserver repeatability, agreement, and symmetry of corneal nerve fiber (NF) morphology in healthy subjects using in vivo corneal confocal microscopy.

**Methods:** Nineteen subjects underwent in vivo corneal confocal microscopy (Heidelberg Retinal Tomograph III Rostock Cornea Module) at baseline and 7 days apart. Bland–Altman plots were generated to assess agreement, and the intraclass correlation coefficient and coefficient of repeatability were calculated to estimate intraobserver and interobserver repeatability for corneal NF density (numbers per square millimeter), nerve branch density (NBD; numbers per square millimeter), NF length (millimeters per square millimeter), and NF tortuosity coefficient. Symmetry between the right and left eyes was also assessed.

**Results:** Intraclass correlation coefficient and coefficient of repeatability for intraobserver repeatability were 0.66 to 0.74 and 0.17 to 0.64, for interobserver repeatability 0.54 to 0.93 and 0.15 to 0.85, and for symmetry 0.34 to 0.77 and 0.17 to 0.63, respectively. NBD demonstrated low repeatability.

**Conclusions:** This study demonstrates good repeatability for the manual assessment of all major corneal NF parameters with the exception of NBD, which highlights the difficulty in defining nerve branches and suggests the need for experienced observers or automated image analysis to ensure optimal repeatability.

**Key Words:** In vivo corneal confocal microscopy, diabetic neuropathy, repeatability

*Cornea 2012;0:1–7*

---

**D**etailed histological analysis of the human cornea before in vivo corneal confocal microscopy (IVCCM) was only possible postmortem using light and electron microscopy. Since the 1980s, IVCCM has been used in ophthalmic research and in clinical practice to assess corneal dystrophies and ectasias; Acanthamoeba, fungal, bacterial, and viral keratitis; the effects of contact lens wear; dry eye disease; and postsurgical follow-up.

Real-time IVCCM has also enabled the characterization of corneal nerves in healthy and keratoconic eyes. Recently, in vitro studies using state-of-the-art immunohistochemical techniques have comprehensively investigated the architecture of the corneal nerves and described novel features. We and others have recently applied this technique to quantify corneal subbasal nerve fibers (NFs) in a variety of peripheral neuropathies including diabetic neuropathy, idiopathic small fiber neuropathy, Fabry disease, anti–myelin-associated glycoprotein neuropathy, chemotherapy-associated peripheral sensory neuropathy, non–length-dependent small fiber neuropathy, and type IV/V hereditary sensory and autonomic neuropathy. Quantification of corneal nerve morphology using 4 key parameters, namely, NF density (NFD), nerve branch density (NBD), NF length (NFL), and the tortuosity coefficient (TC), has allowed the early detection and stratification of peripheral neuropathy and also the assessment of repair after simultaneous pancreas and kidney transplantation. Tavakoli et al recently reported high sensitivity (0.82) and moderate specificity (0.52) for the detection of diabetic neuropathy using IVCCM. However, there is considerable variability for the different corneal nerve parameters assessed because of the subjective criteria applied to identify each structure. Possible solutions include the adoption of internationally accepted criteria and rules to identify the different corneal nerve structures or the development of fully automated image analysis software.

Two recent studies have demonstrated high repeatability of IVCCM but focused primarily on NFL. Hence, the aim of this study was to establish intraobserver, interobserver, and between-eye repeatability and agreement in control subjects for each of the 4 key parameters used to quantify neuropathy.
METHODS

Study Subjects

Nineteen randomly selected healthy subjects aged 23.1 ± 1.2 years, without peripheral neuropathy and/or diabetes, were studied. The study was approved by the North Manchester Research Ethics Committee, and informed written consent was obtained from each subject. None of the subjects had a history of ocular surgery, contact lens wear, corneal infection, or any other systemic disease known to affect the peripheral nervous system. Both eyes of each subject were examined by slit-lamp biomicroscopy and confirmed to be clinically normal. None of the subjects was obese or had abnormal glucose or lipid levels. We used the Toronto consensus criteria36 to exclude peripheral neuropathy by assessing the neuropathy symptom profile, neuropathy deficit score, quantitative sensory testing for vibration perception threshold, cold and warm thresholds, and cold-induced and heat-induced pain.

Corneal Confocal Microscopy

All subjects were scanned with a laser IVCCM [Heidelberg Retinal Tomograph III Rostock Cornea Module (HRT III RCM); Heidelberg Engineering GmbH, Heidelberg, Germany] on 2 occasions separated by a 1-week interval. This IVCCM uses a 670-nm wavelength helium–neon diode laser, which is a class I laser and therefore does not pose any ocular safety hazard. A ×63 objective lens with a numerical aperture of 0.9 and a working distance, relative to the applanating cap (TomoCap; Heidelberg Engineering GmbH), of 0.0 to 3.0 mm was used. The size of each 2-dimensional image produced was 384 × 384 μm, which has a 15 × 15-degree field of view and 10 μm per pixel transverse optical resolution. HRT III RCM uses an entirely digital image capture system, and all images are stored in an external hard drive.

A drop of 0.4% benoxinate hydrochloride (Chauvin Pharmaceuticals, Chefaro, United Kingdom) was used to anesthetize each eye, and Viscotears (0.2% carbomer 980; Novartis UK) was used as the coupling agent between the cornea and the applanating cap. All subjects were asked to fixate on an outer fixation light throughout the IVCCM scan and a CCD [charge-coupled device] camera was used to image the cornea and correctly position the applanating cap to enable image capture strictly from the corneal apex. The overall examination took ~4 minutes for both eyes of each subject at each visit. All images were captured using the “section” mode in the Heidelberg Eye Explorer of the HRT III RCM. The other 2 available modes are “volume” and “sequence.” As Hertz et al35 note, the volume mode may have advantage when inexperienced examiners are using the technique. For the purposes of this study, the same experienced examiner performed all IVCCM scans. There is no general consensus on optimal IVCCM image sampling. We captured 10 [5 left eye (LE), 5 right eye (RE)] images of high clarity at 1-μm intervals from the central cornea of each subject.

Image Analysis

Two observers masked from each other analyzed 380 IVCCM nonoverlapping images, which were randomized before analysis, to assess interobserver repeatability. Observer 1 was experienced in the task of IVCCM image analysis (>2400 images), and observer 2 had no previous experience of corneal nerve quantification. Observer 1 quantified the relevant IVCCM images to assess intraobserver repeatability and symmetry masked for the visit and eye examined. Criteria for image selection were depth, focus position, and contrast. The images were manually analyzed using proprietary purpose-written software (CCMetrics31; M. A. Dabbah, imaging science and biomedical engineering, University of Manchester, Manchester, United Kingdom). The specific parameters measured per frame were those we have previously established18: NFD (numbers per square millimeter), NBD (numbers per square millimeter), NFL (millimeters per square millimeter), and the TC (Fig. 1). NFD is defined as the total number of main NFs per frame divided by the area of the frame in square millimeters (area = 0.16033585 mm²; Fig. 1). NBD is defined as the total number of main nerve branches (NBs; strictly branches that

FIGURE 1. A, An original image as captured with the HRT III RCM. B, An analyzed image using CCMetrics.31 NFD is measured under the red color, which highlights the NFs, and an integrated algorithm measures the value. NBD is measured with the green dots that highlight the junction between NFs and NBs. NFL is the summation of the length of all the nerves highlighted under the blue and red colors. TC—a measure of NF tortuosity—is measured simultaneously with NFD on each NF and is highlighted with the red color. The method is identical to that previously described by Kallinkos et al29 and has been integrated into the current algorithm.
stem from an NF) divided by the area of the frame. NFL is the total length of NFs, NBs, and secondary NBs (branches that stem from an NB) per frame. TC is a mathematical computation of the NF tortuosity as previously described by Kallinikos et al., which is independent of the angle of the nerve in the image. A straight nerve equals a TC of zero and the TC increases with increasing tortuosity of the NF.

**Statistical Analysis**

Data analysis was performed using Microsoft Office Excel 2008 (Microsoft, WA) and StatsDirect version 2.7.7 (StatsDirect Ltd, Cheshire, United Kingdom), and the data are presented as mean ± SD. The data were tested for normality before analysis and appropriate statistical techniques were employed. Differences between groups of measurements were assessed by means of a paired t test. Power analysis was used to calculate the minimum sample size needed to detect an effect. The results showed that for 80%, 85%, and 90% power, 17, 19, and 21 subjects were required, respectively. For the purposes of the present analysis, a 95% confidence interval (CI) was used and a \( P < 0.05 \) considered significant.

The intraclass correlation coefficient (ICC) was calculated to estimate the repeatability of the measurements between and within “occasions” and “observers.” The ICC can be used as an index of the correlation between repeated measures, that is, as an index of repeatability.\(^3\) The ICC was considered excellent if 0.8 to 1 and very good if 0.60 to 0.79. Coefficient of repeatability (CoR) was also calculated as a percentage of an average measurement to estimate the repeatability of the sample. A CoR between 0 and 0.2 was considered good, 0.2 to 0.5 acceptable, and >0.5 poor. The means of the measurements were plotted against the differences between the measurements and the upper and lower limits of agreement were calculated (limits of agreement: 1.96 ± SD, 1.96 – SD), as described by Bland and Altman,\(^3\) to appreciate the between-observer, within-subject, and between-occasion agreement.

**RESULTS**

Subjects in this study had a body mass index of 24.8 ± 4.1 kg/m\(^2\), hemoglobin A1c level of 5.5% ± 0.2%, low-density lipoprotein cholesterol level of 2.7 ± 0.8 mmol/mol, high-density lipoprotein cholesterol level of 1.5 ± 0.3 mmol/mol, and serum triglyceride level of 1.3 ± 0.6 mmol/mol. Subjects had no evidence of peripheral neuropathy: neuropathy deficit score, 0; neuropathy symptom profile, 0; vibration perception threshold, 3.3 ± 1.3 Hz; cold threshold/warm threshold, 28.6 ± 2.4 ± 36 ± 1.8°C; and cold-induced pain/heat-induced pain, 6.4 ± 5.9/47.1 ± 3.9°C.

Intraobserver repeatability was assessed for each parameter using images from the same location and depth of the same eye on 2 separate occasions 7 days apart by the same observer (REv1 vs. REv2 for visit 1 and visit 2, respectively; Table 1). There were no significant differences (\( P > 0.05, 95\% \) CI) between the results from the first scan and the repeated scan. The mean of the values was plotted against the difference between them to derive the Bland–Altman plots (Fig. 2). The relevant ICC values were as follows: NFD, 0.74 (Fig. 2A); NBD, 0.61 (Fig. 2B); NFL, 0.70 (Fig. 2C); and TC, 0.66 (Fig. 2D). The respective CoR values were as follows: NFD, 0.17; NBD, 0.64; NFL, 0.19; and TC, 0.46. The mean differences (±SD) between the 2 assessments were as follows: NFD, 0.1 ± 3.6 numbers per square millimeter; NBD, 5.0 ± 19.4 numbers per square millimeter; NFL, 1.5 ± 2.8 mm/mm\(^2\); and TC, 0.4 ± 3.6.

Interobserver repeatability refers to the assessment of corneal nerve parameters by 2 observers on images of the same eye from the same visit (Table 1). Among the 4 parameters, only NBD showed a significant difference between observers (\( P < 0.0001, 95\% \) CI). The ICC values were as follows: NFD, 0.82 (Fig. 3A); NBD, 0.54 (Fig. 3B); NFL, 0.66 (Fig. 3C); and TC, 0.93 (Fig. 3D). The respective CoR values were as follows: NFD, 0.15; NBD, 0.85; NFL, 0.17; and TC, 0.18. The mean differences (±SD) between the 2 observers were as follows: NFD, 1.1 ± 3.1 numbers per square millimeter; NBD, 56.0 ± 39.0 numbers per square millimeter; NFL, 2.7 ± 2.6 mm/mm\(^2\); and TC, 0.7 ± 1.4.

The symmetry of central corneal nerve morphology was assessed in images from the RE and LE, of the same individual, on the same occasion, and quantified by the same examiner (Table 1). There were no significant differences (\( P > 0.05, 95\% \) CI) in corneal nerve morphology between the RE and LE.

**TABLE 1. NFD, NBD, NFL, and TC to Assess Intraobserver and Interobserver Repeatability, Agreement, and Symmetry Between the RE and the LE**

<table>
<thead>
<tr>
<th></th>
<th>NFD (No./mm(^2))</th>
<th>NBD (No./mm(^2))</th>
<th>NFL (mm/mm(^2))</th>
<th>TC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intraobserver</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RE visit 1</td>
<td>38.3 ± 3.9</td>
<td>58.1 ± 23.0</td>
<td>27.6 ± 4.0</td>
<td>15.8 ± 4.0</td>
</tr>
<tr>
<td>RE visit 2</td>
<td>38.2 ± 5.0</td>
<td>63.1 ± 21.7</td>
<td>29.1 ± 3.8</td>
<td>15.5 ± 4.6</td>
</tr>
<tr>
<td><strong>Interobserver</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observer 1</td>
<td>38.2 ± 5.0</td>
<td>63.1 ± 21.7</td>
<td>29.1 ± 3.8</td>
<td>15.5 ± 4.6</td>
</tr>
<tr>
<td>Observer 2</td>
<td>38.5 ± 5.4</td>
<td>120.0 ± 51.2</td>
<td>31.7 ± 4.8</td>
<td>14.7 ± 3.8</td>
</tr>
<tr>
<td><strong>Symmetry</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RE visit 1</td>
<td>38.3 ± 3.9</td>
<td>58.1 ± 23.0</td>
<td>27.6 ± 4.0</td>
<td>15.8 ± 4.0</td>
</tr>
<tr>
<td>LE visit 1</td>
<td>37.8 ± 4.5</td>
<td>56.3 ± 26.4</td>
<td>27.8 ± 5.2</td>
<td>15.5 ± 1.8</td>
</tr>
</tbody>
</table>

Results are expressed as mean ± SD.

© 2012 Lippincott Williams & Wilkins www.cornejnl.com | 3

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.
the LE. The calculated ICC values were as follows: NFD, 0.77 (Fig. 4A); NBD, 0.73 (Fig. 4B); NFL, 0.45 (Fig. 4C); and TC, 0.34 (Fig. 4D). The respective CoR values were as follows: NFD, 0.17; NBD, 0.63; NFL, 0.36; and TC, 0.48. The mean differences (±SD) between the RE and the LE were as follows: NFD, 0.07 ± 3.9 numbers per square millimeter; NBD, 1.28 ± 18.1 numbers per square millimeter; NFL, 0.1 ± 5.0 mm/mm²; and TC, 0.3 ± 3.7.

**DISCUSSION**

The quantification of corneal subbasal nerves is a rapidly evolving area of special interest to both clinicians and scientists as a surrogate for diagnosing, assessing progression,18 and the benefits of therapeutic intervention in a range of peripheral neuropathies.30 Initial studies provided qualitative evidence of corneal NF alterations or reported changes in the architecture after surgery.11 In the context of using corneal nerve morphology as a surrogate for peripheral neuropathy, a more clearly defined approach has been developed to quantify the 4 key parameters: NFD, NBD, NFL, and TC.18,28,29

Whether individual anatomical variations and intraobserver and interobserver consistency influence the results remains unclear. A recent study has shown that NFL has a very high between-observer and between-occasion repeatability in patients with type 2 diabetes,44 whereas another study showed that NFL had the best reproducibility and validity among all parameters in controls and patients with type 1 diabetes and suggested that the development of IVCCM should focus on the measurement of NFL because of its superiority over the other parameters.35 However, quantifying NFL alone limits the interpretation of corneal nerve damage and repair in the context of disease and particularly when assessing repair after treatment of peripheral neuropathy. Therefore, we have undertaken a detailed assessment of the repeatability and agreement of the 4 main parameters we originally developed and applied18,19,29 in a range of peripheral neuropathies.22,23,28

In this study, corneal NF morphology showed consistency between the RE and the LE. Although NFD and NFL achieved the highest values for intraobserver and interobserver repeatability and agreement, NBD and TC showed
less consistency. Across all assessments, NBD appeared to be the least repeatable parameter, and this finding highlights the importance of accurately defining NBs and NFs. The correct identification of NBs in IVCCM images is especially difficult and mainly depends on background contrast, image clarity, and observer experience and interpretation. In addition, Patel and McGhee40 showed for the first time a continuous centripetal movement of identifiable branch points in the human corneal subbasal nerve plexus of up to 26 μm per week over a 6-week period, which may also cause variability.

A common finding in IVCCM images is crossing X-shaped NFs running from the top to the bottom of the image or Y-shaped appearance of an NF and an NB. In the former case, interpretation is easy and is not expected to vary between observers. However, in the latter case there is no standardized rule to-date to assist the analyzer to correctly define the NF and the NB. Selecting either side to be the NF can affect the outcome because the TC between nerves of the same individual varies. Individual criteria may include the thickness, the continuous pattern, or the reflectivity of the main axon, which differs from that of the NB. In more complicated cases where the pattern is best described by a tree shape (>1 branch stemming from an NF) or an X shape with multiple branches, the variation will clearly increase and this may significantly affect NBD. Hence, both NBD and TC have inherent liability for variability in repeated assessment, as this task is highly subjective, especially when different observers undertake the analysis.

Among the 2 most repeatable parameters, NFD was superior to NFL in all measurements. This finding contrasts that of Hertz et al35 who found NFL to be the most reliable of all IVCCM nerve parameters. NFL is defined as the sum of the total length of NFs and NBs per frame, that is, all nerve structures and may therefore be ideally used as a pan-corneal marker of peripheral neuropathy. However, high or low NFL
does not capture concomitant degeneration and regeneration and may not be as sensitive as NBD, hence limiting the interpretation of subbasal corneal nerve repair. Differences in image collection and sampling techniques may also affect the outcome.

The primary purpose of this study was to quantitatively evaluate the repeatability of assessing subbasal corneal nerves using IVCCM. Possible limitations of this study are the small sample size and the small area of the cornea chosen for analysis. Therefore, the assessment of IVCCM repeatability in multiple corneal areas should also be established. We have demonstrated good intraobserver and interobserver repeatability and consistency between the RE and the LE for NFD and NFL but have identified lower repeatability for NBD and TC when deploying manual image analysis of corneal NF morphology. Both the latter parameters are however important to quantify corneal innervation because they add considerably to the interpretation of disease effect for both nerve degeneration and regeneration. The variability observed with the technique may be improved by applying predefined identification rules for the NFs and their branches. A possible solution for both these issues may lie in the development of a fully automated image analysis system, which would eliminate inconsistencies, enhance repeatability, markedly reduce the analysis time, and hence make IVCCM suitable for clinical practice.

**ACKNOWLEDGMENTS**

The authors thank the Manchester Biomedical Research Centre and Wellcome Trust Clinical Research Facility for supporting this research.

**REFERENCES**
