



# Automation improves repeatability of retinal oximetry measurements

Robert Arnar Karlsson<sup>1,2</sup>, Olof Birna Olafsdottir<sup>1,4</sup>, Soumaya Belhadj<sup>5</sup>, Vedis Helgadóttir<sup>1</sup>, Thorunn Scheving Eliasdóttir<sup>3, 4</sup>, Einar Stefansson<sup>1, 4</sup>, Sveinn Hakon Hardarson<sup>1\*</sup>,

**1** Faculty of Medicine, University of Iceland, Reykjavik, Iceland

**2** Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavik, Iceland

**3** Faculty of Nursing, University of Iceland, Reykjavik, Iceland

**4** The National University Hospital of Iceland, Reykjavik, Iceland

**5** Eberhard Karls Universität, Tübingen, Germany

 These authors contributed equally to this work.

\* sveinnha@hi.is

## Abstract

### Purpose

Retinal oximetry is a technique based on spectrophotometry where images are analyzed with software capable of calculating vessel oxygen saturation and vessel diameter. In this study, the effect of automation of measurements of retinal vessel oxygen saturation and vessel diameter is explored.

### Methods

Until now, operators have had to choose each vessel segment to be measured explicitly. A new, automatic version of the software automatically selects the vessels once the operator defines a measurement area.

Five operators analyzed image pairs from the right eye of 23 healthy subjects with semiautomated retinal oximetry analysis software, Oxymap Analyzer (v2.5.1), and an automated version (v3.0). Inter- and intra-operator variability was investigated using the intraclass correlation coefficient (ICC) between oxygen saturation measurements of vessel segments in the same area of the retina.

### Results

For manual saturation measurements, the inter-rater ICC was 0.80 for arterioles and venules. For automated saturation measurements, the inter-rater ICC was 0.97 for arterioles and 0.96 for venules. For manual diameter measurements, the inter-rater ICC was 0.71 for arterioles and venules. For automated diameter measurements the inter-rater ICC was 0.97 for arterioles and 0.95 for venules. The inter-rater ICCs were different ( $p < 0.01$ ) between the semiautomated and automated version in all instances.

### Conclusion

Automated measurements of retinal oximetry values are more repeatable compared to measurements where vessels are selected manually.

## Introduction

Fundus photographs are widely used to document various ocular diseases such as diabetic retinopathy [1], glaucoma [2], retinopathy of prematurity [3] and retinal vein occlusion [4]. The retinal blood vessels are visible through the optics of the eye and can be imaged with a fundus camera. Changes in retinal vessel oxygen saturation have been found in diabetic retinopathy, glaucoma, age-related macular degeneration, and other common diseases affecting the retina [5–8] as well as systemic and brain disease [9].

Measurements of retinal vessel oxygen saturation were first attempted in 1959 [10], and several groups have experimented with different approaches since then (for review, see [11–14]).

Currently, there are three commercial retinal oximeters available. One of them is the Oxymap T1 which simultaneously acquires two images of the same area of the fundus at two different wavelengths of light. The two spectral images are processed by specialized software, Oxymap Analyzer. However, the Oxymap T1 is limited by the fact that measurements are semiautomated. As a result, users are required to choose each vessel segment being measured and classify it as an arteriole or venule. Manual input can lead to operator-induced variability in the measurement results. In addition, the analysis process is time-consuming and tedious. Therefore, new, automatic software has been developed which automatically detects the vessels, classifies them as arterioles or venules, and selects the vessels once the operator defines a measurement area. This paper evaluates the effect of automation on the repeatability of measurements of retinal vessel oxygen saturation and diameter.

## Methods

Five consecutive images of the right eye were taken with the optic nerve head located in different part of the image. The second and fifth image had the optic nerve head located in the center and those two images were used for analysis. All images were acquired in a dark room with the aiming light set to lowest setting. The flash intensity was set to 50 Ws. The time between flashes (images) never exceeded 30 seconds. The right eye from 23 healthy volunteers (aged between twenty and thirty years) was used for measuring.

Two optic-nerve-centered images were acquired with an interval of no more than two minutes, from the right eye of 23 healthy volunteers aged between twenty and thirty years.

Collection and analysis of images were done with the approval of The Icelandic Data Protection Authority with the informed consent of patients or volunteers and following the tenets of the Declaration of Helsinki.

## Retinal oximetry

The retinal oximeter (Oxymap T1; Oxymap ehf., Reykjavik, Iceland) consists of an image splitter and two digital cameras (Insight IN1800; Diagnostic Instruments Inc, Michigan, USA) attached to a fundus camera (Topcon TRC-50DX; Topcon Corporation, Tokyo, Japan). The device simultaneously acquires two images of the same area of the fundus at two different wavelengths of light, one sensitive to oxyhemoglobin (600 nm) and one isosbestic (570 nm), where oxyhemoglobin and hemoglobin absorb the same amount of light. The images are  $1200 \times 1600$  pixels and cover a  $50^\circ$  field of the retina.

The light absorbance of a solution can be described in terms of optical density (OD) at wavelength  $\lambda$ . Optical density is defined as  $OD_\lambda = \log(I_{0\lambda}/I_\lambda)$  where  $I_{0\lambda}$  is the intensity of the incoming light and  $I_\lambda$  is the intensity of the light after it has interacted with and been absorbed to some degree by the solution in question. Here  $I_{0\lambda}$  and  $I_\lambda$ ,

are estimated from brightness values chosen respectively from reflected light inside vessels, and the perivascular background in the fundus images [7].

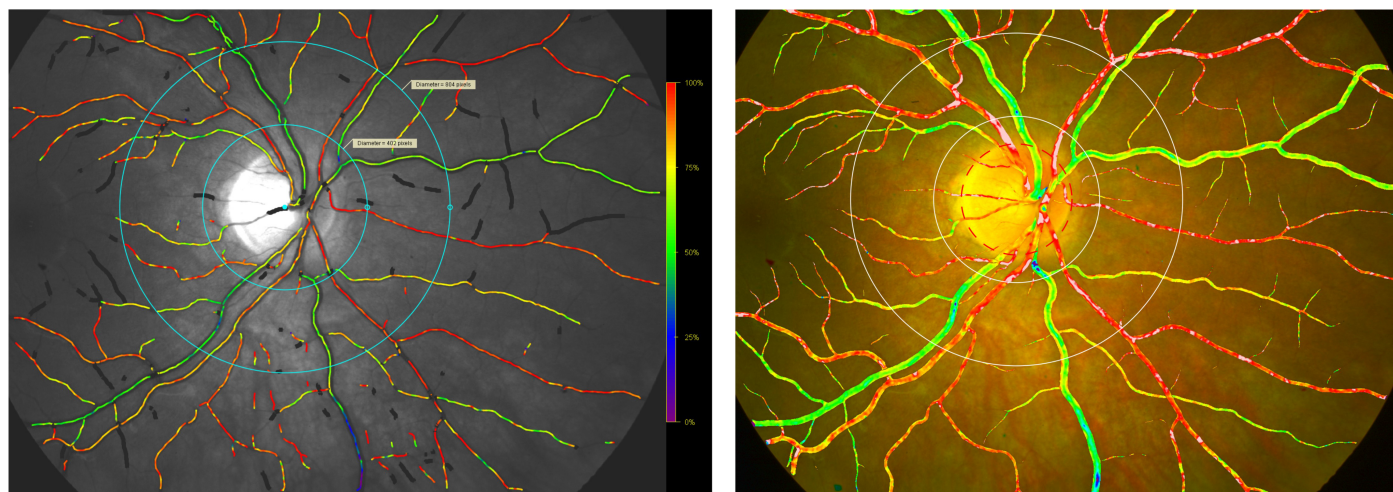
If one of the wavelengths is sensitive to oxyhemoglobin (e.g. 600nm) and one is isosbestic (e.g. 570nm), the ratio of the optical densities  $ODR = OD_{570nm}/OD_{600nm}$  at these two wavelengths will have an inverse and approximately linear relationship to the oxygen saturation (SatO2) [15]. Studies have demonstrated an artifact in the measurements caused by different diameters of the vessels [15–17] making it necessary to add a correction term.

The oxygen saturation is then calculated using  $SatO2 = a \cdot ODR + b + c \cdot D + k$  where  $D$  is the diameter of the vessel in pixels, and the parameters  $a$ ,  $b$ ,  $c$  and  $k$  can be calibrated based on measurements of arteries and veins of healthy volunteers and assuming values from saturation measurements performed in a study with a calibrated device [17, 18].

For version 2.5 of the software the calibration parameters were set to  $a = -1.28$ ,  $b = 1.24$ ,  $c = 0.0097$  and  $k = -0.14$ . For version 3.0 of the software the calibration parameters were set to  $a = -1.28$ ,  $b = 1.24$ ,  $c = 0.0095$  and  $k = -0.107$ .

In version 2.5  $I_{\lambda}$  is selected as the darkest pixel value along the cross-section of a vessel for each wavelength, meaning that saturation values are only calculated for points close to the center of a vessel (unless there is a central reflex from the vessel). For version 3.0, all points within a vessel are used to calculate saturation values.

## Analysis of oximetry images



**Fig 1.** Screenshots of the same image analyzed with version 2.5 (left) version 3.0 (right). Pseudocolor maps are generated automatically. Colors indicate the retinal vessel oxygen saturation. In healthy individuals arteries are normally colored orange to red (approximately 90% - 100% saturation). Veins are colored green to yellow-green (approximately 50% - 60% saturation). Both software versions detect the retinal vessels automatically and select measurement points inside and outside of the retina for automatic calculation of the retinal vessel oxygen saturation.

Images were analyzed with a previously validated semiautomatic software [19–21] (Oxymap Analyzer v.2.5.1, referred to as version 2.5) and new, automatic software (referred to as version 3.0). Figure 1 shows screenshots from both versions where the measurement area is defined.

The first step of the analysis process for both versions is to select the optic nerve head in the image manually. The measurement area is then defined as the area between an inner circle with a diameter 1.5 times the optic nerve diameter and an outer circle

with a diameter 3.0 times the diameter of the optic nerve diameter (Figure 1). The measurement area is defined to avoid reflection of light from the retinal nerve fiber layer around the optic nerve to make the measurements more accurate. For version 2.5, vessel segments with a diameter of 8 pixels or greater between the two circles were manually selected according to a standardized protocol (Oxymap ehf. protocol from November 2013) and classified as arteries or venules by the operator. The saturation and diameter values for each vessel segment would then be exported to an external application for further analysis. Mean values are calculated from all measured arterioles and all measured venules. In version 3.0, the software would automatically summarize the saturation and diameter values for arterioles and venules once the operator had selected the optic nerve head in the image. Median values of all measurement points are used, for arterioles and venules separately.

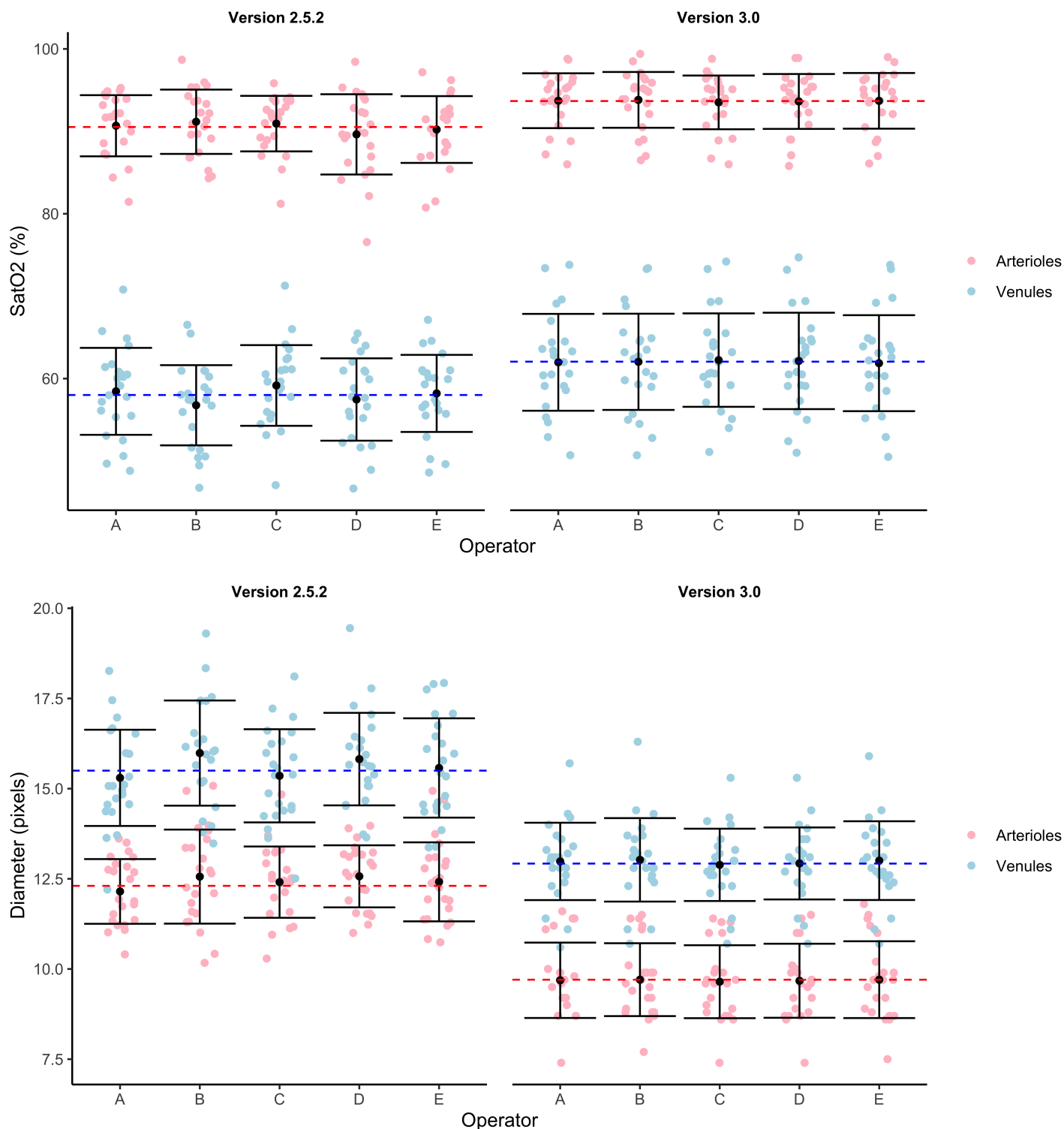
## Statistical analysis

The software versions are compared in several ways. The standard deviation of measured values is compared using an F-test in arterioles and venules for both saturation and diameter. Intraclass correlation coefficients (ICCs), and their 95% confidence intervals, are used to assess the intra ( $ICC_{intra}$ ) and inter-operator ( $ICC_{inter}$ ) reliability using a two-way random factorial absolute agreement ANOVA model [22]. When interpreting reliability using the ICC scores, "values less than 0.5 are indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability" [22]. While the ICC method is widely used when comparing methods or operators, it has been pointed out that it can, under some circumstances, misleadingly indicate good performance [23]. We, therefore, analyze the repeatability of measurements with the two versions using the 95% limits of agreement method (LoA) [24]. The repeated measures are for all operators, separately for oxygen saturation and diameter, and are used to estimate the LoA for the two versions. As there are multiple observations per image the LoAs are calculated for multiple replicates with 95% CIs calculated using the MOVER method [25]. Finally, the time in seconds taken to analyze an image is measured for both versions. The mean and standard deviation are calculated. Statistical significance of the difference is evaluated using a two-tailed, paired t-test.

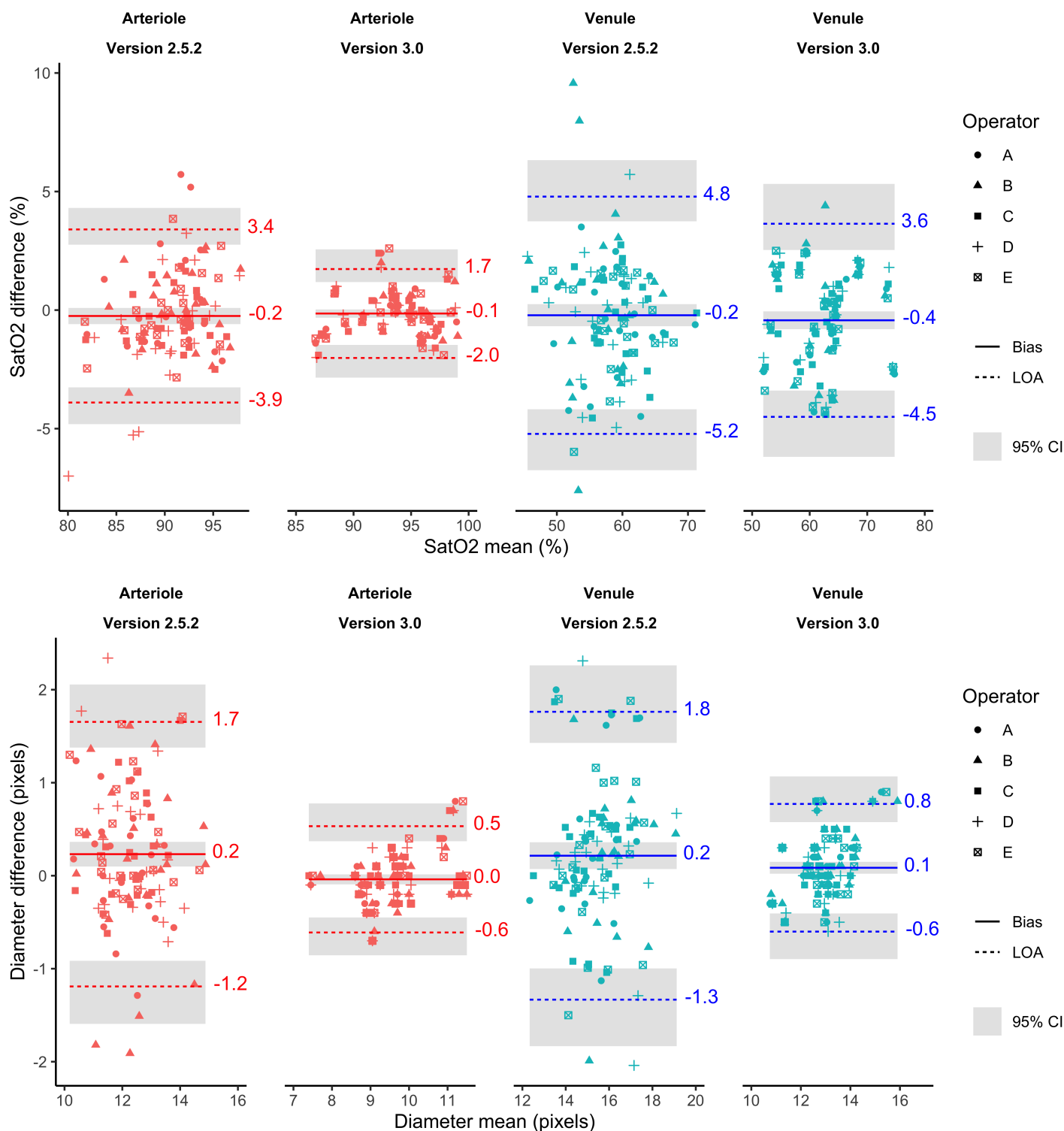
Statistical analysis and generation of plots and graphs was performed in R version 3.6.1 [26] using irrICC [27] version 1.0, ggplot2 [28] version 3.3.2 and boot [29, 30] version 1.3-25.

## Results

Figure 2 shows saturation and diameter measurements for the first image of each patient. Oxygen saturation in retinal arterioles was different between analysis software;  $93.7\% \pm 3.3\%$  for the automatic version 3.0 and  $90.7\% \pm 3.8\%$  for the semi-automatic version 2.5 ( $p < 0.0001$ , mean  $\pm$  SD). Venular oxygen saturation was lower measured with version 2.5,  $58.9\% \pm 5.1\%$  compared to version 3.0;  $62.3\% \pm 5.8\%$ , ( $p < 0.0001$ , mean  $\pm$  SD). Standard deviation between measured subjects was lower in version 3.0, compared to version 2.5 for arterioles ( $F = 1.47$ ,  $p = 0.04$ ) but the difference was statistically insignificant for venules ( $F = 0.74$ ,  $p = 0.12$ ). When diameter measurements were compared between analysis software, arteriolar diameter was decreased with the new version compared to the old version ( $9.7 \pm 1.1$  pixels vs.  $12.3 \pm 1.1$  pixels;  $p < 0.0001$ ) as well as the venular diameter ( $12.9 \pm 1.4$  pixels vs.  $15.5 \pm 1.0$  pixels;  $p < 0.0001$ ). The standard deviation of arteriolar diameter measurements was not significantly altered



**Fig 2.** Measured vessel saturation (top) and diameter (bottom) values from the first image for the two versions and all operators. Light red and blue points show individual measurement values, dashed red and blue lines show the global mean values for arterioles and venules, respectively. Black dots show the mean, and error bars show the standard deviation around the mean for each operator and vessel type.



**Fig 3.** Analysis of repeatability of retinal vessel oximetry (top) and diameter (bottom). The same area in two images acquired within a short time interval is measured using the two software versions for arterioles and venules. The vertical axis shows the difference between two saturation and diameter measurements in the same area (first image minus second image), and the horizontal axis shows their mean. The solid line in the middle shows the mean difference, and the other two broken lines show Limits of Agreement (LoA) or the two standard deviations of the difference. Results are colored red for arterioles and blue for venules. Grey shaded areas show the 95% confidence interval around the bias and LoA.

( $F = 1.22$ ,  $p = 0.14$ ) but standard deviation of venular measurements was significantly less for version 3.0 compared to version 2.5 ( $F = 1.83$ ,  $p < 0.0001$ ).  
123  
124

Figure 3 shows Bland-Altman plots for saturation and diameter measurements.  
125

Table 1 shows intra and inter-rater reliability when measuring vessel oxygen saturation.  
126  
127

**Table 1.** Intra and inter-rater reliability (ICC) when measuring vessel oxygen saturation using the two versions. Measurements from the same area of 23 right eyes in 23 subjects.

	Version	ICC <sub>intra</sub> [95% CI]	ICC <sub>inter</sub> [95% CI]
Arterioles	2.5	0.80 [0.70, 0.90]	0.81 [0.71, 0.90]
Arterioles	3.0	0.97 [0.95, 0.99]	0.97 [0.95, 0.99]
Venules	2.5	0.80 [0.68, 0.90]	0.84 [0.75, 0.92]
Venules	3.0	0.96 [0.94, 0.98]	0.96 [0.94, 0.98]

For arterioles measured with version 2.5 the intra and inter-rater reliability is moderate to good [22]. For venules measured with version 2.5 the intra-rater reliability is moderate to good and the inter-rater reliability is good to excellent. For arterioles and venules measured with version 3.0 the intra and inter-rater reliability is excellent.  
128  
129  
130  
131

Table 2 shows the intra and inter-rater reliability when measuring vessel diameter. For arterioles and venules measured with version 2.5 the intra and inter-rater reliability is moderate to good. For arterioles and venules measured with version 3.0 the intra and inter-rater reliability is excellent.  
132  
133  
134  
135

**Table 2.** Intra and inter-rater reliability (ICC) when measuring vessel diameter using the two versions. Measurements from the same area of 23 right eyes in 23 subjects.

	Version	ICC <sub>intra</sub> [95% CI]	ICC <sub>inter</sub> [95% CI]
Arterioles	2.5	0.71 [0.58, 0.84]	0.73 [0.60, 0.85]
Arterioles	3.0	0.97 [0.96, 0.99]	0.97 [0.96, 0.99]
Venules	2.5	0.71 [0.56, 0.84]	0.76 [0.65, 0.87]
Venules	3.0	0.95 [0.91, 0.98]	0.95 [0.92, 0.98]

The analysis time per image was  $617 \pm 187$  seconds for version 2.5 and  $29 \pm 7$  seconds for version 3. Using version 3.0 the analysis of images was significantly faster than for version 2.5 ( $p = 0.032$ ).  
136  
137  
138

## Discussion and conclusions

  
139

This study compares two versions of software capable of measuring retinal vessel oxygen saturation and vessel diameters. Vessels from the same area were measured in two images taken of the same subject with a short time interval. One version (v3.0) automatically selects the vessels to be measured and classifies them as arterioles or venules, while the other version (v2.5) requires operator input for these tasks.  
140  
141  
142  
143  
144

Version 3.0 measures higher oxygen saturation values for both arterioles and venules than version 2.5 but the variability between measured subjects is similar or slightly lower for version 3.0.  
145  
146  
147

The limits of agreement analysis for oxygen saturation indicate that repeatability is substantially improved for arterioles but less so for venules. The intra and inter-rater ICCs indicate that repeatability, and agreement between operators is improved for version 3.0 compared with version 2.5.  
148  
149  
150  
151

When measuring vessel diameter version, 3.0 performs substantially better for both arterioles and venules compared with version 2.5. The lower measured diameter in  
152  
153

version 3.0 versus version 2.5 can be attributed to superior vessel detection where the boundaries of vessels are more precisely located. Measurements with version 3.0 therefore include some of the smaller vessels that were not detected by version 2.5. As this might lead one to suspect that the lower LoA was simply an artifact of these lower measurement values, the diameter measurements were also compared after the values from version 3.0 had been scaled to have the same mean value as the diameter measurements from version 2.5. Even after this adjustment, the widest LoA for version 3.0, taking into account the outer limits of the 95%CI, was narrower than the narrowest LoA for version 2.5, where the inner limits of the 95% CI were used. We, therefore, conclude that there is substantial evidence that the diameter measurements for version 3.0 are more repeatable than for version 2.5.

This study has several limitations which should be addressed in follow up studies. Here, only images from young, healthy volunteers were used. While anecdotal evidence shows that the automated software works well with images from patients suffering from conditions such as CRVO and diabetic retinopathy the performance should be validated beyond the healthy eye.

In conclusion, while both versions of the software evaluate the oxygen saturation in a similar manner, the measured values are slightly different, which means that results from the two software versions are not directly comparable. The automatic software (Version 3.0) gives more consistent results, both for repeated measurements and between operators, than the manual software (version 2.5). Finally, the automated measurements are substantially faster than the manual measurements.

## 1 Acknowledgments

**Funding:** This research did not receive any specific funding from agencies in the public, commercial, or not- for-profit sectors.

**Competing interests:** This work makes use of images acquired with the Oxymap T1. R. A. Karlsson and S. H. Hardarson are shareholders in Oxymap ehf.

## References

1. Wilkinson CP, Ferris FL, Klein RE, Lee PP, Agardh CD, Davis M, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology*. 2003;110(9):1677–1682. doi:10.1016/S0161-6420(03)00475-5.
2. Chan KKW, Tang F, Tham CCY, Young AL, Cheung CY. Retinal vasculature in glaucoma: A review. *BMJ Open Ophthalmology*. 2017;1(1):1–14. doi:10.1136/bmjophth-2016-000032.
3. Brown JM, Campbell JP, Beers A, Chang K, Ostmo S, Chan RVP, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmology*. 2018;136(7):803–810. doi:10.1001/jamaophthalmol.2018.1934.
4. Hayreh SS, Zimmerman MB. Fundus changes in central retinal vein occlusion. *Retina*. 2015;35(1):29–42. doi:10.1097/IAE.0000000000000256.
5. Pournaras CJ, Rungger-Brandle E, Riva CE, Hardarson SH, Stefansson E. Regulation of retinal blood flow in health and disease. *Prog Retin Eye Res*. 2008;27(3):284–330.



6. Mozaffarieh M, Grieshaber MC, Flammer J. Oxygen and blood flow: Players in the pathogenesis of glaucoma. *Molecular Vision*. 2008;14:224–233. doi:v14/a28 [pii]. 198  
199  
200
7. Hardarson SH. Retinal oximetry. *Acta Ophthalmol*. 2013;91 Thesis:1–47. 201  
doi:10.1111/aos.12086. 202
8. Stefánsson E, Olafsdottir OB, Einarsdottir AB, Eliasdottir TS, Eysteinnsson T, Vehmeijer W, et al. Retinal Oximetry Discovers Novel Biomarkers in Retinal and Brain Diseases. *Investigative ophthalmology & visual science*. 2017;58(6):BIO227–BIO233. doi:10.1167/iovs.17-21776. 203  
204  
205  
206
9. Stefánsson E, Olafsdottir OB, Eliasdottir TS, Vehmeijer W, Einarsdottir AB, Bek T, et al. Retinal oximetry: Metabolic imaging for diseases of the retina and brain. *Progress in Retinal and Eye Research*. 2019;70:1–22. 207  
208  
209  
doi:10.1016/j.preteyeres.2019.04.001. 210
10. Hickam JB, Sieker HO, Frayser R. Studies of retinal circulation and A-V oxygen difference in man. *Trans Am Clin Climatol Assoc*. 1959;71:34–44. 211  
212
11. Beach J. Pathway to Retinal Oximetry. *Translational Vision Science & Technology*. 2014;3(5):2. doi:10.1167/tvst.3.5.2. 213  
214
12. Harris A, Dinn RB, Kagemann L, Rechtman E. A review of methods for human retinal oximetry. *Ophthalmic Surg Lasers Imaging*. 2003;34(2):152–164. 215  
216
13. Linsenmeier RA, Zhang HF. Retinal oxygen: from animals to humans. *Progress in Retinal and Eye Research*. 2017;58:115–151. 217  
218  
doi:10.1016/j.preteyeres.2017.01.003. 219
14. MacKenzie LE, Harvey AR, McNaught AI. Spectroscopic oximetry in the eye: a review. *Expert Review of Ophthalmology*. 2017;12(4):345–356. 220  
221  
doi:10.1080/17469899.2017.1318067. 222
15. Beach JM, Schwenzer KJ, Srinivas S, Kim D, Tiedeman JS. Oximetry of retinal vessels by dual-wavelength imaging: Calibration and influence of pigmentation. *Journal of Applied Physiology*. 1999;86(2):748–758. 223  
224  
225  
doi:10.1152/jappl.1999.86.2.748. 226
16. Hammer M, Vilser W, Riemer T, Schweitzer D. Retinal vessel oximetry-calibration, compensation for vessel diameter and fundus pigmentation, and reproducibility. *Journal of Biomedical Optics*. 2008;13(5):054015. 227  
228  
229  
doi:10.1117/1.2976032. 230
17. Geirsdottir A, Palsson O, Hardarson SH, Olafsdottir OB, Kristjansdottir JV, Stefánsson E. Retinal vessel oxygen saturation in healthy individuals. *Investigative Ophthalmology and Visual Science*. 2012;53(9):5433–5442. 231  
232  
233  
doi:10.1167/iovs.12-9912. 234
18. Hardarson SH, Harris A, Karlsson RA, Halldorsson GH, Kagemann L, Rechtman E, et al. Automatic retinal oximetry. *Investigative Ophthalmology and Visual Science*. 2006;47(11). doi:10.1167/iovs.06-0039. 235  
236  
237
19. Blondal R, Sturludottir MK, Hardarson SH, Halldorsson GH, Stefánsson E. Reliability of vessel diameter measurements with a retinal oximeter. *Graefe's Archive for Clinical and Experimental Ophthalmology*. 2011;249(9):1311–1317. 238  
239  
240  
doi:10.1007/s00417-011-1680-2. 241

20. Palsson O, Geirsdottir A, Hardarson SH, Olafsdottir OB, Kristjansdottir JV, Stefansson E. Retinal oximetry images must be standardized: a methodological analysis. *Invest Ophthalmol Vis Sci.* 2012;53(4):1729–1733. doi:10.1167/iovs.11-8621. 242  
243  
244  
245
21. Olafsdottir OB, Vandewalle E, Pinto LA, Geirsdottir A, De Clerck E, Stalmans P, et al. Retinal oxygen metabolism in healthy subjects and glaucoma patients. *British Journal of Ophthalmology.* 2014;98(3):329–333. doi:10.1136/bjophthalmol-2013-303162. 246  
247  
248  
249
22. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine.* 2016;15(2):155–163. doi:10.1016/j.jcm.2016.02.012. 250  
251  
252
23. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Computers in Biology and Medicine.* 1990;20(5):337–340. doi:https://doi.org/10.1016/0010-4825(90)90013-F. 253  
254  
255  
256
24. J Martin Bland DGA. STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT. *Lancet (London, England).* 1986;1:307–310. doi:10.1016/S0140-6736(86)90837-8. 257  
258  
259  
260
25. Zou GY. Confidence interval estimation for the Bland-Altman limits of agreement with multiple observations per individual. *Statistical Methods in Medical Research.* 2013;22(6):630–642. doi:10.1177/0962280211402548. 261  
262  
263
26. R Core Team. *A Language and Environment for Statistical Computing*; 2018. 264
27. Kilem L Gwet. *irrICC: Intraclass Correlations for Quantifying Inter-Rater Reliability*; 2019. 265  
266
28. Wickham H. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York; 2016. Available from: <https://ggplot2.tidyverse.org>. 267  
268
29. Davison AC, Hinkley DV. *Bootstrap Methods and Their Applications.* Cambridge: Cambridge University Press; 1997. Available from: <http://statwww.epfl.ch/davison/BMA/>. 269  
270  
271
30. Cauty A, Ripley BD. *boot: Bootstrap R (S-Plus) Functions*; 2020. 272