SDC

VALIDATION OF AN AUTOMATED FLUID ALGORITHM ON REAL-WORLD DATA OF NEOVASCULAR AGE-RELATED MACULAR DEGENERATION OVER FIVE YEARS

BIANCA S. GERENDAS, MD, PHD,* AMIR SADEGHIPOUR, PHD,*† MARTIN MICHL, MD,* FELIX GOLDBACH, MD,* GEORGIOS MYLONAS, MD, PHD,* ANASTASIIA GRUBER, MSC,‡ THOMAS ALTEN, MSC,* OLIVER LEINGANG, PHD,* STEFAN SACU, MD,* HRVOJE BOGUNOVIC, PHD,* URSULA SCHMIDT-ERFURTH, MD*

Background/Purpose: To apply an automated deep learning automated fluid algorithm on data from real-world management of patients with neovascular age-related macular degeneration for quantification of intraretinal/subretinal fluid volumes in optical coherence tomography images.

Methods: Data from the Vienna Imaging Biomarker Eye Study (VIBES, 2007–2018) were analyzed. Databases were filtered for treatment-naive neovascular age-related macular degeneration with a baseline optical coherence tomography and at least one follow-up and 1,127 eyes included. Visual acuity and optical coherence tomography at baseline, Months 1 to 3/Years 1 to 5, age, sex, and treatment number were included. Artificial intelligence and certified manual grading were compared in a subanalysis of 20%. Main outcome measures were fluid volumes.

Results: Intraretinal/subretinal fluid volumes were maximum at baseline (intraretinal fluid: 21.5/76.6/107.1 nL; subretinal fluid 13.7/86/262.5 nL in the 1/3/6-mm area). Intraretinal fluid decreased to 5 nL at M1-M3 (1-mm) and increased to 11 nL (Y1) and 16 nL (Y5). Subretinal fluid decreased to a mean of 4 nL at M1-M3 (1-mm) and remained stable below 7 nL until Y5. Intraretinal fluid was the only variable that reflected VA change over time. Comparison with human expert readings confirmed an area under the curve of >0.9.

Conclusion: The Vienna Fluid Monitor can precisely quantify fluid volumes in optical coherence tomography images from clinical routine over 5 years. Automated tools will introduce precision medicine based on fluid guidance into real-world management of exudative disease, improving clinical outcomes while saving resources.

RETINA 42:1673–1682, 2022

The introduction of antivascular endothelial growth factor (anti-VEGF) fundamentally changed the treatment of neovascular age-related macular degeneration (nAMD). Nonetheless, outcomes in real-world setting are poor for AMD.[1] High-resolution optical coherence tomography (OCT) imaging has become the gold standard for guiding treatment decisions in clinical practice for patients with nAMD but is susceptible to subjective evaluation and examiner error in clinical practice.[2] End points in clinical trials are still defined by central retinal subfield thickness (CST). However, the presence and location of intraretinal fluid (IRF) and subretinal fluid (SRF) are gaining ever more interest in the community.[2] Identification of OCT-derived biomarkers has already been shown to play a significant role in the evaluation of visual function, disease activity, and prognosis.[3] Unfortunately, because manual assessment of such imaging biomarkers is time consuming and reading center expertise is unavailable in clinical practice, investigators both in clinical practice and trials detect retinal fluid[4,5] with suboptimal accuracy, resulting in considerable disagreement regarding fluid-related disease activity on OCT images, even between retinal specialists.[6]

This situation has resulted in an intense focus on the development and evaluation of artificial intelligence (AI)–based deep learning algorithms for automated analysis of OCT images.[7] Previous studies using deep

learning approaches have demonstrated automated fluid detection and quantification with high segmentation accuracy.[8] These efforts opened the door for different applications of OCT in AMD, such as diagnosis, home monitoring and referral,[9] and biomarker identification,[10] as well as in other exudative diseases.[11,12] Studies have shown that OCT fluid algorithms can be successfully applied on data from clinical trials[8] and the segmentation of real-world data at one time point.[13] Schmidt-Erfurth et al[10] investigated the correlation of fluid volumes with prognostic outcomes[14] and performed a comprehensive fluid analysis in large prospective clinical trial data. Gerendas et al[11] and Roberts et al[12] investigated the value of AI-based fluid analysis in diabetic macular edema study data. Michl et al[15] performed a large post hoc analysis of data from 5 multicenter trials comprising 2,311 patients with nAMD, diabetic macular edema, or retinal vein occlusion, who received a flexible anti-VEGF therapy over a 12-month period. Keenan et al,[4] who evaluated the performance of retinal specialists in detecting retinal fluid presence in OCT scans of eyes with nAMD and compared their performance with an AI algorithm, showed a higher level of accuracy for the AI-based detection.

In general, many limitations associated with real-world data make it difficult to apply AI-based algorithms in a real-world setting. These limitations include an on average worse image quality, nonstandardized imaging, and treatment protocols and restricted reproducibility between different OCT devices. Applicability of AI-based tools in the real-world is crucial for determining the overall benefit of automated image analysis, especially when evaluating OCT scans over time. In contrast to mere clinical data, imaging data from real-world populations are sparse and derive from diverse cohorts of patients.[16] Moraes et al[13] applied a deep learning algorithm for automated quantification of OCT scans on baseline data from a large real-world cohort and demonstrated major differences regarding retinal features between eyes diagnosed with AMD that were first and second treated, and between eyes in different age and ethnicity groups, but did not report outcomes during therapy. In the current analysis, we investigated the application of an AI-based fluid algorithm on a large real-world data set of eyes with nAMD over a 5-year period. Furthermore, we evaluated automated performance by comparing expert human graders' appraisals of the presence and absence of fluid with quantitative thresholds determined by AI analyses.

From the *Department of Ophthalmology and Optometry, Medical University of Vienna, Spitalgasse, Vienna, Austria; From the †RetInSight GmbH, Elisabethstrasse, Vienna, Austria; and From the ‡Department of Medical Statistics, Medical University of Vienna, Spitalgasse, Vienna, Austria.

B. S. Gerendas: Roche (consultancy), Novartis (consultancy), Bayer (consultancy), D.X.S. (financial research support); S. Sacu: Roche (consultancy), Novartis (consultancy), Bayer (consultancy); H. Bogunovic: Apellis (financial research support), Heidelberg Engineering (financial research support); U. Schmidt-Erfurth: Scientific consultancy to Apellis, Novartis, Genentech, Roche, Kodiak, Heidelberg Engineering, and RetInSight GmbH.

None of the authors has any conflicting interests to disclose.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.retinajournal.com).

Reprint requests: Ursula Schmidt-Erfurth, MD, Department of Ophthalmology, Medical University of Vienna, Waehringer Guertel 18-20, A-1090 Vienna, Austria; e-mail: ursula.schmidt-erfurth@meduniwien.ac.at

Methods

All information about data collection, matching and filtering, patient selection, and subanalysis can be found in the **Supplemental Digital Content 1** (see **Figure S1**, http://links.lww.com/IAE/B753). In short, OCT data and additional patient information about treatments and function of 1,159 eyes were extracted from records of more than 25,000 patients in the VIBES registry at the Medical University of Vienna from 2007 to 2018. The Ethics Committee of the Medical University of Vienna approved the analysis, and no informed consent was necessary for this retrospective analysis. This study conformed the Declaration of Helsinki and the principles of Good Scientific Practice.

Automated Segmentation of Optical Coherence Tomography Scans

As a first step, the baseline OCT images of these 1,159 eyes were segmented with a validated automated deep learning fluid segmentation algorithm, the Vienna Fluid Monitor (RetInSight, Vienna, Austria) described earlier,[8] for the volume of SRF and IRF. In brief, the segmentation algorithm was a convolutional neural network consisting of an encoder and a decoder part and was trained with a series of local B-scan–level patches to predict for every pixel of the patch the likelihood for every pixel of the patch of belonging to one of the four classes: IRF, SRF, retina, or background. The pixels were assigned to IRF or SRF if the corresponding likelihood was above a predetermined threshold, and the resulting fluid compartment volumes were computed. Fluid was analyzed in the central 1-mm area, the central 3-mm and 6-mm disc area, and the total scan area. In addition, central subfield thickness between the internal limiting membrane and the outer border of the retinal pigment epithelium was segmented. After segmentation, all 1,159

scans were manually analyzed for correct foveal positioning and recentered if necessary by a retinal specialist and a grader certified by the Vienna Reading Center. During this process, 32 scans (2.7%) were excluded because of poor image quality that led to implausible fluid segmentations. Eleven thousand one hundred and twenty-seven eyes eyes were used for further analysis.

Statistical Analysis

Of these 1,127 eyes, scans at 1 month (M), 2 months, 3 months, 1 year (Y), 2, 3, 4, and 5 years of follow-up, always in relation to the first injection date, were selected. An available OCT scan at 30, 60, and 90 ± 10 days after the first injection was selected for each of the Months 1 to 3. An OCT scan for each of Years 1 to 5 was selected according to the following definition: first injection + number of years × 365 ± 60 days. Where more than one monthly or yearly scan was available, the scan acquired closest to the related monthly or yearly time point was selected. The mean time frame to the exact time point was calculated in days. All the scans included were segmented in the same way as the baseline scans. Figure 1 shows a representative B-scan with segmented fluid compartments.

Age, sex, and visual acuity values for all time points (defined as the closest visual acuity value available for each OCT scan) and the number of injections in the first 3 months and in each respective year were added from the patient records and the treatment database. A variety of forms of entry had been used for the original visual acuity values. Hence, a retinal specialist converted all available visual acuity values, i.e., letter scores and Snellen entries, to both decimal and logMAR scale according to Lange et al[17] for statistical analysis. Data were first analyzed in a descriptive manner and by using boxplots for each respective fluid compartment and area over time. In addition, for different time points, Spearman correlation coefficients were calculated between visual acuity and, respectively, IRF and SRF volume in the central mm and CST, but only for those eyes where there were no missing values.

Manual Subanalysis of Fluid in a Random Subset of Eyes

Certified graders, supervised by a retinal specialist, analyzed a random subset of 20% of eyes (n = 222) to determine whether intraretinal and/or subretinal fluid was present in the entire scan volume and the central millimeter. Fluid volumes determined by the algorithm

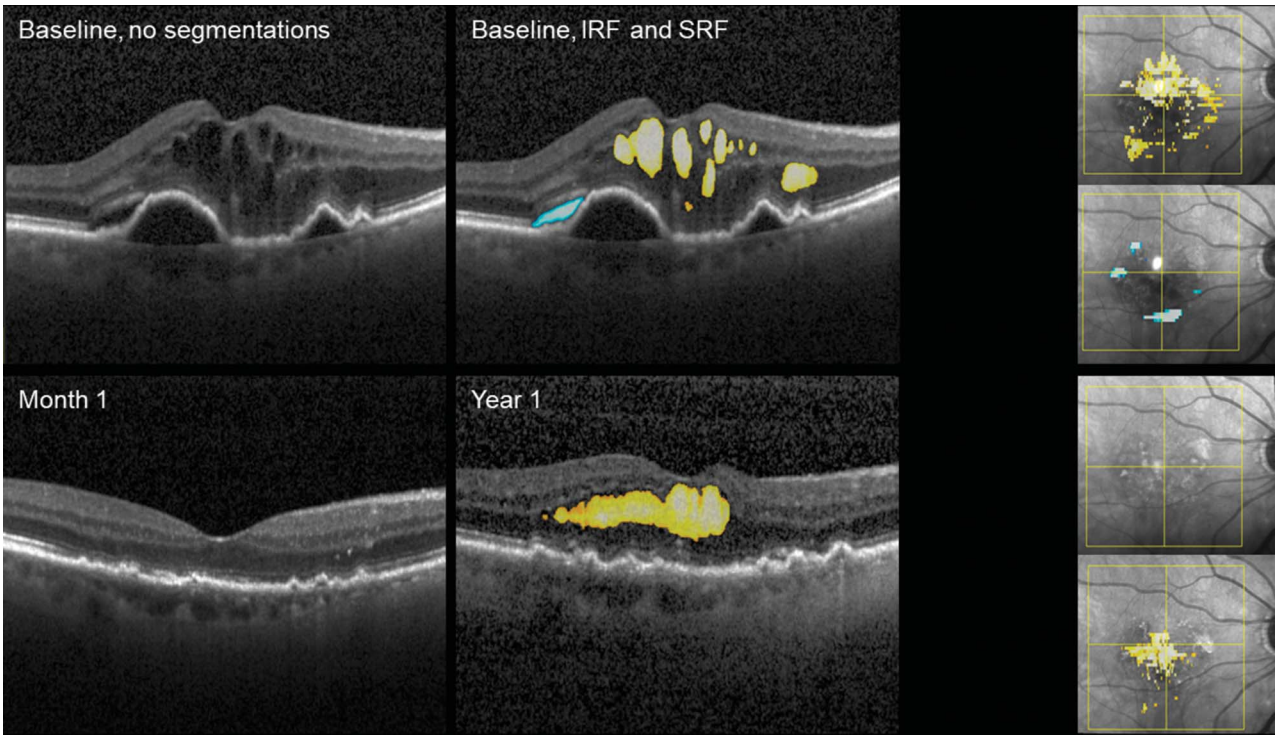

Fig. 1. Representative OCT B-scan with quantified retinal fluid. Representative OCT B-scan of an eye with nAMD showing IRF and SRF (upper left) by automated fluid segmentation (upper middle) for IRF (yellow) and SRF (blue); same patient with no fluid identified at Month 1 (lower left) and with IRF segmentation at Year 1 (lower middle). The panel on the right shows an en-face view of the fluid segmentations of that patient.

were compared in the groups with and without fluid presence as determined by the reading center. To evaluate the detection performance of the automated algorithm, areas under the receiver-operating characteristic curve were calculated and sensitivity and specificity values at the optimal operating point reported.

Results

Eleven thousand one hundred and twenty-seven eyes of 1,049 patients were included in the analysis set at baseline. Demographic information and treatment information as well as information about the OCTs used and the time points for all scans are available in the **Supplemental Digital Content 2** (see **Results, Table and Figures S2–S7**, http://links.lww.com/IAE/B754). The baseline OCT scan was acquired at a mean of 11.2 days before the first injection.

Visual Acuity and Central Retinal Subfield Thickness Values Over Time

The mean baseline visual acuity of all the eyes included was 0.36 (decimal) and 0.63 (logMAR, equals a logMAR score of 53 and a Snellen equivalent of 20/85), increasing to a maximum of 0.6 (decimal) and 0.4 (logMAR, equals a logMAR score of 65 and a Snellen equivalent of 20/50) at M2, and decreasing again to baseline levels or below by year 3 (0.36 decimal, 0.79 logMAR [equals a logMAR score of 46 and a Snellen equivalent of 20/123]). Figure 2 shows boxplots for the mean visual acuity course over time in logMAR visual acuity values (top). The mean central subfield thickness was 358 μm at baseline and decreased to a mean of 280 to 303 μm at all the subsequent time points (Figure 2, bottom). Figure 4 shows visual acuity versus fluid measurements in both compartments versus CST over time.

Volumes of Intraretinal and Subretinal Fluid

Intraretinal and subretinal fluid volumes reached their maximum at baseline. The mean IRF volume at baseline was 22 nL in the central mm, 77 nL in the 3-mm, and 107 nL in the 6-mm disc area. Intraretinal fluid decreased to a mean of 4–5 nL during M1–M3 in the central mm and slightly increased again to 11 nL at Y1 and to 17 nL at Y5. The mean SRF volume at baseline was 13.7 nL in the central mm, 86 nL in the 3-mm, and 262 nL in the 6-mm area. Subsequently, SRF decreased to a mean of 3-5nL during M1–M3 in the central mm and never increased beyond 7.2 nL until Y5. The pattern of fluid resolution was consistent throughout the 1-mm, 3-mm, and 6-mm

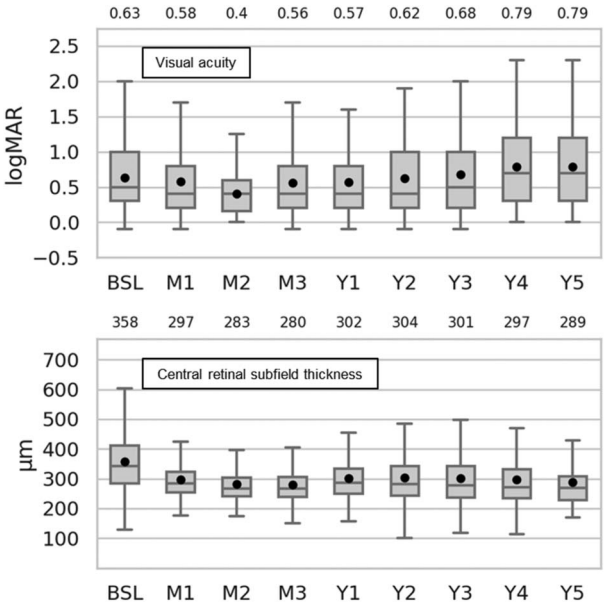

Fig. 2. Visual acuity and central retinal thickness over time. Top: boxplots for visual acuity values over time per time point in logMAR values. The mean values for logMAR visual acuity are given above the boxplots and are also indicated by the black dots in each boxplot. Please note that a logMAR value of 0.63 equals a logMAR score of 53 letters and a Snellen equivalent of 20/85 (0.58/56 and 20/76; 0.4/65 and 20/50; 0.56/58 and 20/73; 0.57/57 and 20/74; 0.62/54 and 20/83; 0.68/51 and 20/96; 0.79/46 and 20/123) Bottom: distribution of central retinal subfield thickness values per time point from baseline to Year 5. The mean central retinal subfield thickness values are given in micrometers above the boxplots and are also indicated by the black dots in each boxplot.

area, confirming the reproducibility of the automated fluid quantification. Fluid, IRF and particularly SRF, was seen in larger volumes in the 3-mm and 6-mm areas. The most intensive resolution for both fluid compartments was obtained during the loading dose. Recurrent fluid during long-term monitoring was most pronounced in the perifoveal and parafoveal area. Figure 3 shows boxplots of IRF and SRF volumes per time point. The therapeutic response pattern for visual acuity and fluid compartments, i.e. IRF, SRF, and CST, revealed the most pronounced time-sensitive fluid-function association between visual acuity and IRF in the central mm, as shown in Figure 4.

Spearman Correlation Coefficients

Spearman correlation coefficients between IRF/SRF volume in the central mm and CST with visual acuity also show that there is a small correlation between IRF volumes (correlation coefficients from up to 0.426) and visual acuity, whereas there is no correlation between SRF volume and visual acuity and CST with visual acuity values at any time point. In the definitely treatment-naive subset, these correlations are even moderate in Y01 to 04 in IRF volumes only. Table 1

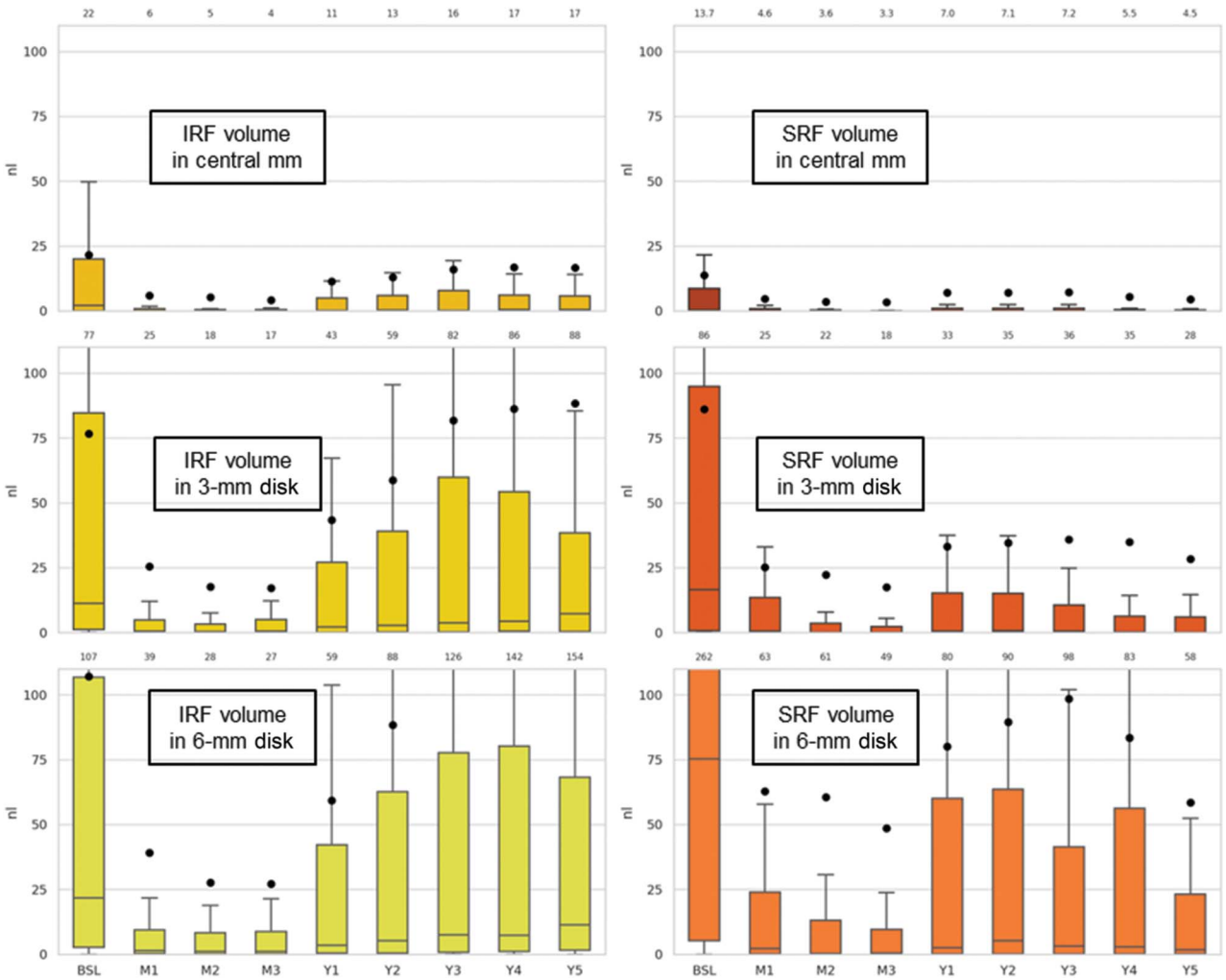

Fig. 3. Fluid quantifications over time. Intraretinal fluid (IRF, yellow) and subretinal fluid (SRF, orange) volumes over time from baseline to Year 5 in the central mm (top), 3-mm (middle), and 6-mm disc (bottom) areas. A consistent pattern is seen throughout all central and paracentral areas with the most pronounced therapeutic response during the loading phase in IRF. Note the preferentially parafoveal distribution of SRF in the 6-mm area. In contrast to IRF, SRF remains lower during the long-term follow-up, although there is more baseline SRF (mean of 262 nL) than IRF (mean of 107 nL). The mean values for fluid volumes are given above the boxplots and are also indicated by the black dots in each boxplot.

presents the Spearman correlation coefficients for both all eyes and definitely treatment-naive subset. The means for visual acuity, IRF, SRF volumes in the central mm, and CST do not differ remarkably from the means of the full data set described in the prior paragraphs.

Subanalysis of Manual Gradings

Human experts from the Vienna Reading Center determined the presence of fluid in the entire scan and in the central mm on each of the baseline scans from the subset of reading center analyzed images. The algorithm measured a mean of 186 nL (total volume) and 39.1 nL (central mm) in the scans where the reading center determined IRF to be present anywhere (n = 103) and in the central mm (n = 89), whereas it measured a mean of 15.6 nL (total volume) and 2.3 nL (central mm) of IRF in the scans where the reading center determined no IRF to be present (n = 106). The algorithm measured a mean of 350 nL (total volume) and 22.9 nL (central mm) in the scans where the reading center determined SRF to be present anywhere (n = 173) and in the central mm (n = 125), whereas it identified a mean of 20.6 nL (total volume) and 1.3 nL (central mm) of SRF in the scans where the reading center determined no SRF to be present (n = 37). Figure 5 presents the area under the curve plots for this subset of patients to determine sensitivity and specificity of the algorithm. Sensitivity was 0.80/0.88 and specificity 0.92/0.84 for the presence of IRF versus the volume of IRF, with an area under the curve of

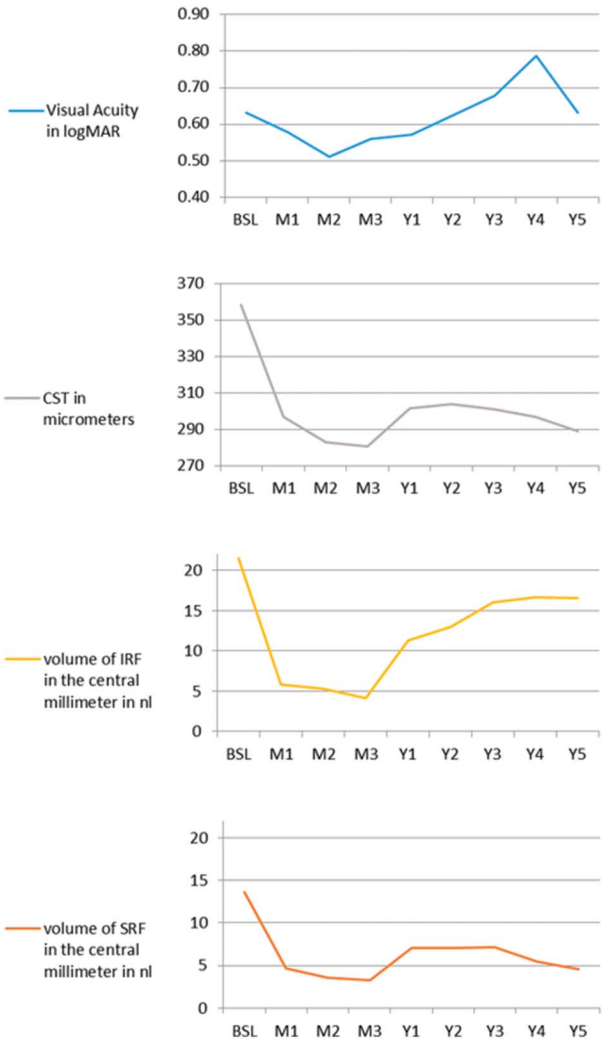

Fig. 4. Comparison of visual acuity, central retinal subfield thickness, and fluid quantifications over time. Comparison of change in function (visual acuity in logMAR, blue), CST (gray) in micrometers, and central millimeter fluid for IRF (yellow) and SRF (orange) in nanoliters (nL) over time. Please note that there were only very few patients in Year 4, contributing to an exceptionally low visual acuity value at that time point.

0.89/0.91 for the central mm/the entire scan. Sensitivity was 0.75/0.80 and specificity 0.73/0.81 for the presence of SRF versus the volume of SRF, with an area under the curve of 0.79/0.87 for the central mm/the entire scan. Segmentation examples of cases, where algorithm and human grader differed from each other, can be found in **Supplemental Digital Content 3** (see **Material**, http://links.lww.com/IAE/B754).

Discussion

The purpose of this study was to apply a validated deep learning algorithm, which has so far only been used on images from clinical studies, over time to real-world imaging data from patients under (real-world) treatment for nAMD. No standardized imaging or treatment protocol is used in such real-world data sets, and many images of suboptimal quality are seen. Applying developed algorithms on such data would be an important step toward their use in clinical routine. In addition, our findings highlight the clinical expectation that IRF correlates best with functional outcomes after treatment and that CST[18] and SRF are biomarkers of little value.

The lack of correlation of CST with clinical outcomes, retinal function, and even realistic fluid volumes has been convincingly demonstrated on clinical study data, particularly in nAMD and diabetic macular edema, which are the leading anti-VEGF indications.[18] There is also solid evidence that OCT evaluation is performed in a most subjective manner in clinical trials and even more so in clinical routine.[2,6] OCT is far away from meeting its initial promise as the "VEGF meter" suggested by Rosenfeld[19] in the first OCT-driven anti-VEGF trial PRONTO. Nevertheless, details regarding retinal fluid visualized by advanced raster scanning OCT, such as fluid localization and its change under therapy, are increasingly recognized as clinically and prognostically relevant.[3,20,21]

The aim of this project was to harness the objective power of automated algorithms for use in evaluating clinical performance in a large real-world population. Of note, it is the first study to our knowledge to use an automated fluid detection algorithm on a routine practice data set over a 5-year period to determine whether such advanced analysis tools can be applied to routine imaging and whether automated segmentations correlate with visual outcomes and treatment decisions. We were able to show that nonstandardized OCT monitoring data from a busy macula clinic can be successfully processed by accurate deep learning–based fluid microvolume quantification. Moreover, a validation of this method by human expert grading confirmed the results. Of note, this manual method is not used in clinical routine, but it replicates what reading centers do for standardized evaluation, showing that the algorithm will be a "more standardized" evaluation tool than today's clinical routine.

For fluid function correlation, our results highlight the concept that IRF volumes correlate well with visual acuity benefits. The pattern of fluid persistence and recurrence during long-term follow-up reflects the progressive loss of fluid control in routine anti-VEGF management, as shown in a multitude of real-world assessments.[22] The initial vision gain is reduced with progressive accumulation of IRF, underlining the need

Table 1. Spearman Correlation Coefficients for Each Time Point Between IRF/SRF Volume/CST With Visual Acuity

		Spearman Correlations All Eyes				Spearman Correlations Subset		
Time point	n	IRF	SRF	CST	n	IRF	SRF	CST
BSL	1,030	0.357	−0.104	0.191	106	0.348	−0.292	0.162
M1	241	0.300	−0.084	0.110	21	0.485	−0.191	−0.016
M2	188	0.220	−0.037	−0.036	19	0.369	−0.026	−0.125
M3	452	0.301	0.011	0.150	75	0.462	0.188	0.443
Y01	663	0.358	−0.121	0.110	103	0.509	−0.111	0.204
Y02	433	0.385	−0.152	0.150	64	0.515	−0.301	0.052
Y03	316	0.426	−0.159	0.166	39	0.603	−0.323	0.112
Y04	237	0.374	−0.066	0.201	25	0.537	−0.356	0.217
Y05	169	0.280	−0.136	0.033	16	−0.102	−0.406	−0.185

for a structured maintenance regimen based on objective fluid quantification. In a recent commentary on poorer outcomes in real-world studies of anti-VEGF therapy for nAMD, Hsu concluded that "treating physicians may be the root cause"[23] of disappointing results. Yet, the dimension of clinical need in real-world patient management can only be handled by physicians who are empowered by the appropriate tools to cope with overwhelming numbers of patients and images.

Realistically, every real-world treatment can be considered adequate if the functional level can be maintained over a period of a few years.[16] The U.S. Electronic Health Records analysis with its 30,106 patients included points to the challenge of large patient and treatment numbers: The overload led to limited visual acuity improvements coupled with only moderate anatomic improvements.[24] Because fluid is clearly the major variable that can be modulated by anti-VEGF therapy, it must be used in the most effective manner as a guidance for efficient therapy. This is the future role of automated fluid quantification in the real world.

We used a deep learning–based algorithm to automatically detect IRF and SRF in OCT scans of patients with nAMD undergoing routine treatment. Its overall performance was convincing, and the fluid volumes determined were comparable with those in nAMD clinical trials. Specifically, our algorithm detected a baseline mean value of 107 nL of IRF and an SRF mean volume of 262 nL in the 6-mm disk area. Schmidt-Erfurth et al described a mean of 123 to 161 nL of IRF and 415 to 464 nL of SRF at baseline in the HARBOR study,[10] and Michl et al[15] described a baseline median of 37 nL of IRF and 157 nL of SRF in the HARBOR and TREND studies. Moraes et al[13] reported a baseline mean/median of 73 to 118/3-7nL of IRF and 258 to 455/54 to 183 nL of SRF in a real-world data set of first-treated and second-treated eyes.

These numbers confirm that IRF measurements and the condition of disease activity are consistently similar in dimensions between trial and routine populations. By comparison, SRF volumes in clinical trial data sets seem to be larger than those seen in the real-world data set. Taking into consideration that Moraes et al distinguished between first-treated (455 nL) and second-treated (258 nL) eyes, their fluid volumes may reflect the duration of disease before initial presentation, while our data set represents more second-treated eyes, which are likely to reflect the real-world setting. The other reason for the discrepancy may be that prior treatment was not indicated in the patients' histories. When we examined the mean volume of SRF in a confirmed treatment-naive subgroup, however, we found a mean volume of SRF in the 6-mm area of 248 nL, i.e. comparable with that of the overall cohort. We also considered that as our algorithm itself was different from that of Moraes et al, it might generally have measured larger values for the volume of SRF, while measuring comparable IRF values. This could likewise be ruled out as the HARBOR data set, which was analyzed with a different algorithm from the one Moraes et al used, provided similar values to those Moraes et al presented. Therefore, we conclude that SRF volumes at baseline are generally higher than IRF volumes and may vary between populations.

Comparison between different reported fluid segmentation algorithms is difficult because they are not publicly available and have not been applied on the same imaging data sets. In particular, the algorithms developed by DeepMind[9,25] and the Notal OCT Analyzer[26] are the most closely related to our Vienna Fluid Monitor. For reported volumetric measurements, a comparison of our algorithm with the DeepMind algorithm is reported above and with Notal OCT Analyzer was part of the analysis reported by Keenan et al.[27] Ideally, the algorithms should be compared directly on the same open data set, like the one from the RETOUCH challenge[28] on fluid detection and quantification.

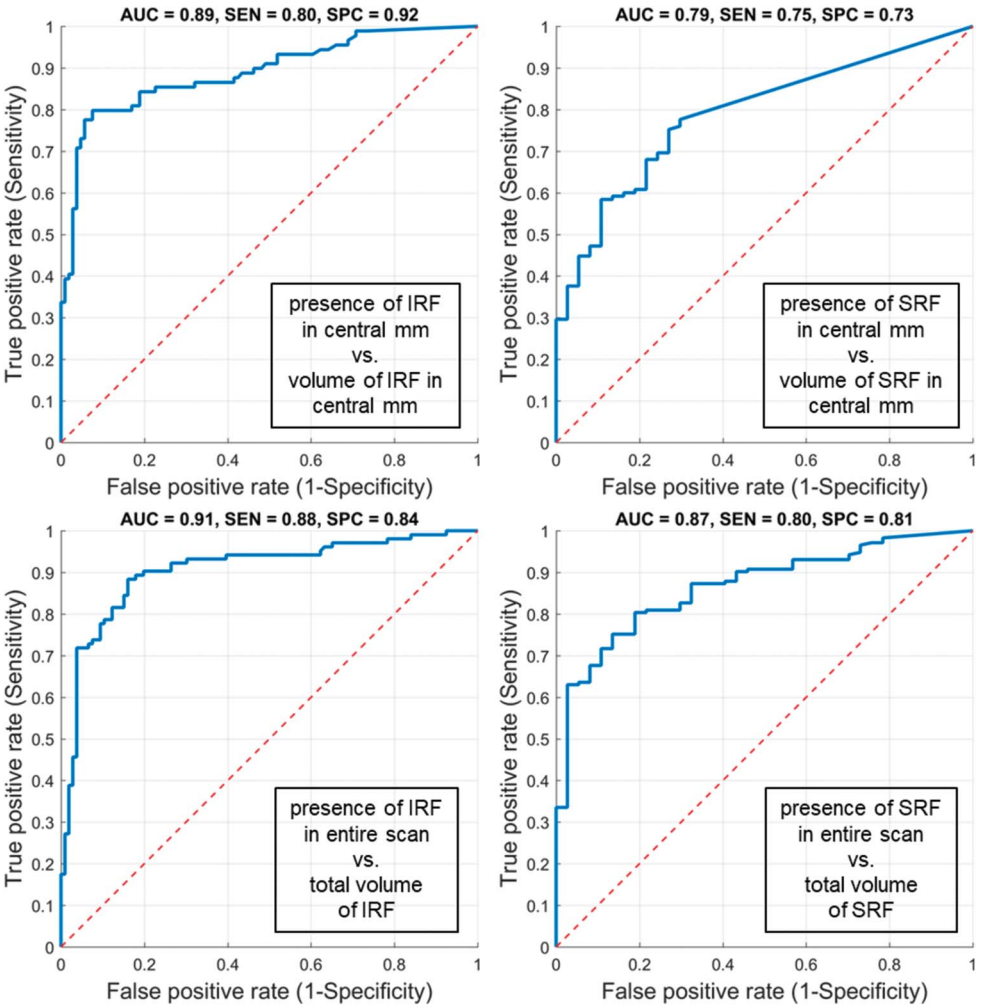

Fig. 5. Accuracy of manual versus automated detection of retinal fluid. Area under the curve plots demonstrating sensitivity and specificity of human expert grading versus automated algorithm for IRF presence in the central mm area (n = 89 presence, n = 106 absence) and the entire scan (n = 103 presence, n = 106 absence) (expert grader) versus volume of central mm and total fluid (algorithm) (left) and SRF presence in the central mm area (n = 125 presence, n = 37 absence) and the entire scan (n = 173 presence, n = 37 absence) (expert grader) versus volume of the central mm and total fluid (algorithm) (right). There are few values available for central SRF, which explains the lower overall accuracy in contrast to the other area under the curve plots.

A major finding of our study is the role of retinal fluid volume measurements compared with CST. A moderate relation between IRF and CST was apparent during the 5-year follow-up. Although CST values remained stable at around 300 μm as early as after the first injection at Month 1, the mean IRF volumes increased close to their baseline values after 3 years and did not decrease thereafter. This persistence impressively reflects the real-world problem of undertreatment.[29] Multiple publications show that visual acuity values are at best maintained for a limited time and in most cases decrease after more years of treatment. However, because OCT was usually not included, none of these real-world publications have provided an anatomic explanation for this pattern. Finger et al[30] looked at CST only in the WAVE study and found that it had decreased at the first follow-up by 99 μm from 349 μm at baseline and never increasing again more than by 20 μm. Our study clearly suggests that although stable CST values imply that the disease activity is under control during these years, there is a large amount of IRF present which spreads over a wider retinal area and does not significantly increase retinal thickness anymore. Yet, persistent IRF is most likely the reason for functional loss and a consequence of undertreatment. Undertreatment has been shown to be a most detrimental issue in the analyses of most real-world data sets where on average, no more than five injections were given during the first year.[31] The visual acuity increased better, with more letters gained in studies by Gillies et al[32] and Holz et al,[33] where the mean number of injections given during the first year was seven or more. Other studies have shown that a functional increase in the real-world population can only be achieved up to 4 months after baseline and visual acuity remains stable or decreased thereafter, as also seen in our cohort with a mean of 4.32 injections during the first year. The goal in an optimal treatment regimen is obviously to apply antifluid medication when fluid is truly present, and disease is active. This principle of disease activity is reflected in the definition of "fluctuations" and can ideally be managed by

precision measurement of small fluid volume changes such as IRF in nanoliters by the use of sensible AI tools.[34,35] A section on limitations of this study can be found in the supplementary material to this publication.

Conclusion

Our study provides proof-of-principle in a real-world space that our algorithm does not only detect fluid presence or absence in a qualitative manner, as routinely performed by the human expert, but also does so quantitatively at an IRF as well as SRF level. The comparison between reading center grading of fluid presence and absence and the low (almost zero) fluid volumes of the "absence" group confirms that the algorithm can measure very discrete volumes, indicating a low false positive rate.

In conclusion, we have demonstrated reliable performance of AI-based automated fluid detection and quantification in a large representative nAMD cohort with an extended real-world follow-up. Future prospective guidance of disease management using such precision tools presents a big promise for daily clinical routine by supporting clinicians with quantitative fluid volumetrics for their treatment decisions. This approach not only takes into consideration a single B-scan, as is often the case in a busy clinical practice, but also offers valuable pixel information of an entire OCT volume in real time. It not only introduces precision medicine into clinical practice but also provides enormous savings in healthcare budgets. Real-world practice is currently undergoing significant paradigm shifts: A growing proportion of retina experts chose to build their treatment decisions exclusively on OCT images (American Society of Retina Specialists, 2020, Global Trends in Retina). Trivizki et al[36] demonstrated that eliminating visual acuity assessments and ophthalmoscopic examinations in patients undergoing OCT-guided retreatment with intravitreal injections resulted in substantially decreased exposure times for patients and clinic staff, markedly decreased cost per encounter, and doubled patient volumes per clinic day. This may be an acceptable choice, provided the evaluation of the pixel-rich volumetric scan is performed in an objective and reliable manner, which is often beyond human capacity. Undertreatment because of an overworked healthcare system currently causes enormous deficits and psychophysical burdens in patient care. Intensive efforts to design and introduce pharmacological innovation and increased efficacy of substances are ongoing but often with marginal improvement in outcomes. Fully automated, broadly available, AI-based precision tools to guide treatment in a most efficient manner, ideally by a mouse click, will offer wide-ranging superior benefits to patients, physicians, and healthcare institutions.

Key words: deep learning, fluid monitoring, neovascular AMD, OCT, real-world management.

Acknowledgement

The authors would like to express their sincere gratitude to all physicians and orthoptists who contributed to the wealth of data in the VIBES registry, particularly the medical doctors from the Macula Clinic of the Department of Ophthalmology, Medical University of Vienna. Further they thank the staff of VRC, especially for diagnosis filtering from patient records and image gradings and Elise Langdon-Neuner for language editing of the manuscript.

References

1. Lim LS, Mitchell P, Seddon JM, et al. Age-related macular degeneration. Lancet 2012;379:1728–1738.
2. Schmidt-Erfurth U, Chong V, Loewenstein A, et al. Guidelines for the management of neovascular age-related macular degeneration by the European Society of Retina Specialists (EURETINA). Br J Ophthalmol 2014;98:1144–1167.
3. Schmidt-Erfurth U, Waldstein SM. A paradigm shift in imaging biomarkers in neovascular age-related macular degeneration. Prog Retin Eye Res 2016;50:1–24.
4. Keenan TD, Clemons TE, Domalpally A, et al. Retinal specialist versus artificial intelligence detection of retinal fluid from OCT: age-related eye disease study 2: 10-year follow-on study. Ophthalmology 2020;128:100–109.
5. Toth CA, Decroos FC, Ying GS, et al. Identification of fluid on optical coherence tomography by treating ophthalmologists versus a reading center in the comparison of age-related macular degeneration treatments trials. Retina 2015; 35:1303–1314.
6. DeCroos FC, Toth CA, Stinnett SS, et al. Optical coherence tomography grading reproducibility during the comparison of age-related macular degeneration treatments trials. Ophthalmology 2012;119:2549–2557.
7. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, et al. Artificial intelligence in retina. Prog Retin Eye Res 2018;67:1–29.
8. Schlegl T, Waldstein SM, Bogunovic H, et al. Fully automated detection and quantification of macular fluid in OCT using deep learning. Ophthalmology 2018;125:549–558.
9. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med 2018;24:1342–1350.
10. Schmidt-Erfurth U, Vogl W-D, Jampol LM, Bogunović H. Application of automated quantification of fluid volumes to anti–VEGF therapy of neovascular age-related macular degeneration. Ophthalmology 2020;127:1211–1219.

11. Gerendas BS, Bogunovic H, Sadeghipour A, et al. Computational image analysis for prognosis determination in DME. Vis Res 2017;139:204–210.
12. Roberts PK, Vogl WD, Gerendas BS, et al. Quantification of fluid resolution and visual acuity gain in patients with diabetic macular edema using deep learning: a post hoc analysis of a randomized clinical trial. JAMA Ophthalmol 2020;138:945–953.
13. Moraes G, Fu DJ, Wilson M, et al. Quantitative analysis of OCT for neovascular age-related macular degeneration using deep learning. Ophthalmology 2021;128:693–705.
14. Schmidt-Erfurth U, Bogunovic H, Sadeghipour A, et al. Machine learning to analyze the prognostic value of current imaging biomarkers in neovascular age-related macular degeneration. Ophthalmol Retina 2018;2:24–30.
15. Michl M, Fabianska M, Seeböck P, et al. Automated quantification of macular fluid in retinal diseases and their response to anti-VEGF therapy. Br J Ophthalmol 2022;106:113–120.
16. Daien V, Eldem BM, Talks JS, et al. Real-world data in retinal diseases treated with anti-vascular endothelial growth factor (anti-VEGF) therapy—a systematic approach to identify and characterize data sources. BMC Ophthalmol 2019;19:206.
17. Lange C, Feltgen N, Junker B, Schulze-Bonsel K, Bach M. Resolving the clinical acuity categories "hand motion" and "counting fingers" using the Freiburg Visual Acuity Test (FrACT). Graefe's Arch Clin Exp Ophthalmo 2009;247:137–142.
18. Pawloff M, Bogunovic H, Gruber A, et al. A systematic correlation of central subfield thickness (CSFT) with retinal fluid volumes quantified by deep learning in the major exudative macular diseases. Retina 2021;42:831–841.
19. Rosenfeld PJ. Optical coherence tomography and the development of antiangiogenic therapies in neovascular age-related macular degeneration. Invest Ophthalmol Vis Sci 2016;57: OCT14–OCT26.
20. Jaffe GJ, Ying GS, Toth CA, et al. Macular morphology and visual acuity in year five of the comparison of age-related macular degeneration treatments trials. Ophthalmology 2019;126:252–260.
21. Sharma S, Toth CA, Daniel E, et al. Macular morphology and visual acuity in the second year of the comparison of age-related macular degeneration treatments trials. Ophthalmology 2016;123:865–875.
22. Mehta H, Tufail A, Daien V, et al. Real-world outcomes in patients with neovascular age-related macular degeneration treated with intravitreal vascular endothelial growth factor inhibitors. Prog Retin Eye Res 2018;65:127–146.
23. Hsu J, Regillo CD. Poorer outcomes in real-world studies of anti-vascular endothelial growth factor therapy for neovascular age-related macular degeneration. Ophthalmology 2020;127: 1189–1190.
24. Kiss S, Campbell J, Almony A, et al. Management and outcomes for neovascular age-related macular degeneration: analysis of United States electronic health records. Ophthalmology 2020;127:1179–1188.
25. Wilson M, Chopra R, Wilson MZ, et al. Validation and clinical applicability of whole-volume automated segmentation of optical coherence tomography in retinal disease using deep learning. JAMA Ophthalmol 2021;139:964–973.
26. Chakravarthy U, Goldenberg D, Young G, et al. Automated identification of lesion activity in neovascular age-related macular degeneration. Ophthalmology 2016;123:1731–1736.
27. Keenan TDL, Chakravarthy U, Loewenstein A, et al. Automated quantitative assessment of retinal fluid volumes as important biomarkers in neovascular age-related macular degeneration. Am J Ophthalmol 2021;224:267–281.
28. Bogunovic H, Venhuizen F, Klimscha S, et al. RETOUCH: the retinal OCT fluid detection and segmentation benchmark and challenge. IEEE Trans Med Imaging 2019;38:1858–1874.
29. Holz FG, Figueroa MS, Bandello F, et al. Ranibizumab treatment in treatment-naive neovascular age-related macular degeneration: results from luminous, a global real-world study. Retina 2020;40:1673–1685.
30. Finger RP, Wiedemann P, Blumhagen F, et al. Treatment patterns, visual acuity and quality-of-life outcomes of the WAVE study—a noninterventional study of ranibizumab treatment for neovascular age-related macular degeneration in Germany. Acta Ophthalmol 2013;91:540–546.
31. Cohen SY, Mimoun G, Oubraham H, et al. Changes in visual acuity in patients with wet age-related macular degeneration treated with intravitreal ranibizumab in daily clinical practice: the LUMIERE study. Retina 2013;33:474–481.
32. Gillies MC, Walton R, Simpson JM, et al. Prospective audit of exudative age-related macular degeneration: 12-month outcomes in treatment-naive eyes. Invest Ophthalmol Vis Sci 2013;54:5754–5760.
33. Holz FG, Tadayoni R, Beatty S, et al. Key drivers of visual acuity gains in neovascular age-related macular degeneration in real life: findings from the AURA study. Br J Ophthalmol 2016;100:1623–1628.
34. Chakravarthy U, Havilio M, Syntosi A, et al. Impact of macular fluid volume fluctuations on visual acuity during anti-VEGF therapy in eyes with nAMD. Eye 2021;35:2983–2990.
35. Evans RN, Reeves BC, Maguire MG, et al. Associations of variation in retinal thickness with visual acuity and anatomic outcomes in eyes with neovascular age-related macular degeneration lesions treated with anti-vascular endothelial growth factor Agents. JAMA Ophthalmol 2020;138:1043–1051.
36. Trivizki O, Karp MR, Chawla A, et al. Eliminating visual acuity and dilated fundus examinations improves cost efficiency of performing optical coherence tomogrpahy–guided intravitreal injections. Am J Ophthalmol 2020;219:222–230.