



Published in final edited form as:

Ophthalmology. 2021 January ; 128(1): 100–109. doi:10.1016/j.ophtha.2020.06.038.

Retinal Specialist versus Artificial Intelligence Detection of Retinal Fluid from OCT:

Age-Related Eye Disease Study 2: 10-Year Follow-On Study

Tiarnan D.L. Keenan, BM BCh, PhD¹, Traci E. Clemons, PhD², Amitha Domalpally, MD, PhD³, Michael J. Elman, MD⁴, Moshe Havilio, PhD⁵, Elvira Agrón, MA¹, Gidi Benyamini, MBA⁵, Emily Y. Chew, MD¹

¹Division of Epidemiology and Clinical Applications, National Eye Institute, National Institutes of Health, Bethesda, Maryland. ²The EMMES Company, LLC, Rockville, Maryland. ³Fundus Photograph Reading Center, University of Wisconsin, Madison, Wisconsin. ⁴Elman Retina Group, Baltimore, Maryland. ⁵Notal Vision Ltd., Tel Aviv, Israel.

Abstract

Purpose: To evaluate the performance of retinal specialists in detecting retinal fluid presence in spectral domain OCT (SD-OCT) scans from eyes with age-related macular degeneration (AMD) and compare performance with an artificial intelligence algorithm.

Design: Prospective comparison of retinal fluid grades from human retinal specialists and the Notal OCT Analyzer (NOA) on SD-OCT scans from 2 common devices.

Participants: A total of 1127 eyes of 651 Age-Related Eye Disease Study 2 10-year Follow-On Study (AREDS2-10Y) participants with SD-OCT scans graded by reading center graders (as the ground truth).

Methods: The AREDS2-10Y investigators graded each SD-OCT scan for the presence/absence of intraretinal and subretinal fluid. Separately, the same scans were graded by the NOA.

Correspondence: Tiarnan D.L. Keenan, BM BCh, PhD, NIH, Building 10, CRC, Room 10D45, 10 Center Drive, MSC 1204, Bethesda, MD 20892-1204. tiarnan.keenan@nih.gov.

Author Contributions:

Conception and design: Keenan, Clemons, Domalpally, Havilio, Benyamini, Chew

Data collection: Keenan, Domalpally, Elman, Havilio, Benyamini, Chew

Analysis and interpretation: Keenan, Clemons, Domalpally, Elman, Havilio, Agrón, Benyamini, Chew

Obtained funding: Benyamini, Havilio; Study was performed as part of 7 of 8 authors' regular employment duties. No additional funding was provided.

Overall responsibility: Keenan, Clemons, Domalpally, Elman, Havilio, Agrón, Benyamini, Chew

Supplemental material available at www.aaojournal.org.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The author(s) have made the following disclosure(s): G.B. and M.H.: Employees of Notal Vision. No relevant conflicting relationship exists for the other authors.

HUMAN SUBJECTS: Human subjects were included in this study. Institutional review board approval was obtained at each clinical site, and written informed consent was obtained from all study participants. All research adhered to the tenets of the Declaration of Helsinki.

No animal subjects were used in this study.

Main Outcome Measures: Accuracy (primary), sensitivity, specificity, precision, and F1-score.

Results: Of the 1127 eyes, retinal fluid was present in 32.8%. For detecting retinal fluid, the investigators had an accuracy of 0.805 (95% confidence interval [CI], 0.780–0.828), a sensitivity of 0.468 (95% CI, 0.416–0.520), a specificity of 0.970 (95% CI, 0.955–0.981). The NOA metrics were 0.851 (95% CI, 0.829–0.871), 0.822 (95% CI, 0.779–0.859), 0.865 (95% CI, 0.839–0.889), respectively. For detecting intraretinal fluid, the investigator metrics were 0.815 (95% CI, 0.792–0.837), 0.403 (95% CI, 0.349–0.459), and 0.978 (95% CI, 0.966–0.987); the NOA metrics were 0.877 (95% CI, 0.857–0.896), 0.763 (95% CI, 0.713–0.808), and 0.922 (95% CI, 0.902–0.940), respectively. For detecting subretinal fluid, the investigator metrics were 0.946 (95% CI, 0.931–0.958), 0.583 (95% CI, 0.471–0.690), and 0.973 (95% CI, 0.962–0.982); the NOA metrics were 0.863 (95% CI, 0.842–0.882), 0.940 (95% CI, 0.867–0.980), and 0.857 (95% CI, 0.835–0.877), respectively.

Conclusions: In this large and challenging sample of SD-OCT scans obtained with 2 common devices, retinal specialists had imperfect accuracy and low sensitivity in detecting retinal fluid. This was particularly true for intraretinal fluid and difficult cases (with lower fluid volumes appearing on fewer B-scans). Artificial intelligence–based detection achieved a higher level of accuracy. This software tool could assist physicians in detecting retinal fluid, which is important for diagnostic, re-treatment, and prognostic tasks.

Age-related macular degeneration (AMD) is the leading cause of legal blindness in all developed countries.^{1,2} The standard of care for neovascular AMD is repeated intravitreal injection of anti-vascular endothelial growth factor drugs.³ Re-treatment decisions are based predominantly on frequent imaging with spectral domain OCT (SD-OCT).^{4,5} Specifically, the exudative activity of the neovascular disease is typically judged by qualitative assessments of intraretinal fluid and subretinal fluid presence by retinal specialists. Best practice requires that physicians should evaluate each of many individual B-scans, in the form of a macular cube, for the presence or absence of fluid. However, this is time-consuming, and it is possible that accuracy in routine clinical practice is suboptimal. Even at the clinical trial level, studies have demonstrated that investigators often miss retinal fluid on OCT imaging (compared with reading center evaluation).⁵ Indeed, disagreement regarding fluid presence exists even between expert graders at the reading center level.⁸

Multiple factors are creating a pressing demand for computer assistance with OCT analysis. Demographic changes mean that the prevalence of neovascular AMD is increasing substantially,^{9,10} and patients often need treatment and OCT monitoring for many years.^{11,12} Clinic burdens are driving the development of home-based OCT devices,¹³ but these will actually increase the generation of OCT data requiring expert evaluation. For these reasons, evaluation of the current performance of manual OCT interpretation and the development and evaluation of automated algorithms that assist in OCT analysis but do not replace clinical decision-making are high priorities.

So far, OCT technology has undergone rapid hardware improvements,¹⁴ but less attention has been devoted to software advances. Most of the software available on the commercial OCT viewing platforms relates primarily to the automated segmentation of retinal layers and thickness measurements of these layers.^{15–18} However, in neovascular AMD, thickness

measurements alone are not very useful for assisting in disease management or guiding visual prognosis.¹⁹ By contrast, the presence, location, and quantity of intraretinal fluid and subretinal fluid predict visual outcomes more accurately^{19–21} and are critical for re-treatment decisions.^{3,6,7} Software algorithms that help identify the presence, location, and quantity of fluid could be highly useful in clinical practice.

Artificial intelligence (AI) is attracting increasing attention in medicine. In ophthalmology, the potential advantages of AI-based image analysis include the rapid speed, high consistency, and quantitative nature of the analyses. Thus, it is possible that AI may provide improved software for helping clinicians perform OCT evaluation; in this respect, AI is used to provide a tool, rather than replacing human decision-making or making autonomous clinical judgments. The Notal OCT Analyzer (NOA, Notal Vision Ltd, Tel Aviv, Israel) is one such AI machine learning-based software tool. It automatically analyzes OCT data for the presence of intraretinal and subretinal fluid, in addition to multiple other qualitative and quantitative features.²² In a previous study of the NOA on 142 OCT macular cube scans obtained with the Cirrus device (Carl Zeiss Meditech, Dublin, CA) from one UK center, it had similar accuracy in the identification of fluid to that of 3 retinal specialists. It also showed robust performance in ranking B-scans by order of importance for the identification of fluid, which may be a useful aid for clinicians in handling large numbers of B-scans in their daily routine. Recent analysis of NOA performance on OCT data from the Tel Aviv Medical Center demonstrated similar results for the Spectralis device (Heidelberg Engineering, Heidelberg, Germany), in addition to quantification of the volume of intraretinal fluid and subretinal fluid.²³

The first aim of this study was to evaluate the performance of a large group of retinal specialists in the assessment of intraretinal and subretinal fluid presence in SD-OCT macular volume scans obtained with 2 commonly used devices (Cirrus and Spectralis), as part of the Age-Related Eye Disease Study 2 10-year Follow-On Study (AREDS2-10Y), which comprised data from 19 US study centers. The second aim was to compare the performance of the NOA and to evaluate the potential benefit of assisting physicians with this AI-based decision support tool.

Methods

Study Population

The population used for this study was participants in the AREDS2-10Y. The AREDS2 was a multicenter phase III randomized clinical trial that analyzed the effects of nutritional supplements on the course of AMD in people at moderate-to-high risk of progression to late AMD.²⁴ Its study design has been described previously.²⁴ In short, 4203 participants aged 50 to 85 years were recruited between 2006 and 2008 at 82 retinal specialty clinics in the United States. Inclusion criteria at enrollment were the presence of bilateral large drusen or late AMD in 1 eye and large drusen in the fellow eye. The primary outcome was the development of late AMD, defined as central geographic atrophy or neovascular AMD.

After close-out of the main study at 5 years, a subset of 709 participants from 19 of the study sites underwent a single repeat evaluation at 10 years in the form of the AREDS2-10Y.

For the follow-on study, institutional review board approval was obtained at each clinical site and written informed consent for the research was obtained from all study participants. The research was conducted under the Declaration of Helsinki and complied with the Health Insurance Portability and Accountability Act.

Study Procedures

The AREDS2-10Y visit comprised a comprehensive eye examination by certified study personnel using standardized protocols. This included SD-OCT imaging of the macula in both eyes, using the Cirrus (with a cube scan comprising 512 A-scans in each of the 128 B-scans covering a 20×20 degree area) or the Spectralis (with a high-speed volume scan comprising 97 B-scans, ART 9 [max 15] with standard orientation [0°], covering a 20×20 degree area). The study investigators evaluated the OCT images and recorded the presence or absence of intraretinal fluid and of subretinal fluid in the macula of each eye. The investigators were not aware of the current study, that is, that the same OCT images were to be read later by both a reading center and an AI-based algorithm.

The same OCT images were sent to the University of Wisconsin Fundus Photograph Reading Center, where they underwent evaluation by expert graders for the same features (i.e., presence or absence of intraretinal fluid and of subretinal fluid in the macula of each eye). Intraretinal fluid was defined as intraretinal cysts or thickening of the outer nuclear layer with loss of the normal retinal contour; outer retinal tubules and tractional cyst-like structures were not defined as intraretinal fluid. Subretinal fluid was defined as well-demarcated, usually bell-shaped, hyporeflective spaces between the ellipsoid zone and the retinal pigment epithelium (RPE). Presence in at least 2 consecutive B-scans was required. The graders were masked to the reports from the AREDS2-10Y investigators, to the grades from previous retinal imaging, and to all clinical data. Each OCT image was evaluated independently by 2 graders. In the case of disagreement between the 2 graders, a senior grader at the reading center adjudicated the final grade. These grades provided the ground truth labels for the OCT scans.

Notal OCT Analyzer

The development of the NOA has been described.²² In brief, a machine learning and image recognition computational technique was used to develop a classifier that distinguishes normal morphologic features from elevated or distorted contours that occur as a result of fluid presence between or within retinal tissue compartments. The algorithmic process includes 3 major steps:

1. Delineation of internal limiting membrane (ILM) and RPE, using several local and global image processing techniques including pixel-graph optimization.
2. Candidate fluid-region identification using standard image processing techniques.
3. Machine learning feature-based classification of the candidate regions to distinguish true from false fluid regions.

After these steps, several more classification steps occur, including distinguishing regions of intraretinal versus subretinal fluid, identifying vitreo-macular interface abnormalities such as epiretinal membrane, and identifying and quantifying RPE irregularities. The algorithm allows fully automated detection and quantification of fluid in the various tissue compartments, when applied to a macular volume scan from Cirrus or Spectralis SD-OCT devices. Validation of this algorithm has been reported.²² That level of performance justified the implementation of the NOA as an analytical tool included in the AREDS2-10Y, in the form of the current preplanned and prospective external validation study.

Evaluation of the Performance of the Human Clinical Practitioners and the Notal OCT Analyzer

The primary outcome was the performance of the investigators and the NOA, in terms of correctly identifying the presence or absence of fluid (intraretinal or subretinal) from the SD-OCT macular cube. The secondary outcomes were the performance in the separate identification of (i) intraretinal fluid and (ii) subretinal fluid. For each of the 3 analyses, the following metrics were calculated: accuracy, sensitivity, specificity, precision, and FI-score (which incorporates sensitivity and precision into a single metric). The primary performance metric was accuracy; the rest of the metrics were secondary performance metrics.

For exploratory analyses, we performed error analysis by examining characteristics of the true-positive and false-negative cases of both the human investigators and the NOA. This analysis included the following as proxies for case difficulty: (i) proportion of cases requiring reading center senior adjudication for fluid presence, and, for the eyes with fluid identified by the NOA; (ii) NOA-estimated quantity of retinal fluid; and (iii) NOA-estimated number of B-scans with fluid present. Another exploratory aspect of the study was to evaluate the potential benefits of adding an AI-based algorithm to clinical care. Therefore, we considered the potential increase in fluid detection possibly achievable through adding an AI-based algorithm to human performance; for this analysis, we examined cases with retinal fluid that were missed by the human investigators but correctly identified by the NOA. Statistical analysis was performed using SAS version 9.4 (SAS Institute Inc, Cary, NC).

Results

For the analysis of retinal fluid presence, of the 1400 eyes with OCT data available to the reading center, 136 (9.7%) were excluded because of a reading center assessment for retinal fluid of ungradable or questionable. Of the remaining 1264 eyes, 137 (10.8%) were excluded by the NOA. These were excluded for 1 of the 3 following reasons that are built into the NOA as an automated quality filter: (i) poor image quality; (ii) suspected erroneous ILM or RPE delineation; and (iii) presence of vitreo-macular interface abnormalities to a degree that prevents fluid assessment.

Thus, 1127 eyes of 651 participants were eligible for the primary analysis. The characteristics of these eyes and participants are shown in Table 1. Mean age was 80.0 years (standard deviation [SD], 7.6), and 59.5% were female. The proportion of eyes with neovascular AMD (defined by positive reading center grading from SD-OCT or color fundus photography, or history of anti-vascular endothelial growth factor or laser therapy

for neovascular AMD) was 45.3%. The proportion of eyes with retinal fluid (intraretinal or subretinal) present, according to the ground truth of reading center grading, was 32.8%. The proportion of eyes whose SD-OCT imaging required senior adjudication at the reading center for fluid presence was 19.1% (corresponding to 27.0% of the eyes with retinal fluid and 15.2% of those without fluid). The numbers of eyes eligible for the secondary analyses were 1147 (evaluation of intraretinal fluid) and 1194 (evaluation of subretinal fluid). Of the 1147 eyes, the proportion with intraretinal fluid present (according to reading center grading) was 28.3%; of the 1194 eyes, the proportion with subretinal fluid present was 7.0%.

Primary Analysis: Performance of the Human Investigators and the Notal OCT Analyzer in Identifying Retinal Fluid

The performance of the human investigators and the NOA in identifying retinal fluid presence, relative to the ground truth, was quantitated using multiple performance metrics. The results are shown in Table 2. For the investigators, the accuracy (the primary performance metric) was 0.805 (95% confidence interval [CI], 0.780–0.828). The sensitivity was 0.468 (95% CI, 0.416–0.520), specificity was 0.970 (95% CI, 0.955–0.981), precision was 0.883 (95% CI, 0.829–0.924), and F1-score was 0.611. For the NOA, the accuracy was 0.851 (95% CI, 0.829–0.871). The sensitivity was 0.822 (95% CI, 0.779–0.859), specificity was 0.865 (95% CI, 0.839–0.889), precision was 0.749 (95% CI, 0.704–0.790), and F1-score was 0.784. The receiver operating characteristic curve for the NOA is shown in Figure 1; the area under the curve was 0.925. For comparison, the performance of the human investigators (pooled performance) on this same data set is shown as a single point.

The performance of the human investigators and the NOA were also evaluated on the Cirrus and Spectralis scans, considered separately. The results are shown in the Appendix (available at www.aaojournal.org). The accuracy of the NOA was numerically superior to the investigators for both the Cirrus scans (0.874 vs. 0.799) and Spectralis scans (0.840 vs. 0.807), that is, with some overlap of the 95% CIs between the investigators and the NOA, when each device was considered separately.

The results of sensitivity analyses are described in the Appendix (available at www.aaojournal.org). These include analyses in which the dataset was limited to the subset of 511 eyes with neovascular AMD (Appendix, available at www.aaojournal.org) and analyses in which the dataset was expanded to include the 116 eyes with a reading center grade of questionable for retinal fluid.

Secondary Analyses: Performance of the Human Investigators and Notal OCT Analyzer in Identifying Intraretinal Fluid and Subretinal Fluid

Likewise, the performance of the human investigators and the NOA in identifying the presence of (i) intraretinal fluid and, separately, (ii) subretinal fluid was quantitated using the same performance metrics. The results are shown in Table 3. Regarding intraretinal fluid, the accuracy of the investigators was 0.815 (95% CI, 0.792–0.837), sensitivity was 0.403 (95% CI, 0.349–0.459), specificity was 0.978 (95% CI, 0.966–0.987), precision was 0.879, and F1-score was 0.553. For the NOA, the accuracy was 0.877 (95% CI, 0.857–0.896), sensitivity was 0.763 (95% CI, 0.713–0.808), specificity was 0.922 (95% CI, 0.902–0.940),

precision was 0.795 (95% CI, 0.746–0.838), and F1-score was 0.779. Regarding subretinal fluid, the accuracy of the investigators was 0.946 (95% CI, 0.931–0.958), sensitivity was 0.583 (95% CI, 0.471–0.690), specificity was 0.973 (95% CI, 0.962–0.982), precision was 0.620 (95% CI, 0.504–0.727), and F1-score was 0.601. For the NOA, the accuracy was 0.863 (95% CI, 0.842–0.882), sensitivity was 0.940 (95% CI, 0.867–0.980), specificity was 0.857 (95% CI, 0.835–0.877), precision was 0.332 (95% CI, 0.272–0.396), and F1-score was 0.491.

Error Analyses

We performed exploratory error analysis for both the human investigators and the NOA. In particular, because low sensitivity was the main factor driving lower accuracy for the investigators, we considered all cases with retinal fluid (according to the ground truth) and divided these into true-positive and false-negative cases. For each group, we used the metrics described as proxies for case difficulty (i.e., proportion of cases requiring reading center senior adjudication for fluid presence, NOA-estimated quantity of fluid, and NOA-estimated number of B-scans with fluid).

The results are shown in Table 4. Of the 173 eyes with fluid correctly identified by the investigators, the proportion with intraretinal fluid only, proportion requiring adjudication, mean retinal fluid volume, and number of B-scans with fluid present were each significantly lower ($P < 0.001$, $P < 0.0001$, $P < 0.0001$, and $P < 0.0001$, respectively) than those for the 197 eyes with fluid missed by the investigators. Similar analyses were performed while restricting the cases to those correctly identified by the NOA as having retinal fluid present (Table 4). Again, of the 161 eyes with fluid correctly identified by the investigators, significant differences were observed for the same 4 characteristics, compared with those for the 143 eyes with fluid missed by the investigators. The results for the NOA are shown in Table 4. Of the 304 eyes with fluid correctly identified by the NOA, the proportion with intraretinal fluid only and proportion requiring adjudication were each significantly lower ($P < 0.001$ and $P = 0.001$, respectively) than those for the 66 eyes with fluid missed by the NOA.

Overall, the scans with retinal fluid missed by the investigators were more likely to have lower total fluid volumes and to contain intraretinal fluid only. Thus, as expected, the cases in which the investigators missed retinal fluid appeared more challenging (in terms of the metrics used) than those in which the fluid was correctly identified. Likewise, the cases in which the NOA missed retinal fluid were characterized by a high proportion requiring reading center adjudication and to contain intraretinal fluid only.

We also examined the false-positive cases for the NOA. For the 102 false-positive cases, the mean NOA-estimated fluid volume was 5.4 nl (SD, 22) and the mean number of B-scans with fluid present was 7.5 (SD, 11.3; median, 4). By contrast, for the 304 true positive cases, the mean NOA-estimated fluid volume was 110 nl (SD, 32; median, 12; $P = 0.001$) and the mean number of B-scans with fluid present was 24.8 (SD, 23.2; median, 17; $P < 0.001$). Thus, when the NOA falsely predicted that retinal fluid was present, it tended to make these predictions with very low estimated fluid volumes and appearing on a significantly lower number of B-scans.

Representative Examples of OCT Scans with the Accompanying Notal Report

In Figure 2, 3 representative examples of SD-OCT scans for which the NOA correctly identified the presence of retinal fluid (of varying volumes) are shown. In each case, also shown are the 2 Early Treatment Diabetic Retinopathy Study grid heatmaps automatically generated by the NOA, 1 for intraretinal fluid and 1 for subretinal fluid; these provide rapid visualization of the location and extent of fluid separately for each tissue compartment. The total volume of intraretinal and subretinal fluid is also displayed in nanoliters. In addition, a single representative B-scan is shown in each case, demonstrating how the NOA identifies and color-codes intraretinal and subretinal fluid on every B-scan in each cube.

Discussion

Main Results, Implications, Interpretation

In this study, we prospectively evaluated the performance of a large group of retinal specialists working in a clinical trial setting for the identification of retinal fluid on macular OCT scans from a large cohort of eyes with AMD. The OCT scans were from 2 commonly used devices. After this, we compared the performance of the retinal specialists with that of an AI-based algorithm, using reading center grades as the ground truth. The accuracy of the NOA was not only non-inferior but also superior numerically to that of the retinal specialists. The higher accuracy of the NOA was derived from substantially higher sensitivity (0.822 vs. 0.468), with only moderately lower specificity (0.865 vs. 0.970). Thus, the retinal specialists correctly identified retinal fluid in less than half of the cases. Regarding the cases with retinal fluid, those that were missed by the human investigators appeared to be more challenging cases, as might be expected. However, these cases of missed fluid might still be clinically important, particularly in eyes with intraretinal fluid and a relatively central location.

The lower specificity of the NOA in this cohort is likely to be acceptable because the purpose of the algorithm would be to assist the physician by acting as an additional diagnostic tool (rather than to replace the physician by autonomous grading). In this way, the physician would review the NOA report and could over-rule it as appropriate. Indeed, the en face heatmaps and B-scan color-coding allow the physician to examine areas of NOA-suspected fluid for potential agreement or disagreement. The substantially higher sensitivity of the NOA is an important advantage here, because the NOA heatmaps and color-codes would draw the attention of the physician to areas of fluid that might otherwise be missed. In addition, one of the NOA features that was validated previously²² is its ability to rank all B-scans by level of likelihood of fluid presence. This would allow efficient implementation of “NOA consulting” to the physician by means of a single view of the 3 B-scans with the highest probabilities of retinal fluid. In the hypothetical scenario of implementing the NOA in the setting of the current study, a substantial increase in the sensitivity of detecting retinal fluid of more than 80% (from 173 to 316 eyes) might be expected.

The accuracy of retinal specialists in real-world settings is likely to be lower than that measured in this clinical trial setting. In real-world practice, retinal specialists may not have time to perform detailed assessment of every individual B-scan across the whole macular

volume separately for both intraretinal fluid and subretinal fluid. Likewise, the accuracy of general ophthalmologists may be lower again. Thus, the use of those retinal specialists who participate in clinical trials was deliberately set as a high bar as comparison between the NOA and the highest standards of clinical practice. Indeed, worse visual outcomes have been consistently observed in real-world practice compared with those from clinical trial settings.^{12,25–29} It is possible that the relatively lower sensitivity of intraretinal fluid detection in real-world practice has contributed to this phenomenon, although this is difficult to assess with existing data. If this were true, a diagnostic tool that assists physicians in detecting retinal fluid with high sensitivity might help improve visual outcomes. However, evaluating this possibility would require its own dedicated studies. These would comprise prospective trials of traditional versus AI-assisted care in neovascular AMD, as has been conducted in other medical disciplines.^{30–32}

This test set was large and diverse; it contained more than 1000 eyes drawn from multiple sites across the United States, comprising a wide variety of participants, OCT devices, and operators. Of note, this represents a robust demonstration of external validity and generalizability, because the test set was drawn from a different population than the training and initial validation sets. In addition, the test set comprised a potentially challenging mixture of cases, reflected in the unusually high rates of reading center senior grader adjudication required. The participant mean age was high, and because some eyes had neovascular AMD at AREDS2 baseline and others developed neovascular disease soon after, the 10-year follow-on nature of the study meant that neovascular lesions were often mature, many with accompanying atrophy or fibrosis.

The tissue location of the retinal fluid is important. In neovascular AMD, the assessment of intraretinal fluid appears to be more important than that of subretinal fluid; in a randomized controlled trial, the visual outcome of eyes where re-treatment decisions were based on intraretinal fluid alone was similar to those where the decisions were based on both fluid locations.³³ Likewise, previous studies have demonstrated that intraretinal fluid is more highly predictive of poor visual outcomes than subretinal fluid.^{21,34} In that context, this study found that the accuracy of the NOA for intraretinal fluid detection was numerically superior to that of the retinal specialists (0.877 vs. 0.815); the sensitivity of the retinal specialists was particularly low at 0.403 versus 0.763 for the NOA.

The NOA may be useful as an additional diagnostic tool to assist physicians. Its advantages include rapid delivery of information on not just the presence or absence of retinal fluid but the compartment(s) involved. In addition, as demonstrated in Figure 2, the heatmaps demonstrate to the physician the location and extent of fluid separately for the 2 compartments; the presence of central involvement may be particularly important. Finally, the total estimated fluid volume is presented in nanoliters; this quantitative information might provide a useful metric for titrating treatments and follow-up intervals, as well as helping distinguish between exudative activity (with fluctuating volume) and degenerative cysts (with stable volume).

Comparison with Literature

In one previous study,³⁵ a deep learning approach was used to perform automated detection and quantification of retinal fluid on SD-OCT scans. Unlike the current analysis, this report was a retrospective study and did not represent external validation. In the previous report, only cases with clear consensus annotation between the graders were used for the study sample, whereas the current study sample had a high proportion of cases requiring senior grader adjudication because of disagreement over fluid presence. Thus, direct comparison is difficult. However, the performance of the NOA appeared only marginally inferior to that of the deep learning approach reported in the previous study.³⁵ The mean NOA sensitivity on all 4 categories (Cirrus/Spectralis; intraretinal/subretinal fluid) was 85.5% compared with 86.4% in the previous study; the mean NOA specificity was 89.3% compared with 93.8% in the previous study.

Another previous study³⁶ applied deep learning approaches to macular OCT datasets. However, the main aim of this study was to perform automated referral recommendations. Thus, although the algorithm performed automated segmentation of the OCT data into multiple volumes as an intermediate step, the performance of the algorithm was tested on the final referral recommendation and the suggested diagnosis, not by comparison of predicted versus genuine presence of retinal fluid.

Although multiple studies have examined various AI-based approaches to the detection of retinal fluid,^{35,37-46} we are not aware of any prospectively undertaken studies. Most previous reports have been small studies with relatively low case numbers in the test set.³⁷⁻⁴⁴ Except for one study,⁴⁵ all previous studies used training and testing OCT data from a single center. Likewise, almost no previous studies have performed generalizability testing using external datasets.⁴⁵ With only 3 exceptions,^{35,36,45} all previous studies used a single OCT device only. These considerations are important, because performance might decrease substantially when algorithms developed using 1 OCT device on a particular patient population are applied to different patient populations imaged with different devices. Thus, studies that perform robust prospective external validation tests, as in this report, are important. In addition, with only 2 exceptions,^{35,37} most previous studies have used approaches that pertain only to intraretinal fluid or do not distinguish between intraretinal and subretinal fluid.

Strengths and Limitations

The strengths of this study include its prospective, preplanned nature, large size, use of masked reading center grading for the ground truth, and ability to compare the performance of the NOA with that of a large group of retinal specialists. Because the test data were drawn from a diverse population at multiple centers across the United States (comprising OCT scans from 2 commonly used devices) and from a different population than the training data, this study represents a demonstration of robust generalizability and external validation. Limitations include the exclusion of a small minority of OCT scans by the NOA. However, we consider that it is preferable for the NOA to have a quality filter like this to avoid forcing it to provide an assessment when a reliable one may not be possible. In addition, validating the quantitative estimates of fluid volume provided by the NOA would require comparison

with ground truth fluid volumes from human expert grading; this was not conducted in this study but has been robustly validated in separate analyses, with correlation coefficients of 0.95 (Spectralis) and 0.92 (Cirrus) for agreement between NOA retinal fluid volumes and those from human expert grading.²³ Finally, the AREDS2-10Y was not representative of real-world practice. Thus, despite the prospective nature of the study and diversity of the study population, NOA performance in real-world practice remains unclear.

In conclusion, in this large and challenging sample of SD-OCT macular volume scans obtained with 2 commonly used devices, retinal specialists had imperfect accuracy in detecting retinal fluid, with low sensitivity. This was particularly true for (i) intraretinal fluid and (ii) challenging cases (with low fluid volume and fluid appearing on fewer B-scans). Artificial intelligence–based detection achieved a higher level of accuracy. This AI software tool could assist physicians in detecting retinal fluid, which is important for diagnostic, re-treatment, and prognostic tasks in neovascular AMD.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support to the study was provided by Notal Vision Ltd., through a service agreement with the EMMES Company, LLC. The AREDS2 study was supported by intramural program funds and contracts from the National Eye Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland (Contract Number HHSN263201300005C). The funding organizations participated in the design of the study, interpretation of the data, and preparation of the manuscript.

Abbreviations and Acronyms:

AI	artificial intelligence
AMD	age-related macular degeneration
AREDS2-10Y	Age-Related Eye Disease Study 2 10-year Follow-On Study
CI	confidence interval
ILM	internal limiting membrane
NOA	Notal OCT Analyzer
RPE	retinal pigment epithelium
SD	standard deviation
SD-OCT	spectral domain OCT

References

1. Quartilho A, Simkiss P, Zekite A, et al. Leading causes of certifiable visual loss in England and Wales during the year ending 31 March 2013. *Eye (Lond)*. 2016;30:602–607. [PubMed: 26821759]

2. Congdon N, O'Colmain B, Klaver CC, et al. Causes and prevalence of visual impairment among adults in the United States. *Arch Ophthalmol*. 2004;122:477–485. [PubMed: 15078664]
3. Flaxel CJ, Adelman RA, Bailey ST, et al. Age-Related Macular Degeneration Preferred Practice Pattern®. *Ophthalmology*. 2020;127:P1–P65. [PubMed: 31757502]
4. Group CR, Martin DF, Maguire MG, et al. Ranibizumab and bevacizumab for neovascular age-related macular degeneration. *N Engl J Med*. 2011;364:1897–1908. [PubMed: 21526923]
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. [PubMed: 26102433]
6. Schmidt-Erfurth U, Klmscha S, Waldstein SM, Bogunovic H. A view of the current and future role of optical coherence tomography in the management of age-related macular degeneration. *Eye (Lond)* 2017;31:26–44. [PubMed: 27886184]
7. Gale RP, Mahmood S, Devonport H, et al. Action on neovascular age-related macular degeneration (nAMD): recommendations for management and service provision in the UK hospital eye service. *Eye (Lond)* 2019;33(Suppl 1):1–21.
8. 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. [PubMed: 22939114]
9. Wong WL, Su X, Li X, et al. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *Lancet Glob Health*. 2014;2:e106–116. [PubMed: 25104651]
10. Colijn JM, Buitendijk GHS, Prokofyeva E, et al. Prevalence of age-related macular degeneration in Europe: the past and the future. *Ophthalmology*. 2017;124:1753–1763. [PubMed: 28712657]
11. Singer MA, Awh CC, Sadda S, et al. HORIZON: an open-label extension trial of ranibizumab for choroidal neovascularization secondary to age-related macular degeneration. *Ophthalmology*. 2012;119:1175–1183. [PubMed: 22306121]
12. Keenan TD, Vitale S, Agron E, et al. Visual acuity outcomes after anti-vascular endothelial growth factor treatment for neovascular age-related macular degeneration: Age-Related Eye Disease Study 2 Report Number 19. *Ophthalmol Retina*. 2020;4:3–12. [PubMed: 31395505]
13. Maloca P, Hasler PW, Barthelmes D, et al. Safety and feasibility of a novel sparse optical coherence tomography device for patient-delivered retina home monitoring. *Transl Vis Sci Technol*. 2018;7:8.
14. Fujimoto J, Swanson E. The development, commercialization, and impact of optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2016;57:OCT1–OCT13. [PubMed: 27409459]
15. Zeiss. The New Cirrus family, https://www.zeiss.com/meditec/us/_masterpages/products/ophthalmology-optometry/cirrus-family.html; 2019. Accessed November 27, 2019.
16. Zeiss. Zeiss Cirrus 6000. 2019. Available at: <https://www.zeiss.com/meditec/us/products/ophthalmology-optometry/retina/diagnostics/optical-coherence-tomography/oct-optical-coherence-tomography/cirrus-6000-performance-oct.html>. Accessed 11/27/2019.
17. Heidelberg Engineering. Spectralis. Multimodal imaging platform optimized for the posterior segment. 2019, Available at: <https://business-lounge.heidelbergengineering.com/us/en/products/spectralis/>. Accessed 11/27/2019.
18. Heidelberg Engineering. Heidelberg Eye Explorer. Next-generation platform for Heidelberg engineering device management and integration. 2019. Available at: <https://business-lounge.heidelbergengineering.com/uss/en/products/heyex/>. Accessed 11/27/2019.
19. 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, [PubMed: 26307399]
20. 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. [PubMed: 31047298]
21. Waldstein SM, Simader C, Staurengi G, et al. Morphology and visual acuity in aflibercept and ranibizumab therapy for neovascular age-related macular degeneration in the VIEW Trials. *Ophthalmology*. 2016;123:1521–1529. [PubMed: 27157149]

22. Chakravarthy U, Goldenberg D, Young G, et al. Automated identification of lesion activity in neovascular age-related macular degeneration. *Ophthalmology*. 2016;123: 1731–1736. [PubMed: 27206840]
23. Goldstein HM, Rafaeli O, Loewenstein A. A novel AI-based algorithm for quantifying volumes of retinal pathologies in OCT scans. San Francisco, California: American Academy of Ophthalmology Meeting; 2019. October 12, 2019.
24. AREDS2 Research Group, Chew EY, Clemons T, et al. The Age-Related Eye Disease Study 2 (AREDS2): study design and baseline characteristics (AREDS2 report number 1). *Ophthalmology*. 2012;119:2282–2289. [PubMed: 22840421]
25. Rosenfeld PJ, Brown DM, Heier JS, et al. Ranibizumab for neovascular age-related macular degeneration. *N Engl J Med*. 2006;355:1419–1431. [PubMed: 17021318]
26. Brown DM, Michels M, Kaiser PK, et al. Ranibizumab versus verteporfin photodynamic therapy for neovascular age-related macular degeneration: two-year results of the ANCHOR study. *Ophthalmology*. 2009;116:57–65 e55. [PubMed: 19118696]
27. Comparison of Age-related Macular Degeneration Treatments Trials Research Group, Martin DF, Maguire MG, et al. Ranibizumab and bevacizumab for treatment of neovascular age-related macular degeneration: two-year results. *Ophthalmology*. 2012;119:1388–1398. [PubMed: 22555112]
28. Writing Committee for the UK Age-Related Macular Degeneration EMR Users Group. The neovascular age-related macular degeneration database: multicenter study of 92 976 ranibizumab injections: report 1: visual acuity. *Ophthalmology*. 2014;121:1092–1101. [PubMed: 24461586]
29. 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. [PubMed: 29305324]
30. Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ*. 2020;368: m689. [PubMed: 32213531]
31. Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut*. 2019;68:1813–1819. [PubMed: 30814121]
32. NIH U.S. National Library of Medicine [ClinicalTrials.gov](https://clinicaltrials.gov). Breast ultrasound image reviewed with assistance of deep learning algorithms. <https://clinicaltrials.gov/ct2/show/NCT03706534>. Accessed May 18, 2020.
33. Arnold JJ, Markey CM, Kurstjens NP, Guymer RH. The role of sub-retinal fluid in determining treatment outcomes in patients with neovascular age-related macular degeneration—a phase IV randomised clinical trial with ranibizumab: the FLUID study. *BMC Ophthalmol*. 2016;16:31. [PubMed: 27009515]
34. 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. [PubMed: 26783095]
35. 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. [PubMed: 29224926]
36. 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. [PubMed: 30104768]
37. Lee H, Kang KE, Chung H, Kim HC. Automated segmentation of lesions including subretinal hyperreflective material in neovascular age-related macular degeneration. *Am J Ophthalmol*. 2018;191:64–75. [PubMed: 29655643]
38. Lee CS, Tying AJ, Deruyter NP, et al. Deep-learning based, automated segmentation of macular edema in optical coherence tomography. *Biomed Opt Express*. 2017;8:3440–3448. [PubMed: 28717579]
39. Wang J, Zhang M, Pechauer AD, et al. Automated volumetric segmentation of retinal fluid on optical coherence tomography. *Biomed Opt Express*. 2016;7:1577–1589. [PubMed: 27446676]

40. Xiayu X, Kyungmoo L, Li Z, et al. Stratified sampling voxel classification for segmentation of intraretinal and subretinal fluid in longitudinal clinical OCT data. *IEEE Trans Med Imaging*. 2015;34:1616–1623. [PubMed: 25769146]
41. Chiu SJ, Allingham MJ, Mettu PS, et al. Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. *Biomed Opt Express*. 2015;6: 1172–1194. [PubMed: 25909003]
42. Chen X, Zhang L, Sohn EH, et al. Quantification of external limiting membrane disruption caused by diabetic macular edema from SD-OCT. *Invest Ophthalmol Vis Sci*. 2012;53: 8042–8048. [PubMed: 23111607]
43. Wilkins GR, Houghton OM, Oldenburg AL. Automated segmentation of intraretinal cystoid fluid in optical coherence tomography. *IEEE Trans Biomed Eng*. 2012;59: 1109–1114. [PubMed: 22271827]
44. Fernandez DC. Delineating fluid-filled region boundaries in optical coherence tomography images of the retina. *IEEE Trans Med Imaging*. 2005;24:929–945. [PubMed: 16092326]
45. Venhuizen FG, van Ginneken B, Liefers B, et al. Deep learning approach for the detection and quantification of intraretinal cystoid fluid in multivendor optical coherence tomography. *Biomed Opt Express*. 2018;9:1545–1569. [PubMed: 29675301]
46. Motozawa N, An G, Takagi S, et al. Optical coherence tomography-based deep-learning models for classifying normal and age-related macular degeneration and exudative and non-exudative age-related macular degeneration changes. *Ophthalmol Ther*. 2019;8:527–539. [PubMed: 31407214]

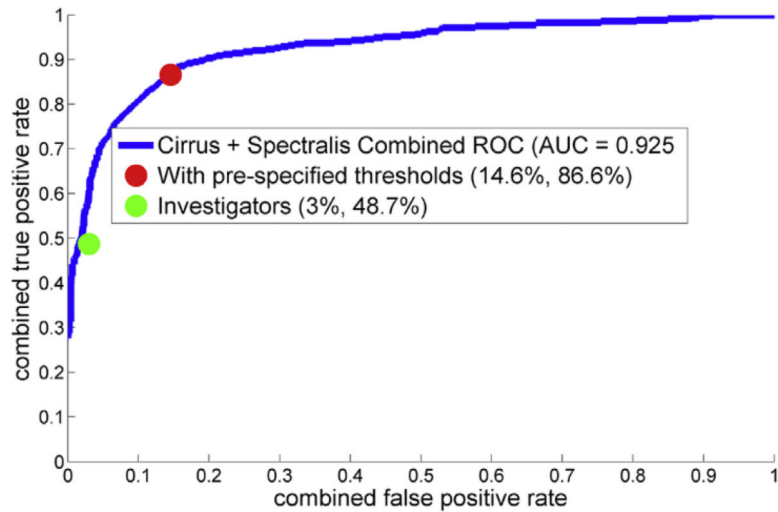


Figure 1.

Receiver operating characteristic curve for the Notal OCT Analyzer (NOA): spectral domain OCT (SD-OCT) macular cubes obtained from both the Heidelberg Spectralis (Heidelberg Engineering, Heidelberg, Germany) and Cirrus Zeiss devices (Carl Zeiss Meditech, Dublin, CA). The area under the curve was 0.925. The receiver operating characteristic curve was based on 1051 of the 1127 eyes eligible for the primary analysis. The other 76 eyes were those classified by the NOA as having artefactual retinal fluid actually caused by vitreo-macular interface abnormalities (thus given a negative grade for retinal fluid by the NOA). Because the NOA first allocates a high probability of retinal fluid to these scans but subsequently and automatically grades them as negative, they are excluded from the receiver operating characteristic analysis (but included in all other performance metrics). AUC = area under the curve; ROC = receiver operating characteristic.

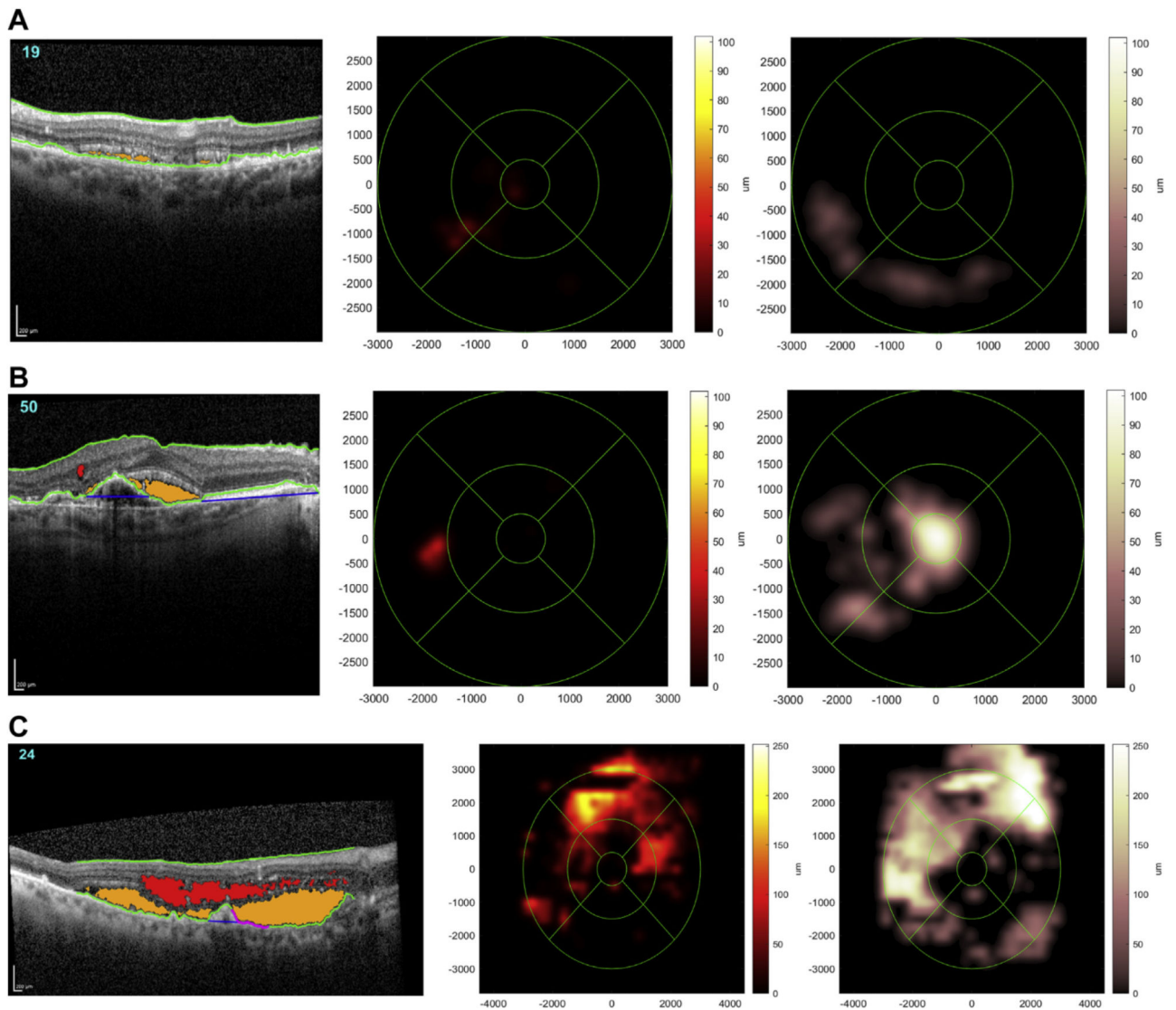


Figure 2. Representative examples of spectral domain OCT (SD-OCT) scans where the Notal OCT Analyzer (NOA) correctly identified the presence of retinal fluid. A representative B-scan is shown (**left**) in each case, demonstrating how the NOA identifies and color-codes intraretinal fluid (**red**) and subretinal fluid (**orange**) on every B-scan. Also shown for each case are the 2 Early Treatment Diabetic Retinopathy Study grid heatmaps automatically generated by the NOA, 1 for intraretinal fluid (**middle**) and 1 for subretinal fluid (**right**); these provide rapid visualization of the location and extent of fluid separately for each tissue compartment. The total volumes of intraretinal fluid and subretinal fluid estimated by the NOA are displayed in nanoliters. **A**, Low total fluid volume (22 nl). **B**, Moderate total fluid volume (136 nl). **C**, High total fluid volume (3374 nl).

Table 1.

Characteristics of the 1127 Study Eyes Eligible for the Primary Analysis

	Proportion (%)
Age (yrs), mean (SD)	80.0 (7.6)
Female sex	59.5
White race	97.4
Education	
High school or less	23.2
At least some college	49.3
Postgraduate	24.0
Smoking	
Never	48.9
Former	45.6
Current	5.5
Best-corrected visual acuity (ETDRS letters), mean (SD)	64.5 (24.4)
Neovascular AMD	45.3
Senior reading center grader adjudication for fluid presence	19.1
Retinal fluid present on macular SD-OCT (reading center grading)	32.8

AMD = age-related macular degeneration; SD = standard deviation; SD-OCT = spectral domain OCT.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Retinal Specialist versus Artificial Intelligence-Based Detection of Retinal Fluid (Intraretinal or Subretinal)

	Investigators		Notal OCT Analyzer	
	<i>Estimate</i>	<i>95% CI</i>	<i>Estimate</i>	<i>95% CI</i>
Accuracy	0.805	0.780–0.828	0.851	0.829–0.871
Sensitivity	0.468	0.416–0.520	0.822	0.779–0.859
Specificity	0.970	0.955–0.981	0.865	0.839–0.889
Precision	0.883	0.829–0.924	0.749	0.704–0.790
F1-score	0.611	–	0.784	–

CI = confidence interval.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Secondary Analyses: Retinal Specialist versus Artificial Intelligence–Based Detection of Intraretinal Fluid and Subretinal Fluid

	Investigators		Notal OCT Analyzer	
	<i>Estimate</i>	<i>95% CI</i>	<i>Estimate</i>	<i>95% CI</i>
Detection of Intraretinal Fluid				
Accuracy	0.815	0.792–0.837	0.877	0.857–0.896
Sensitivity	0.403	0.349–0.459	0.763	0.713–0.808
Specificity	0.978	0.966–0.987	0.922	0.902–0.940
Precision	0.879	0.816–0.927	0.795	0.746–0.838
F1-score	0.553	–	0.779	–
Detection of Subretinal Fluid				
Accuracy	0.946	0.931–0.958	0.863	0.842–0.882
Sensitivity	0.583	0.471–0.690	0.940	0.867–0.980
Specificity	0.973	0.962–0.982	0.857	0.835–0.877
Precision	0.620	0.504–0.727	0.332	0.272–0.396
F1-score	0.601	–	0.491	–

CI = confidence interval.

Table 4.

Error Analyses

	Notal OCT Analyzer (All Cases)			Investigators (All Cases)			Investigators (Only Considering Cases Correctly Identified by the NOA as Having Retinal Fluid)		
	True-positives	False-negatives	P	True-positives	False-negatives	P	True-positives	False-negatives	P
No.	304	66		173	197		161	143	
Age (yrs), mean (SD)	80.6 (7.4)	81.7 (6.9)	0.29	80.6 (7.3)	80.7 (7.4)	0.87	80.5 (7.4)	80.6 (7.4)	0.88
Female sex, %	57.4	58.7	0.85	63.8	54.7	0.10	61.4	54.2	0.23
Best-corrected visual acuity (ETDRS letters), mean (SD)	53.5 (27.9)	59.7 (22.0)	0.05	53.3 (28.9)	55.8 (25.3)	0.38	52.4 (29.0)	54.8 (26.6)	0.46
Cirrus (Carl Zeiss Meditec, Dublin, CA) (vs. Spectralis [Heidelberg Engineering, Heidelberg, Germany]) scans, %	32.2	27.2	0.47	30.1	32.5	0.65	31.1	33.6	0.71
Intraretinal fluid present only (%) [*]	67.8	90.9	<0.001	59.5	82.7	<0.001	57.8	79.0	<0.001
Subretinal fluid present only (%)	12.5	4.5	0.08	14.5	8.1	0.07	14.3	10.5	0.39
Both intraretinal and subretinal fluid present (%)	12.5	1.5	0.01	17.9	4.1	<0.001	19.3	4.9	<0.001
Reading center senior grader adjudication for fluid presence (%)	23.4	43.9	0.001	15.0	37.6	<0.0001	14.3	33.6	<0.0001
Total NOA-estimated fluid volume (nl): mean (SD)	110 (320)	–	–	156 (387)	33 (154)	<0.0001	168 (398)	46 (179)	<0.001
NOA-estimated number of B-scans with fluid: mean (SD); %	24.8 (23.2); 25.4	–	–	30.4 (25.3); 31.9	11.7 (16.5); 11.3	<0.0001	32.7 (24.8); 34.2	15.9 (17.5); 15.5	<0.001

NOA = Notal OCT Analyzer; SD = standard deviation.

^{*} The percentages do not necessarily sum to 100% because some cases had reading center grades of definite fluid for one fluid type but ungradable or questionable for the other fluid type, so are not included in these categories.