

Repeatability of a Virtual Reality Headset Perimeter in Glaucoma and Ocular Hypertensive Patients

Rafaella Nascimento e Silva¹, Julia A. Kim¹, Yihao Li¹, Chen Chen¹, Anisa F. Chaudhry², Andrew R. Berneshawi², Miao Zhang¹, Adam Villarreal¹, Jocelyn Liu², Ann Shue², Dolly S. Chang^{1,2}, and Robert T. Chang²

¹ Genentech, Inc., South San Francisco, CA, USA

² Byers Eye Institute, Stanford University School of Medicine, Palo Alto, CA, USA

Correspondence: Robert T. Chang, Byers Eye Institute, Stanford University, 2452 Watson Court, Palo Alto, CA 94303, USA. e-mail: viroptic@gmail.com.

Received: September 10, 2023

Accepted: April 15, 2024

Published: June 20, 2024

Keywords: perimetry; virtual reality; glaucoma; ocular hypertension

Citation: Nascimento e Silva R, Kim JA, Li Y, Chen C, Chaudhry AF, Berneshawi AR, Zhang M, Villarreal A, Liu J, Shue A, Chang DS, Chang RT. Repeatability of a virtual reality headset perimeter in glaucoma and ocular hypertensive patients. *Transl Vis Sci Technol.* 2024;13(6):14. <https://doi.org/10.1167/tvst.13.6.14>

Purpose: The VisuALL S is an automated, static threshold, virtual reality–based perimeter for mobile evaluation of the visual field. We examined same-day and 3-month repeatability.

Methods: Adult participants with a diagnosis of glaucoma or ocular hypertension underwent two VisuALL 24-2 Normal T- Full threshold strategy tests at baseline and one additional exam at 3 months for each eligible eye. Spearman, intraclass correlation coefficients (ICCs), and Bland–Altman plots were used to assess the correlation of individual point sensitivities and mean deviation (MD) among three tests.

Results: Eighty-eight eyes (44 participants) were included. Average age was 68.1 ± 14.3 years, and 60.7% were male. VisuALL MD was highly correlated between tests (intravisit: $r = 0.89$, intervisit: $r = 0.82$; $P < 0.001$ for both). Bland–Altman analysis showed an average difference in intravisit MD of -0.67 dB (95% confidence interval [CI], -6.04 to 4.71 dB) and -0.15 dB (95% CI, -8.04 to 7.73 dB) for intervisit exams. Eight-five percent of pointwise intravisit ICCs were above 0.75 (range, 0.63 to 0.93), and 65% of pointwise intervisit ICCs were above 0.75 (range, 0.55 to 0.91).

Conclusions: VisuALL demonstrated high correlation of MD between tests and good repeatability for individual point sensitivities among three tests in 3 months, except at the points around the blind spot and superiorly.

Translational Relevance: The preliminary reproducibility results for VisuALL are encouraging. Its portable design makes it a potentially useful tool for patients with glaucoma, enabling more frequent assessments both at home and in clinical settings.

Introduction

Standard automated perimetry (SAP) has long remained a cornerstone in the diagnosis and management of vision-threatening diseases such as glaucoma. Conventional SAP, including the Humphrey Field Analyzer (HFA; Carl Zeiss Meditec, Dublin, CA, USA) and Octopus perimeter (Haag-Streit AG, Koeniz, Switzerland), is commonly used to test visual function by presenting static white light stimuli of constant size but adjustable brightness in various locations throughout the visual field (VF). By determining the minimum threshold necessary for a patient

to detect presented stimuli, SAP can identify the loss of sensitivity to light, or contrast sensitivity.¹ Despite the capabilities of SAP, several limitations of these perimeters are evident. For one, the HFA and Octopus are office-based devices that require skilled technicians to improve reliability, which limits the frequency of testing; patients must travel to the doctor's office on multiple occasions to undergo repeat examinations, which is inefficient and sometimes impractical. Moreover, the test itself is known to be uncomfortable² due to positioning constraints and with one eye patched at a time, which may compromise the reliability of results and make it difficult to determine when true VF progression has occurred. In response,

novel headset-based perimeters that allow more neck freedom and bilateral testing have recently been developed to overcome some of these conventional SAP limitations. Another key differentiator of virtual reality (VR) perimetry is the ability to perform the test at home, multiple times, which increases the chance of detecting change over time earlier for a repeatable test.

The VisuALL S Analyzer (Olleyes, Inc., Summit, NJ, USA) is a US Food and Drug Administration–registered class I device VR headset with software simulating perimetry testing. Unlike the HFA and Octopus VF tests that require patients to keep their heads immobilized in a perimeter bowl, the VisuALL test utilizes a small, lightweight, head-mounted VR device without the need for a dark room. Since patients can be tested with both eyes at the same time and move their heads during the exam, VisuALL’s design is especially attractive for patients with musculoskeletal problems or claustrophobia who may find conventional SAP to be stressful and difficult. A small three-subject pilot study evaluating VisuALL versus Humphrey visual field (HVF) 24-2 reported a small incidence of fixation losses, with losses detected in only 2 of 24 total VF tests.³ In terms of how this new perimeter compares to conventional SAP, previous studies have demonstrated that global VF measurements obtained with VisuALL are significantly correlated and have overall good agreement with measurements obtained from the HFA. In addition, VisuALL mean sensitivity has been shown to have high diagnostic performance in discriminating normal versus glaucomatous eyes with a comparable receiver operating characteristic curve (0.98 vs. 0.93, $P = 0.06$) to HFA mean sensitivity.⁴ When comparing user experience, Groth et al.⁵ evaluated VisuALL-K, a pediatric protocol of the VR perimeter, in a normal cohort ($N = 50$) of children aged 8 to 17 years and reported higher patient satisfaction compared to HFA.

Although VisuALL’s advantageous design optimized for portability, comfort, and improved user experience has enabled it to emerge as a promising alternative for VF assessment, its repeatability has not yet, to our knowledge, been evaluated in patients with glaucoma. In this study, we assessed the same-day and 3-month repeatability both globally and on a pointwise level of the VisuALL S Analyzer using the VisuALL S 24-2 protocol in a cohort of adult patients diagnosed with glaucoma or ocular hypertension. Determining these measures is crucial to increasing clinician confidence in VisuALL variability since each patient result may serve as their own baseline if more frequent, at-home VF monitoring is performed to detect glaucoma progression earlier.

Materials and Methods

This observational cohort study was approved by the Institutional Review Board at Stanford University and conducted from 2021 to 2022. The study protocol adhered to all tenets of the Declaration of Helsinki. Patients with a diagnosis of glaucoma or ocular hypertension who had recently performed, or who were scheduled to perform, a Humphrey VF 24-2 exam at Byers Eye Institute during regular clinic visits were enrolled prospectively. All study participants had prior experience taking Humphrey VF exams at Byers Eye Institute and had the diagnosis defined by trained glaucoma specialists (AS and RTC) based on all clinical data in the chart notes, including fundus photos, optical coherence tomography, HVF, intraocular pressure, and risk factors.

Male and female patients ≥ 22 years of age were eligible to participate in the study. A diagnosis of glaucoma was defined by clinical evidence of progressive retinal ganglion cell dysfunction and degeneration using at least one VF test and at least one structural modality, such as spectral-domain optical coherence tomography imaging and/or fundus photography. Ocular hypertension was defined as intraocular pressure of >21 mm Hg with normal structural and functional testing. If two eyes of a patient met the eligibility criteria, both eyes were enrolled. All participants signed an informed consent; the document was read aloud to participants in its entirety if they were unable to read due to impaired vision.

Patients with a clinical history of an ocular disease, disorder, or other condition that would likely interfere with the interpretation of the study results, such as a corneal transplant, macular disease, or nonglaucomatous optic neuropathy, were excluded from the study. Additional exclusion criteria included best-corrected visual acuity (BCVA) of 20/200 or worse in either eye due to glaucoma, intraocular surgery in the study eye within 12 weeks prior to the first study visit, and the inability to complete VF exams for any reason, such as being unable to tolerate the VR headset due to a physical or mental condition.

Previous publications have described VisuALL in detail, but its features and specifications are summarized here for reference.^{3,4} The VisuALL testing system consists of three main components: a head-mounted VR device; a Bluetooth-connected, handheld response button; and a web-capable device, such as a laptop, phone, or tablet to access the website where tests are administered and managed. In terms of software, the testing system uses the Olleyes cloud-based server to store the VF data, the VisuALL web application, and

the Unity algorithms, which are all Health Insurance Portability and Accountability Act compliant.

The VR headset is lightweight (approximately 300 g) and features two organic, light-emitted diode screens (one for each eye) of 1920×2160 pixels. When worn, the resulting display measures 125.40×70.56 mm and sits at a distance of 60.50 mm from the eyes that subtends a field of view up to 100 degrees.³ Due to the use of separate screens, occlusion of the eye is not required during testing; both eyes are kept open during the VF exam and are randomly tested within the same session.

VisuALL's 24-2 protocol (Normal T-Full threshold strategy) was used in this study with white Goldmann size III test stimuli shown against a scotopic background (1 cd/m^2) and tested both eyes simultaneously.^{3,5} In the 24-2 protocol, 54 points are tested in the central 24 degrees of the VF, with each test location 6 degrees apart. Similar to the Swedish Interactive Thresholding Algorithm Standard strategy that was developed for the HFA, Normal T employs a double-crossover method to establish thresholds for an anchor point in each quadrant of the VF. A proprietary testing strategy is then used to determine the threshold values of predetermined adjacent locations. In the resulting VF report, threshold values are reported to the clinician and patient in a range of 0 to 49 decibels (dB).³

This study took place over two clinic visits: at the first visit, participants underwent two VF test sessions using the VisuALL T-24 protocol to investigate the perimeter's intravisit repeatability. Three months (± 1 week) after the baseline visit, participants returned to the clinic to perform one additional VisuALL T-24 protocol VF test to investigate intervisit repeatability. Participants were instructed to remain seated upright and to wear their usual close-distance prescription eyeglasses while taking the VF exam; those who did not bring their own eyeglasses but needed refraction correction were provided with a trial lens to use for the duration of the test. With the assistance of study staff, participants put on the VR headset and adjusted its straps accordingly, ensuring that the device was fitted securely. Participants were then given the response button to hold in their dominant hand and instructed to begin the test when they were ready. Immediately prior to the VF exam, a brief tutorial with audio was displayed within the VR headset that instructed participants to fixate at the red target in the center of the screen and to depress the "trigger" button on the handpiece each time that they detected a flash of light.

Summary statistics for numerical values were reported as mean and standard deviation or median (interquartile range [IQR]), after performing the

Shapiro–Wilk test; it was determined that the VF data were not normally distributed. Kruskal–Wallis rank-sum and Pearson χ^2 tests were used to compare continuous and categorical variables, respectively, between VF test sessions. For intravisit repeatability, the two exams taken at the baseline visit were compared; for intervisit analysis, the first VF exam at baseline was compared with the repeat VF exam taken within 3 months. The first exam was selected to be used as a baseline, as there is the potential reported fatigue effect.⁶ Intraclass correlation coefficients (ICCs) with "single-rater" type and differences in threshold sensitivities between intravisit and intervisit exams were calculated to measure the absolute agreement at each test location in the VF and displayed in a color-coded, graphical format in the pointwise analysis. In this analysis, the left eyes were inverted to the right-eye format. To account for the physiologic blind spot, two points were excluded in the corresponding area of the VF. In addition, Bland–Altman plots using 95% limits of agreement (mean difference ± 1.96 standard deviation) and Spearman correlation test and a plot with line of equality ($x = y$) were created to assess the agreement of mean deviation (MD) values between intravisit and intervisit tests. Finally, a similar method as described in Artes et al.⁷ was used to examine how intervisit reproducibility varied according to the sensitivity of different locations in the VF⁸: mean absolute difference between the first exam and the VF exam taken at 3 months was plotted against the average baseline sensitivity in 5-dB intervals. All statistical analyses were performed using R software version 4.1.3 (R Foundation for Statistical Computing, Vienna, Austria).

Results

In total, 112 eyes of 56 patients were enrolled in the study across all glaucoma severity or ocular hypertensive diagnosis. Ninety-four eyes of 47 participants completed all three VisuALL VF exams. The patients who discontinued the study were not able to complete all the exams. All eyes' data met the reliability criteria (false-positive and false-negative rates $\leq 33\%$), but six eyes of three participants were excluded due to technical reasons; upon individual review of the VF data, these participants were found to have total scotomas in their third exams. The remaining cohort, consisting of 44 participants (88 eyes), was 60.7% male and had a mean age of 68.1 ± 14.3 years at baseline. In addition, the mean test duration of the first VF exam was 4.9 minutes per eye, and participants' median (IQR)

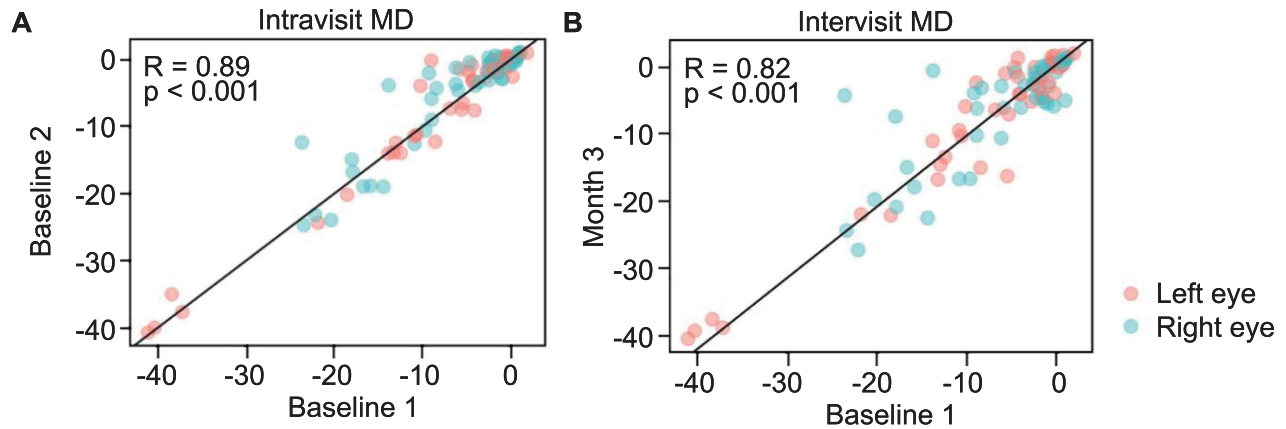


Figure 1. The association of MD values obtained with VisuALL between intravisit (A) and intervisit (B) exams assessed with Spearman correlation test and a line of equality ($x = y$).

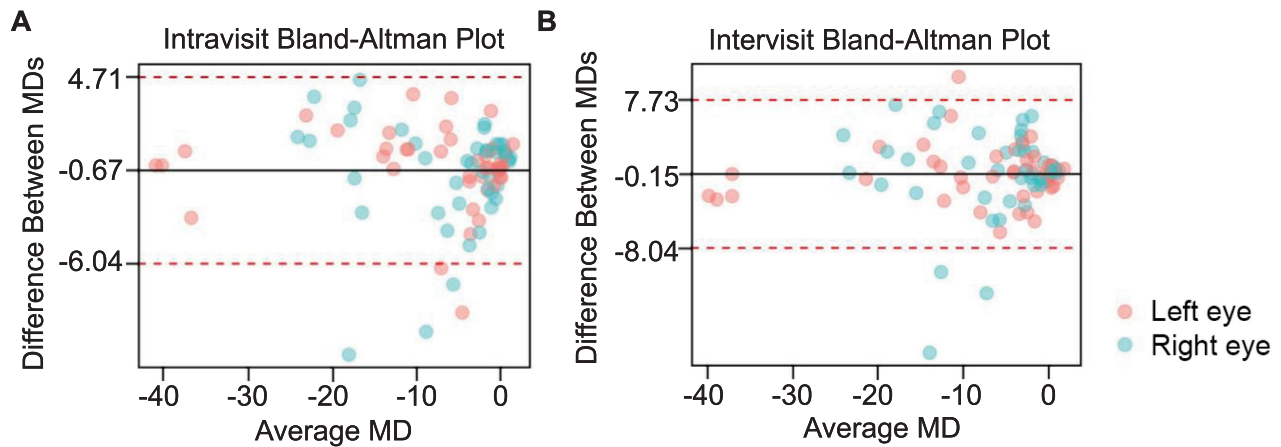


Figure 2. The repeatability of MD measurements obtained with VisuALL for intravisit (A) and intervisit (B) exams. *Solid line* denotes the average difference of MD between tests (bias); *dotted lines* denote the 95% limits of agreement.

MD and pattern standard deviation (PSD) values were -4.04 ($-10.70, -0.86$) dB and 5.46 ($3.07, 12.71$) dB, respectively. There was no statistically significant difference in mean test duration ($P = 0.271$), MD ($P = 0.893$), or PSD ($P = 0.997$) across the three VisuALL tests.

Mean deviation values demonstrated high correlations between both intravisit and intervisit VF exams: the intravisit correlation of MD was 0.89 , and the intervisit correlation of MD was 0.82 ($P < 0.001$ for both; Fig. 1). Bland–Altman analysis further revealed an average difference of MD between intravisit tests of -0.67 dB, with five eyes that fell outside the limits of agreement (95% confidence interval [CI], -6.04 to 4.71 dB; Fig. 2). For intervisit exams, the average difference of MD was found to be -0.15 dB, and four eyes

fell outside the limits of agreement (95% CI, -8.04 to 7.73 dB; Fig. 2).

Point-by-point analysis of threshold sensitivities also demonstrated strong correlations, with 85% of points achieving good ($0.75 \leq ICC < 0.90$) to excellent ($ICC \geq 0.90$) ICC values⁹ in the intravisit analysis: across all the VFs, ICCs ranged from 0.63 to 0.93 (Fig. 3). Out of the 52 reported test locations, 19 points (36.5%) had excellent ICCs, 25 points (48%) had good ICCs, and 8 points (15.4%) had moderate ICCs ($0.5 \leq ICC < 0.75$). In addition, the mean difference in threshold sensitivities ranged from -2.10 to 0.32 dB, and 47 out of 52 points (90.4%) showed mean differences that were negative, with sensitivities values lower in the baseline test 1 compared with baseline test 2.

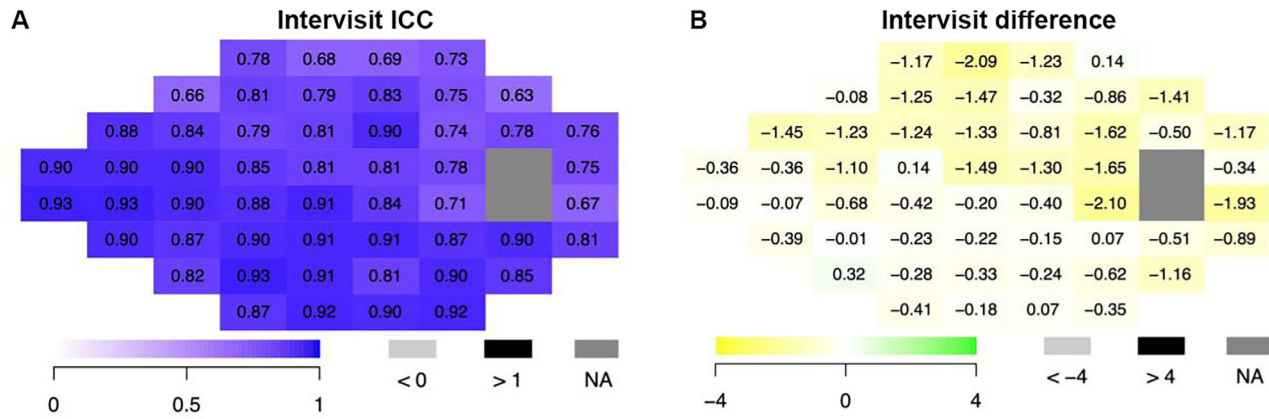


Figure 3. Point-by-point analysis of threshold sensitivities at each VF test location comparing intravisit VisuALL exams ($N = 88$ eyes) by ICC (A) and difference (B). Left eyes were inverted to the right eye matching format.

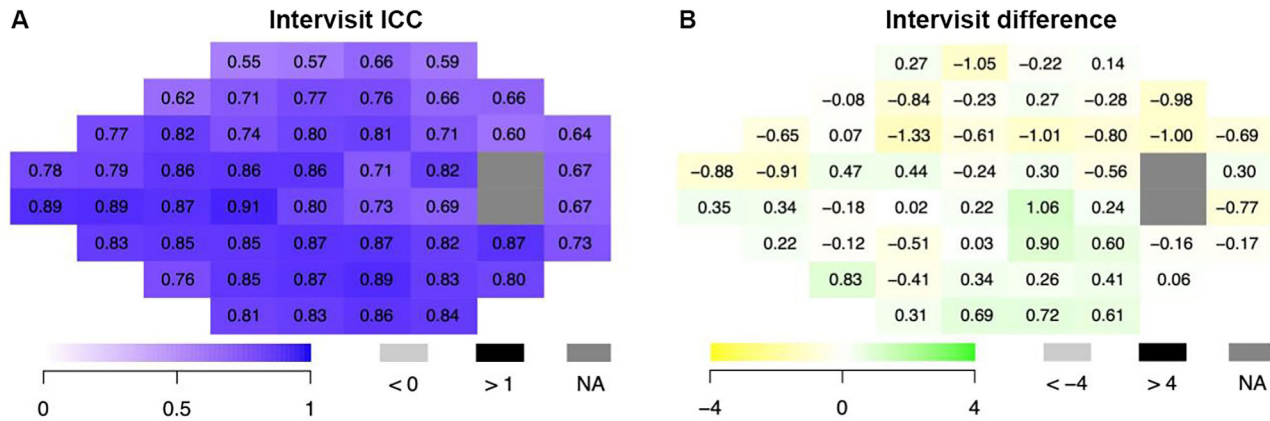


Figure 4. Point-by-point analysis of threshold sensitivities at each VF test location comparing intervisit VisuALL exams ($N = 88$ eyes) by ICC (A) and difference (B). Left eyes were inverted to the right eye matching format.

As for the intervisit point-by-point analysis, 65% of ICC values were good to excellent with the range from all the fields 0.55 to 0.91 (Fig. 4). Thus, the intervisit exams demonstrated slightly lower, but still

strong, pointwise correlations in comparison to the intravisit analysis: 1 point (1.9%) had an excellent ICC, 33 points (63.5%) had good ICCs, and 18 points (34.6%) had moderate ICCs. Among the 10 points that

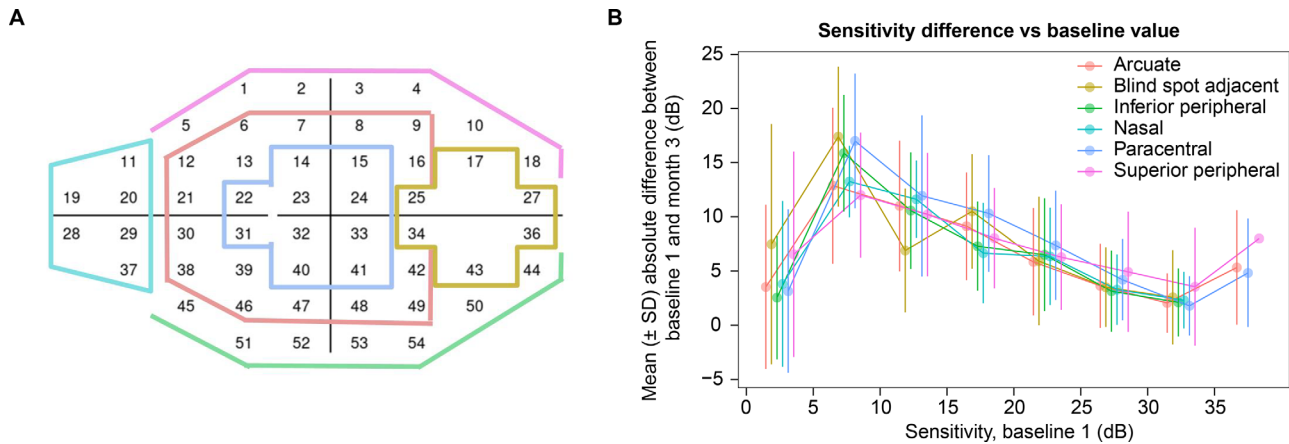


Figure 5. Arcuate, blind spot adjacent, inferior peripheral, nasal, paracentral, and superior peripheral areas of the visual field (A). Mean absolute difference and error bars between baseline and 3-month sensitivities against average baseline sensitivity across all the 52 points subdivided in the respective areas of the visual field (B). Threshold data were binned in intervals of 5 dB.

were adjacent to the two blind spot points in the VF, 7 showed moderate ICC values and 3 showed good ICC values, consistent with typical increased variation around the blind spot. In terms of the mean difference in threshold sensitivities of all 52 points, values ranged from -1.33 to 1.06 dB. Figures 3 and 4 also highlight that the ICCs are high in the important nasal, cecocentral, and arcuate regions.

Figure 5 reveals how intervisit repeatability varied according to average baseline sensitivities in arcuate, blind spot adjacent, inferior peripheral, nasal, paracentral, and superior peripheral areas of the VF (Fig. 5A). In all of these six areas, intervisit exams demonstrated the highest variability in the baseline sensitivity interval of 5 to 10 dB (very severe reduction in sensitivity), whereas the lowest variability was observed in the baseline sensitivity interval of 30 to 35 dB (typical normal healthy value, Fig. 5B).

Discussion

Conventional SAP is the leading method of functional glaucoma testing from a historical perspective, having become widely available on a standardized testing platform.¹⁰ However, there are some limitations of the HFA and similar perimeters, such as their uncomfortable, bulky designs that limit them to in-office use only, eye patching requirement, high cost, and trained technicians assistance requirement. VR headset-based perimetry, a potential alternative glaucoma functional test, can overcome some of those limitations by mimicking the gold standard perimetry in a more ergonomic form factor with the possibility of more economical remote self-testing. In this study, we assessed the same-day and intervisit repeatability of mean deviation and pointwise VF sensitivities obtained with a new VR-based perimeter, the VisuALL S Analyzer, in patients with all stages of glaucoma or ocular hypertension. Our results suggest that VisuALL's VF 24-2 threshold protocol exam can generate VFs that are reproducible both globally and on a pointwise level.¹¹

Establishing the repeatability and prospective diagnostic performance of up-and-coming perimeters like VisuALL is critical to their potential future adoption by clinicians. Previous studies have shown that detecting meaningful clinical change in glaucoma depends not only on the frequency of VF testing and the rate at which a patient's disease is progressing but also on the reliability of VF measurements.^{12,13} As reported by Chauhan et al.,¹² lower variability of VF exams and a greater number of tests, with a repeat-

able baseline, lessen the time period needed to detect change of any parameter. The widely used Humphrey VF test is known to be prone to both short- and long-term fluctuations, similar to any subjective threshold contrast sensitivity testing algorithm, underscoring the importance of new perimeters being able to perform just as well, or ideally with less test-retest variability, than the HFA.¹⁴ Our Bland-Altman analysis showed good agreement in VisuALL MD difference values (intravisit: -0.67 dB, 95% CI, -6.04 to 4.71 dB; intervisit: -0.15 dB, 95% CI, -8.04 to 7.73 dB), indicating strong repeatability. The Spearman correlation for VisuALL was also high (intravisit: $r = 0.89$; intervisit: $r = 0.82$). However, since our protocol did not include the same HFA protocol as VisuALL, a direct comparison between the devices was not possible. These results are in line with HFA studies, where a 2-month interval Bland-Altman analysis showed similar limits of agreement, and the Collaborative Initial Glaucoma Treatment Study indicated a learning effect between the first and second baseline visits within 30 days (MD difference: -0.26 ± 1.9 dB; correlation: 0.91 , 95% CI, 0.88 to 0.92).¹¹ In our study, 84.6% of intravisit and 65.4% of intervisit VF points demonstrated good to excellent repeatability in threshold sensitivities (ICC values ≥ 0.75). Coupled with previous reports of VisuALL's high correlation with HFA mean sensitivity measurements in patients with glaucoma ($r = 0.8$),⁴ our findings would support the potential of VisuALL for clinical disease monitoring. However, it is important to note that we did not directly compare VisuALL's repeatability with conventional HFA due to the patient's burden performing multiple tests on the same day, and there are differences in testing algorithms and the specifications between the two devices. VisuALL presents the stimulus intensity between 3 and 120 cd/m² on a background luminescence of 1 cd/m², and HFA's stimulus has a range of 10 to 3183.1 cd/m² against a background of 10 cd/m².^{3,4}

Interestingly, while previous studies investigating the HFA have reported the presence of a large learning effect between patients' initial VF test sessions, participants who underwent VisuALL testing in our study did not appear to demonstrate a similarly substantial effect. This finding might be explained by the inclusion criteria for the enrollment of experienced Humphrey VF takers. Heijl and Bengtsson¹⁵ administered repeated Humphrey 30-2 full-threshold VF tests (five tests at about 1-week intervals) to 25 patients with newly detected glaucoma who had no prior perimetric testing experience and found evidence of significant perimetric learning: the mean of MD values improved by 2.81 dB ($P < 0.001$) between the first and second

VF tests. After the second test, no significant differences were observed. By contrast, we found no statistically significant difference in the mean of MD values between the three VisuALL tests in our study ($P = 0.893$). The VisuALL reports of the seven eyes of four patients that were out of the limit of agreement range in the intravisit and intervisit Bland–Altman analysis are presented in Supplementary Figure S1, and the large MD differences may be justified by the learning effect (cases 1 OD, 4 OD and OS, 5 OD and OS) or the cloverleaf pattern artifact due to reduced response along the test (cases 2 OD, 3 OS). While it is important to note that all of our participants were experienced Humphrey VF test takers, the VR headset was also an entirely new testing environment for them, so the learning effect needs to be taken into account.

Although VisuALL performed well globally and at the majority of points, lower repeatability was observed at certain locations in the VF. A previous study with 4044 participants that evaluated the pointwise test–retest variability of the HFA within 30 days found an ICC range of 0.66 to 0.89, with peripheral points showing lower correlation than central ones.¹⁴ In our study, the same-day variability of sensitivity points, rather than total deviation (TD) points, showed ICCs ranging from 0.63 to 0.93. Notably, Figures 3 and 4 show more moderate ICCs at points in the superior hemifield, particularly adjacent to the physiologic blind spot or at the periphery field. Among the eight points that had moderate ICCs (<0.75) in the intravisit analysis (Fig. 3), five were in the superior periphery field, and three were immediately adjacent to the blind spot. The lack of repeatability observed in the superior peripheral field could be attributed to factors such as the eyelid or ptosis effect.¹⁶ Additionally, a previous study evaluating the pointwise intravisit and 2-month intervisit variability in peripheral HFA testing showed that fluctuations were higher in the superior quadrant compared to other areas.¹⁷ The lower ICC values on the superior hemifield may also be attributed to the lower sensitivity values on these test points at test 1, as shown in Supplementary Figure S2. Supplementary Table S1 demonstrates that among the three ranges grouped by tertile analysis (<22 dB, 22–26 dB, >26 dB), the lower intra- and intervisit ICC values were found in the points with <22 dB in the baseline test 1. The lower repeatability observed at adjacent locations to the two blind spot points also makes sense in the context of previously published work. Fluctuations in VF measurements have been reported to increase at the border of the blind spot, similar to how greater fluctuations are seen at the edges of glaucomatous scotomas.^{18,19} Kang et al.²⁰ also recently compared two novel perimeters to the HFA, including a similar

VR-based perimeter called IMOVifa (CREWT Medical Systems Inc., Tokyo, Japan), and reported that large differences in VF measurements between devices and tests were particularly seen at peripheral locations near the physiologic scotoma. The authors theorized that these discrepancies may be due to factors such as the naturally variable size and shape of the blind spot between patients and either increased or decreased fixation instability due to the novel testing methods used by the perimeters.²¹ The VisuALL now has eye tracking to help reduce fixation losses that was not available at the time of the study. Refractive error associated with increased axial length is also theorized to impact the shape and location of the blind spot.^{21,22} Although VisuALL incorporates new testing features in an attempt to produce more accurate VF results, our findings suggest that the VR-based perimeter is still subject to increased test–retest variability at the physiologic scotoma's borders like conventional SAP.

For the intervisit analysis, the 18 locations that had moderate ICCs also followed the general pattern of the intravisit analysis, and the amount of test–retest variability tended to increase with decreasing baseline sensitivity (Fig. 5). A number of potential reasons to explain these observations are offered by the existing literature. For one, many studies on the HFA and Octopus have shown that test–retest variability of threshold sensitivity increases with decreasing VF sensitivity; peripheral points in the VF, which are less sensitive due to the natural hill of vision, are therefore more prone to variability than centrally located points.^{23–25} Moreover, glaucomatous vision loss is known to affect peripheral vision first before progressing toward central fixation,²⁶ which may further increase the amount of test–retest variability seen in the periphery.¹⁸

In alignment with the literature, participants in our study with known or suspected glaucoma demonstrated lower repeatability in peripheral locations; interestingly, however, lower repeatability was observed specifically at superior areas of the periphery. A possible reason to account for this finding is if the trial lens or glasses were displaced within the VR headset during testing.²⁷ Specifically, if the lenses sat too low on a participant's face, the upper rim may have interfered with stimuli being displayed and caused threshold sensitivities in the superior periphery to fluctuate between VF tests. Another possible explanation is if the VR headset was not secured properly to the face, which may have allowed some light to leak into the top of the device. However, given the ease of testing due to device portability, VisuALL could be repeated more frequently to obtain more accurate baselines than traditional annual office testing.

Our study is subject to certain limitations. First, our study could have been improved by increasing the sample size ($N = 44$ participants, 88 eyes) to cover more glaucoma severities by HVF cutoffs. The average MD in our study population was -7.4 dB. Therefore, applying these results to patients with moderate-to-severe glaucomatous damage should be done cautiously, especially considering the expected higher test–retest variability in such cases. Moreover, it is not possible to examine whether the intervisit variability is truly due to testing or to the natural fluctuation of the disease states without additional clustered testing. In addition, our study may have benefited from conducting a more detailed analysis to investigate how known factors, as, for example, refractive error²⁸ and pupil size,²⁹ which affect VF repeatability, could impact the VisuALL analysis. We purposefully kept our eligibility criteria broad: patients with any type of glaucoma and a broad range of severity (BCVA $<20/200$) were able to enroll in our study. While this allowed for more generalizable results in theory, it would be useful for future studies to also explore specific subsets of patients among the overall glaucoma population, as the lower range of stimulus of VisuALL compared to HFA may lead to more variability in more advanced cases. While the study was conducted in a clinical setting, the potential impact of environmental background distractions on the test execution was not evaluated. Although the VR headset has no light leak during the fit and should not get affected by the ambient lighting, and the tests were performed in a clinic room away from other people, other factors such as the environment noise should be considered in future assessments, as it can influence the outcomes. Additionally, the VisuALL's background luminescence, which is notably dimmer compared to the HFA with different contrast sensitivity, should be considered when this technology is compared to HFA. This reduced background luminance may affect the sensitivity across different visual field eccentricities, potentially impacting the reproducibility of results when compared to those obtained with HFA.⁵ The varying levels of background brightness can also influence the duration required for a patient to adjust to the background light intensity.

In our analysis of repeatability, we must acknowledge the possibility that changes in the disease could have occurred between the first and third VF exams for some participants. However, this influence is likely minimal, considering the short 3-month period, with median rates of VF loss in clinical populations ranging from -0.05 to -0.62 dB per year.^{30–32} It is also important to note that our evaluation of VisuALL was confined to a clinical setting, leaving its perfor-

mance in home environments untested. Additionally, a direct comparison of VisuALL's performance with the HVF was not feasible in our study, as intravisit and intervisit testing from using the latter was not conducted due to patient time constraints and test fatigue. Further research is necessary for a more comprehensive comparison between these two devices.

In conclusion, our results demonstrate that the VisuALL S Analyzer T-24 protocol test generates repeatable VF results in patients with glaucoma or ocular hypertension who have taken prior HVF, with better repeatability metrics in the points with nonsevere sensitivity loss. This suggests that VisuALL may serve as an alternative to conventional SAP for assessing patients' visual function, especially in situations where traditional bowl perimetry is not available or impractical (e.g., hospital inpatients). Future studies should examine the performance of VisuALL in patients' own homes to give further confidence that VR-based perimetry is reproducible and suitable for at-home monitoring of the VF.

Acknowledgments

The authors thank the participants and their families who took part in the study, as well as the staff, research coordinators, and investigators at each participating institution. Writing assistance provided by Genentech, Inc.

Supported by Genentech, Inc., with participation in the study design, management, analysis, and interpretation of the data; National Institutes of Health Grant P30 EY026877 and Research to Prevent Blindness, Inc. Unrestricted Grant (Stanford Ophthalmology); and an American Glaucoma Society Mentoring for the Advancement of Physician Scientists grant (AS).

Data Sharing: For up-to-date details on Roche's Global Policy on the Sharing of Clinical Information and how to request access to related clinical study documents, see <https://www.roche.com/innovation/process/clinical-trials/data-sharing>.

Disclosure: **R. Nascimento e Silva**, Genentech, Inc. (E); **J.A. Kim**, Genentech, Inc. (E); **Y. Li**, Genentech, Inc. (E); **C. Chen**, Genentech, Inc. (E); **A.F. Chaudhry**, None; **A.R. Berneshawi**, None; **M. Zhang**, Genentech, Inc. (E); **A. Villarreal**, Genentech, Inc. (E); **J. Liu**, None; **A. Shue**, None; **D.S. Chang**, Genentech, Inc. (E); **R.T. Chang**, Genentech, Inc. (F)

References

- Jampel HD, Singh K, Lin SC, et al. Assessment of visual function in glaucoma: a report by the American Academy of Ophthalmology. *Ophthalmology*. 2011;118:986–1002.
- Gardiner SK, Demirel S. Assessment of patient opinions of different clinical tests used in the management of glaucoma. *Ophthalmology*. 2008;115:2127–2131.
- Montelongo M, Gonzalez A, Morgenstern F, et al. A virtual reality-based automated perimeter, device, and pilot study. *Transl Vis Sci Technol*. 2021;10:20.
- Razeghinejad R, Gonzalez-Garcia A, Myers JS, Katz LJ. Preliminary report on a novel virtual reality perimeter compared with standard automated perimetry. *J Glaucoma*. 2021;30:17–23.
- Groth SL, Linton EF, Brown EN, et al. Evaluation of virtual reality perimetry and standard automated perimetry in normal children. *Transl Vis Sci Technol*. 2023;12:6.
- Kelly SR, Bryan SR, Crabb DP. Does eye examination order for standard automated perimetry matter? *Acta Ophthalmol*. 2019;97:e833–e838.
- Artes PH, Iwase A, Ohno Y, et al. Properties of perimetric threshold estimates from Full Threshold, SITA Standard, and SITA Fast strategies. *Invest Ophthalmol Vis Sci*. 2002;43:2654–2659.
- Lau LI, Liu CJ, Chou JC, et al. Patterns of visual field defects in chronic angle-closure glaucoma with different disease severity. *Ophthalmology*. 2003;110:1890–1894.
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155–163.
- Camp AS, Weinreb RN. Will perimetry be performed to monitor glaucoma in 2025? *Ophthalmology*. 2017;124:S71–S75.
- Gillespie BW, Musch DC, Guire KE, et al. CIGTS (Collaborative Initial Glaucoma Treatment Study) Study Group. The collaborative initial glaucoma treatment study: baseline visual field and test-retest variability. *Invest Ophthalmol Vis Sci*. 2003;44:2613–2620.
- Chauhan BC, Garway-Heath DF, Goñi FJ, et al. Practical recommendations for measuring rates of visual field change in glaucoma. *Br J Ophthalmol*. 2008;92:569–573.
- Turpin A, McKendrick AM. What reduction in standard automated perimetry variability would improve the detection of visual field progression? *Invest Ophthalmol Vis Sci*. 2011;52:3237–3245.
- Choi EY, Li D, Fan Y, et al. Predicting global test-retest variability of visual fields in glaucoma. *Ophthalmol Glaucoma*. 2021;4:390–399.
- Heijl A, Bengtsson B. The effect of perimetric experience in patients with glaucoma. *Arch Ophthalmol*. 1996;114:19–22.
- Taniguchi A, Yunoki T, Otsuka M, Hayashi A. Visual field changes in glaucoma patients after blepharoptosis surgery. *Eur J Ophthalmol*. 2022;32:3353–3357.
- Young WO, Stewart WC, Hunt H, Crosswell H. Static threshold variability in the peripheral visual field in normal subjects. *Graefes Arch Clin Exp Ophthalmol*. 1990;228(5):454–457.
- Gardiner SK. Differences in the relation between perimetric sensitivity and variability between locations across the visual field. *Invest Ophthalmol Vis Sci*. 2018;59:3667–3674.
- Haefliger IO, Flammer J. Increase of the short-term fluctuation of the differential light threshold around a physiologic scotoma. *Am J Ophthalmol*. 1989;107:417–420.
- Kang J, De Arrigunaga S, Freeman SE, et al. Comparison of perimetric outcomes from a tablet perimeter, smart visual function analyzer, and Humphrey field analyzer. *Ophthalmol Glaucoma*. 2023;6:509–520.
- Wang M, Shen LQ, Boland MV, et al. Impact of natural blind spot location on perimetry. *Sci Rep*. 2017;7:6143.
- Lin F, Chen S, Song Y, et al. Glaucoma Suspects with High Myopia Study Group. Classification of visual field abnormalities in highly myopic eyes without pathologic change. *Ophthalmology*. 2022;129:803–812.
- Artes PH, Hutchison DM, Nicoleta MT, et al. Threshold and variability properties of matrix frequency-doubling technology and standard automated perimetry in glaucoma. *Invest Ophthalmol Vis Sci*. 2005;46:2451–2457.
- Boeglin RJ, Caprioli J, Zulauf M. Long-term fluctuation of the visual field in glaucoma. *Am J Ophthalmol*. 1992;113:396–400.
- Stewart WC, Hunt HH. Threshold variation in automated perimetry. *Surv Ophthalmol*. 1993;37:353–361.
- Elze T, Pasquale LR, Shen LQ, et al. Patterns of functional vision loss in glaucoma determined with archetypal analysis. *J R Soc Interface*. 2015;12:20141118.
- Keltner JL, Johnson CA, Cello KE, et al. Ocular Hypertension Treatment Study Group. Classification of visual field abnormalities in the ocu-

- lar hypertension treatment study. *Arch Ophthalmol*. 2003;121:643–650.
28. Weinreb RN, Perlman JP. The effect of refractive correction on automated perimetric thresholds. *Am J Ophthalmol*. 1986;101:706–709.
 29. Lindenmuth KA, Skuta GL, Rabbani R, Musch DC. Effects of pupillary constriction on automated perimetry in normal eyes. *Ophthalmology*. 1989;9:1298–1301.
 30. Saunders LJ, Medeiros FA, Weinreb RN, Zangwill LM. What rates of glaucoma progression are clinically significant? *Expert Rev Ophthalmol*. 2016;11:227–234.
 31. Chauhan BC, Malik R, Shuba LM, et al. Rates of glaucomatous visual field change in a large clinical population. *Invest Ophthalmol Vis Sci*. 2014;55:4135–4143.
 32. Heijl A, Buchholz P, Norrgren G, Bengtsson B. Rates of visual field progression in clinical glaucoma care. *Acta Ophthalmol*. 2013;91:406–412.