Inter-Method and Inter-Observer Agreement of Retinal Oximetry Image Processing

Pooja D. Jani, Jean-Claude Mwanza, Alexa M. Waters, Kyle B. Billow, and Seema Garg*

Department of Ophthalmology, University of North Carolina, USA

Abstract

Purpose: To evaluate the agreement between two independent observers and two retinal oximetry image processing methods.

Methods: Optic disc-centered retinal oximetry images acquired using the Oxymap T1 retinal oximeter in 59 normal eyes were processed by two independent observers using two methods (Full-Field and Ring methods). Agreement between processing methods and observers was assessed using Bland-Altman analysis. Parameters analyzed included arterial (Sao2) and venous (Svo2) oxygen saturation, and arterial and venous width.

Results: Significant differences were observed between processing methods, but not between observers in Svo2, venous width, and arterial width (all P < 0.001). Agreement analysis showed that mean differences between global Svo2 measurements obtained by the two observers were about zero using either method (Full-Field method range: -0.047% to -0.032%; Ring method range: -0.071% to -0.062%), with narrow 95% confidence limits. Mean differences between Svo2 measurements generated using the two methods by either observer were also small (range: -1.70% to -0.049% for observer #1 and -1.76% to -0.049% for observer #2), but the 95% confidence limits of agreement were slightly wider.

Conclusions: The two methods, but not observers, showed significant differences in some parameters, indicating different observers may reliably obtain retinal oxygenation parameters using either processing method. Despite reasonably good agreement between the two methods, the wide range of differences between them may limit their interchangeability. The Ring method is faster and easier to use, and may be better suited for clinical applications in the future.

INTRODUCTION

Retinal tissue hypoperfusion and hypoxia have been implicated in the pathogenesis of vision-threatening disorders such as diabetic retinopathy, [1-4] age-related macular degeneration (AMD), [5,6] glaucoma, [7,8] and vaso-occlusive retinal disease [9-11]. The ability to detect subtle changes in retinal oxygenation and function might be important for early detection, management, and monitoring of these diseases. A complete understanding of the role of oxygen in human retinal pathology has been limited by the lack of non-invasive and sensitive methods for measuring intraretinal oxygen levels.

Retinal oximetry is an emerging non-invasive retinal imaging technique for determining retinal vessel oxygen saturation (SO2) and may ultimately become a diagnostic tool for tracking disease progression and evaluating the effectiveness of treatment. It is based on the same principles as standard pulse oximetry, utilizing the differential light absorbance of oxyhemoglobin and deoxyhemoglobin, in this case at 570 nm and 600 nm light wavelengths, respectively. The technical aspects of retinal oximetry have been previously described in the literature [12,13]. The Oxymap T1 retinal oximeter (Oxymap ehf, Reykjavik, Iceland) holds potential for future clinical use because it has been shown to be sensitive to variations in hemoglobin oxygenation [14].

However, despite available reports on good reproducibility of SO2 measurements in normal and diseased eyes, [14-16] no systematic study has been performed to evaluate inter-method and inter-observer agreement of retinal oximetry image processing methods. The purpose of this study was to determine whether retinal SO2 measurements obtained using two images processing methods by the same observer as well as using the same processing method by two observers are comparable.
METHODS

Subjects

A total of 59 healthy subjects were enrolled in this study. Approval was obtained from the Institutional Review Board of the University of North Carolina (UNC) School of Medicine and the study was performed in accordance with the Health Insurance Portability and Accountability Act. Subjects were recruited from general ophthalmology clinics at UNC. Enrolled subjects were 18 years of age or older with a documented normal ophthalmic examination. Subjects from diverse ethnic backgrounds were enrolled, with the goal of representing African-Americans, Asians, Caucasians, and Hispanics equally. Exclusion criteria included a history of ocular conditions such as cataracts, glaucoma, diabetic retinopathy, macular degeneration, retinal vascular occlusions, or any other ocular disease that could confound retinal SO₂ measurements. Additional exclusion criteria included a history of diabetes, chronic obstructive pulmonary disease, severe anemia, hemoglobinopathies such as sickle cell anemia, uncontrolled systemic hypertension, or intraocular radiation therapy.

Image acquisition

After the subject’s pupils were dilated with tropicamide 1% and phenylephrine 2.5%, 50- degree dual-wavelength (570 nm and 600 nm) retinal oximetry images were obtained under uniform lighting conditions with the Oxymap T1 Retinal Oximeter, which is mounted on a Topcon TRC50-VT fundus camera (Topcon Co, Tokyo, Japan). All images were obtained in a closed photography room where overhead lights and other instrument lights were turned off so that the sole light source was from a computer monitor associated with the retinal oximeter equipment. One optic disc-centered retinal oximetry image was acquired per eye. Image quality was judged during acquisition; images with poor focus, bright reflections or shadows, poor contrast, or poor positioning were excluded. Standard optic disc-centered 50-degree color fundus photographs (FF 450Plus IR, Carl Zeiss Meditec, Dublin, CA) were then obtained. These photos were used to document the absence of pathology as well as to help differentiate arterioles from venules in cases where the distinction was difficult on retinal oximetry images alone.

Retinal oximetry image processing and measurement

For the purposes of this study, only right eye optic disc-centered images were processed due to the high correlation between both eyes in a single patient. Two processing methods, referred to as the Full-Field method and the Ring method were refined by our research group from a previously described method [1]. Each image was processed by two trained observers (AMW and KBB) using Oxymap Analyzer software version 2.3.1. The software was set to detect a minimum vessel width of 8.0 pixels. Step-by-step descriptions of the two image processing methods are outlined below.

Full-field method

Step 1 – Optic Disc Exclusion: A measurement circle 300 pixels in diameter is centered over the optic disc. This circle is expanded in 50-pixel increments until all vessel segments in proximity to the optic disc margin are included within the circle. This process defines an area around the optic disc where vessel identification as arteriole or venule is ambiguous. The circle is also expanded to include vessel branchings and vessel crossings in proximity to the optic disc margin (Figure 1A). All vessels within the optic disc-centered circle are then excluded.

Step 2 – Quadrant Division: Perpendicular lines are drawn through the center of the optic disc-centered exclusion circle and kept parallel with the edges of the image. The optic disc-centered image is thus divided into four quadrants: superotemporal, supersonal, inferonasal, and inferotemporal (Figure 1A).

Step 3 – Exclusion Mapping: A margin with a width of 31 pixels is excluded from the peripheral edges of the image. All branchings and vessel crossings are then excluded using a 19 pixel circle (Figure 1B).

Step 4 – Vessel Selection: Vessels to be included in the analysis are then manually selected by either single-clicking the segment of interest or clicking and dragging over the desired vessel length. A vessel that straddles two quadrants is assigned to the quadrant in which the majority of the vessel lies and that the vessel apparently supplies or drains. All available arterioles in the superotemporal quadrant are first selected and added to the analysis table (Figure 1C) to generate SO₂ and arteriole width measurements for this quadrant. This procedure is repeated for all arterioles in the superonasal, inferonasal, and inferotemporal quadrants. All quadrants are then simultaneously selected in the analysis table to obtain the global arteriole SO₂ and arteriole width measurements of the image (Figure 1D). The entire process is repeated for all venules in each of the four quadrants.

Ring method

Step 1 – Optic Disc Exclusion: A circle measuring 250 pixels
in diameter is centered over the optic disc. The circle is then expanded in 50 pixel increments in order to best approximate the optic disc margin in a manner similar to Step 1 of the Full-Field method (Figure 2A).

Step 2 – Quadrant Division: The quadrants are divided as described in Step 2 of the Full-Field method (Figure 2A).

Step 3 – Measurement Area Demarcation: A concentric inner circle two times the diameter of the optic disc-centered circle is drawn around the central circle. Another concentric outer circle 4 times the diameter of the optic disc-centered circle is then drawn. The region between the inner and outer concentric circles defines the measurement area for analysis. All vessels within the central and inner circles are excluded (Figure 2A and 2B).

Step 4 – Exclusion Mapping: Exclusion mapping is carried out as described in Step 3 of the Full-Field method. The area of the image beyond the outer circle is not included in processing (Figure 2C).

Step 5 – Vessel Selection: Vessel selection is carried out as specified in Step 4 of the Full-Field method. Only vessel segments within the ring measurement area are selected for processing (Figure 2C). Global measurements are generated separately for all arterioles and venules within the ring measurement area (Figure 2D).

It is important to note that the Ring method includes additional exclusion strategies that are used to increase accuracy. These strategies are applied during Step 4 and are as follows: 1) branch points and crossings that should be excluded are first identified on the black-and-white fundus image by removing the SO2 overlay map (Figure 3A). After reapplying the SO2 overlay, exclusion points are retained if they mark the intersection of two or more detected (colored or gray) vessels. Exclusion points that do not mark the intersection of detected vessels are removed (Figure 3B). 2) Regions of the measurement area where vessel detection is inappropriate (i.e., the SO2 overlay does not reflect vessel anatomy, or vessel anatomy prohibits accurate analysis) are omitted. Intertwining, crossing, or overlapping vessel segments are excluded using adjacent 19 pixel circles placed in a proximal-to-distal direction along the vessel segment (Figure 3C1). 3) If the software detects a bifurcation within a single linear vessel, the entire length of the apparently bifurcated vessel is excluded (Figure 3C2). 4) If two exclusion areas are in close proximity (<19 pixels), the vessel segment between them is excluded.

Statistical analysis

Statistical analyses were performed using SPSS software (version 20.0; SPSS, Inc., Chicago, IL). Student t-test for paired samples was used to test the significance level of differences between measurements obtained by the two observers using each analysis method and between measurements generated with the two methods by each observer. The agreement between observers and between methods was assessed in the 59 normal eyes by calculating the bias (mean difference between observers or methods) and the lower (mean difference – 2 x standard deviation of the differences) and upper (mean difference + 2 x standard deviation of the differences) 95% limits of agreement as suggested by Bland and Altman [17]. The agreement was assessed based on the assumption of equal imprecision between operators and between methods. All differences between observers and between methods were calculated as measurements from observer #1 minus those from observer #2 using the same method or measurements generated from the Full-Field method.
minus those from the Ring method using the same observer, respectively. The parameters of SaO\textsubscript{2} and SvO\textsubscript{2} and arterial and venous widths were analyzed for this study. The mean differences between observers and methods were compared using the paired Student t-test. A minimum of 21 subjects were required to detect a significant difference of 3% in SaO\textsubscript{2} between the two methods, with a standard deviation of 4% at a significance level of 0.05, with a power of 90%.

**RESULTS**

A total of 19 males and 40 females were enrolled in this study. The mean age of enrolled subjects was 41.3 ± 14.2 years (range: 20 to 74 years) and the mean spherical equivalent was -1.3 ± 2.7 diopters (range: -9.75 to +3 diopters).

Pair-wise comparisons of oxygen saturation and vessel width measurements in normal eyes between observers on each method are displayed in (Table 1). Global measurements obtained by the two observers were comparable for SaO\textsubscript{2} and SvO\textsubscript{2} using the Full-Field method, and for SvO\textsubscript{2}, SaO\textsubscript{2}, venous width and arterial width using the Ring method (all P > 0.05). In contrast, venous width (P = 0.03) and arterial width (P = 0.04) measured by the two observers using the Full-Field method were significantly different. On the quadrant level (results not shown in Table 1), measurements by the two observers were significantly different on the Full-Field method for superotemporal SvO\textsubscript{2} (P = 0.03), superotemporal venous width (P = 0.03), inferonasal arterial width (P = 0.001), and superonasal arterial width (P = 0.02). All quadrant measurements from the two observers were comparable on the Ring method (all P > 0.05).

The comparison of global measurements obtained using both methods by each observer (Table 1) showed significant differences for SvO\textsubscript{2}, venous width, and arterial width (all P < 0.001). Neither observer found significant differences in SaO\textsubscript{2} measurements generated by the two methods (all P > 0.05). For quadrant measurements (not shown in Table 1), both observers found significant differences between measurements generated by the two methods for inferonasal SvO\textsubscript{2} (P < 0.001), superonasal SvO\textsubscript{2} (P < 0.001), inferotemporal SvO\textsubscript{2} (P = 0.006), superonasal venous width (P = 0.03), inferotemporal venous width (P < 0.001), superonasal arterial width (P = 0.001), inferotemporal arterial width (P = 0.001), and superonasal arterial width (P = 0.03).

The results of the agreement between observers and between processing methods in normal eyes are shown in (Table 2). The mean differences between global SO\textsubscript{2} measurements obtained by the two observers were about zero and ranged from -0.047% to -0.032% with the Full-Field method and from -0.071% to -0.06% with the Ring method. Similarly, interobserver mean differences in vessel width ranged from 0.025 pixels to 0.035 pixels with the Full-Field method and from -0.013 pixels to 0.006 pixels with the Ring method. The 95% limits of agreement were narrow for all parameters. The mean differences between SO\textsubscript{2} measurements generated using the two methods ranged between -1.70% and -0.049% for observer #1 and between -1.76% and -0.049% for observer #2. In both scenarios, 95% of the differences were within the 95% limits of agreement, suggesting good agreement between observers and between methods (Figure 4). However, the differences between observers and between methods were significantly different for SaO\textsubscript{2}, venous, arterial, and arterial width (all P < 0.05).
as shown in (Figure 4), the 95% limits of agreement between methods were slightly wider for most parameters. The comparison of mean differences between observers using both methods or between methods used by both observers did not reveal a statistically significant difference for any of the parameters in global (Table 2) or quadrant measurements (not shown).

**DISCUSSION**

Since retinal hypoperfusion and hypoxia have been implicated in the development of a number of potentially blinding diseases, there has been increased interest in developing tools for objective and non-invasive measurement of retinal SO₂ over the last several decades. Oxymap T1 is one of the devices that have emerged as a result of this interest; this device provides a technology capable of measuring retinal SO₂ levels, and it may potentially become a useful adjunct in the management of hypoxia-induced retinal diseases. However, prior to establishing any technology for use in the diagnostic confirmation of disease stability or progression, it is critical to assess measurement variability, which can be influenced by factors such as using different operators, devices, or image acquisition and processing methods. In the clinical setting, images are often acquired by different operators using the same device, and they may be processed using different methods by different observers during the course of a disease. The purpose of this study was to evaluate the inter-method agreement of retinal SO₂ measurements between two image processing methods used by the same observer as well as the inter-observer agreement between two observers using the same image processing method. Although previous work has been done to determine reproducibility and repeatability of retinal SO₂ measurements of vessel segments, [14-16] no studies have been conducted to determine the agreement between different retinal oximetry image processing methods. Herein we evaluated for the first time the agreement between two image processing methods and the agreement between observers in measuring retinal SO₂.

Interpretation of the Bland-Altman analysis is based on the magnitude of the difference between values obtained using two methods or two observers to quantify the same variable. This difference between repeated measurements is considered non-significant when it is small and not sufficient to cause concerns with clinical interpretation [18]. In other words, if the results of the repeated measurements are constant, the results of either of the image processing methods will not be significantly affected by random error, [19] and both methods can be used interchangeably. In the present study, the absolute differences between observers for all arterial and venous parameters were almost zero and the 95% confidence limits of agreement contained 95% of the differences and were narrow for both image processing methods. This finding represents good agreement between observers, suggesting different observers may use either processing method without concern for introducing significant measurement discrepancies.

Greater inter-observer differences with the older Oxymap software existed because observers had to both identify and then manually select vessels in order to generate SO₂ measurements. Although the newer version of the Oxymap software automatically identifies vessels, it is still necessary for observers to manually select these vessels in order for the software to generate SO₂ values. Since manual vessel selection is still a necessary part of retinal oximetry image processing and data generation, some degree of variability and human error is still attributable to the observer. Thus, observers must be well-trained with both processing methods before using them.

Although there was good overall agreement between the two methods based on the small differences between processing methods used by either observer, the 95% limits of agreement were slightly larger compared to those of the agreement between observers. For example, limits of agreement of ±3.02% for SvO₂ or SaO₂ (observer #1) and ±2.66% for SaO₂ (observer #2) were wide and may not be clinically acceptable. Thus, although there was some degree of agreement between the two image processing methods, switching between methods during patient follow-up should be done cautiously. We therefore recommend adhering to only one processing method so potential measurement discrepancies can be avoided. There are two possible explanations for the significant differences observed on paired t-tests (Table 1) and for the wide confidence limits of agreement between the two methods. First, the size of the measurement area between methods differs: Full-Field method measurements are generated from a larger area that includes the peripapillary and peripheral retina, whereas the Ring method has a smaller effective measurement area that is limited to the peripapillary zone. The large exclusion area in the Ring method therefore decreases the amount of peripheral, higher-order vessels available for processing. Differences in SO₂ values generated by the Full-Field versus Ring methods may therefore be attributed to the generally narrower diameter of these peripheral vessels. Second, the slightly wide limits of agreement may solely reflect the difference between algorithms on which the two methods are based. However, if this latter explanation were true, the wide variability in measurements would have also been seen on inter-observer comparison.

Variability of biological parameters within the same research group, between research groups, between instruments, and between methods has long been a practical concern. Currently, research groups using the Oxymap T1 device are utilizing various methods for image processing. However, if retinal oximetry is to be used in clinical decision-making, comparability of results among different institutions and healthcare providers is paramount. Our group has outlined two methods for retinal oximetry image processing in a standard operating procedure format to facilitate the comparison. While the decision of which of the two processing methods to employ should be left to the discretion of the researcher, it is important to note that the Full-Field method requires more time (an average of 30 minutes per image) to accurately select retinal vessels extending into the periphery of the fundus image. The Ring method is less time-consuming (an average of 15 minutes per image) since it includes a smaller portion of the image for processing and measurement, thereby making it easier to use. Also, additional exclusion criteria in the Ring method suggest that this method may render more accurate measurements since vessel areas that are known to be highly variable and inaccurate are excluded. As retinal oximetry may be utilized for diagnostic purposes in the future, it is practical to utilize the least time-consuming processing method in order to maximize efficiency in the clinical setting.
Figure 4 Bland-Altman plots for agreement between observers in measuring average venous oxygen saturation (SvO₂) using the Full-Field method (A) and the Ring method (B), and in measuring average arterial oxygen saturation (SaO₂) with the Full-Field (C) and the Ring (D) method. Plots E through H depict agreement between the two methods in measuring average SvO₂ by observer #1 (E) and by observer #2 (F), and in measuring SaO₂ by observer #1 (G) and observer #2 (H). The plain horizontal lines in each plot indicate the mean difference whereas the upper and lower dashed lines represent the upper and lower 95% limits of agreement, respectively.
The fact that our study included only healthy eyes may limit the generalizability of our findings to diseased eyes, which were outside the scope of our current study. Since retinal vessel autoregulation is believed to play a key role in maintaining adequate oxygen supply despite changes in perfusion pressure and metabolic needs up to a certain threshold, it is possible that under certain pathologic conditions, dysregulation of retinal vessels may result in instability of oxygen saturation levels. Dysregulatory effects may lead to higher differences in \( SO_2 \) measurements between observers and between methods. Nevertheless, this study provides a framework to examine the agreement between observers and between methods to evaluate retinal oximetry measurements in various eye diseases in the future.

In conclusion, the two methodologies for retinal oximetry image processing described here can be reliably used to evaluate retinal \( SO_2 \). The agreement between observers was better than between processing methods; thus, the wide range of differences between the processing methods may limit their interchangeability. Also, we have demonstrated that multiple observers can measure retinal \( SO_2 \) parameters using either processing method without concern for introducing measurement variability.

**REFERENCES**


