Journal Pre-proof

Visual field endpoints for neuroprotective trials: a case for AI driven patient enrichment

Andrew Chen , Giovanni Montesano , Randy Lu , Cecilia S. Lee , David P. Crabb , Aaron Y. Lee

PII: S0002-9394(22)00283-5
DOI: https://doi.org/10.1016/j.ajo.2022.07.013
Reference: AJOPHT 12295

To appear in: *American Journal of Ophthalmology*

Received date: January 10, 2022
Revised date: June 27, 2022
Accepted date: July 18, 2022

Please cite this article as: Andrew Chen , Giovanni Montesano , Randy Lu , Cecilia S. Lee , David P. Crabb , Aaron Y. Lee , Visual field endpoints for neuroprotective trials: a case for AI driven patient enrichment, *American Journal of Ophthalmology* (2022), doi: https://doi.org/10.1016/j.ajo.2022.07.013

Visual field endpoints for neuroprotective trials: a case for AI driven patient enrichment

Andrew Chen[1], Giovanni Montesano[2,3], Randy Lu[1], Cecilia S. Lee[1], David P. Crabb[2], Aaron Y. Lee[1]

[1] Department of Ophthalmology, University of Washington, Seattle, Washington, United States
[2] Optometry and Visual Sciences, City, University of London, London, UK
[3] NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK

Short Title: Neuroprotective trial VF endpoints and AI patient enrichment

Supplemental Material available at AJO.com

Corresponding Author:
Andrew Chen, MD
Assistant Professor
Department of Ophthalmology
University of Washington
Box 359608, 325 Ninth Avenue
Seattle WA 98104
Phone: (206) 221-5931
Email: achen20@uw.edu

ABSTRACT

Purpose: To evaluate if an artificial intelligence (AI) model can better select candidates that would demonstrate visual field (VF) progression in order to shorten the duration or the number of patients needed for a clinical trial

Design: Retrospective cohort study

Methods: 7,428 eyes of 3,871 patients from the University of Washington Department of Ophthalmology VF Dataset were included. Progression was defined as at least 5 locations with greater than 7 dB of change compared to baseline on two consecutive tests. Progression for all patients, a subgroup of the fastest progressing based on survival curves, and patients selected based on an elastic net Cox regression model were compared. The model was trained on pointwise threshold deviation values of the first VF, age, gender, laterality and the mean total deviation (MD) at baseline.

Results: 13% of all patients met the criteria for progression at five years. Differences in survival were observed when stratified by MD and age ($p < 0.0001$). Those at risk of progression included patients 60 to 80 years old with an initial MD < -5.0. This subgroup decreased the sample size required to detect progression compared to the entire cohort. The AI model-selected patients required the lowest number of patients for all effect sizes and trial lengths. For a trial length of 3 years and effect size of 30%, the number of patients required was 1656 [95% confidence interval (CI), 1638–1674] , 903 [95% CI, 884–922], and 636 [95% CI, 625–646] for the entire cohort, the subgroup, and the model-selected patients, respectively.

Conclusion: An AI model can identify high risk patients to substantially reduce the number of patients needed or study duration required to meet clinical trial endpoints.

INTRODUCTION

Glaucoma is a slowly progressive, sight-threatening optic neuropathy that affects over 80 million people worldwide.[1] The intraocular pressure (IOP) is the only modifiable risk factor for

glaucoma progression. Although many treatments to decrease the IOP exist such as topical therapy, laser, and surgical interventions, a substantial portion of patients will progress with continuous optic nerve damage and visual field loss despite being on treatments.[2,3] Therefore, given the number of people with glaucoma, additional benefit from neuroprotective treatments, even if modest, could prevent blindness in thousands of people.

Progression of visual field (VF) loss is an important clinical endpoint in glaucoma, in particular when assessing the efficacy of treatments. The Federal Drug Administration (FDA) has stated that VF loss has likely occurred if ≥5 VF locations have significant change beyond the 5% probability level or if there is at least a 7 dB between-group mean difference for the entire field.[4] Subsequently, the FDA has defined the criteria for VF progression as at least 5 locations with greater than 7 dB of change in prior regulatory studies. However, conducting clinical trials of new therapy using the FDA progression criteria has led to significant follow-up time and resources resulting in key challenges. One solution would be to recruit patients that are more likely to exhibit progression, thus magnifying the differences and increasing the power of clinical trials. These patients are often identified based on their baseline features, such as age, ethnicity, IOP and initial visual field damage.[3,5–12] However, it can be challenging to find the most informative combination of these baseline parameters to identify patients that are more likely to exhibit VF damage progression.

Artificial intelligence (AI) has shown many promising applications in healthcare recently and could offer a better solution for selecting ideal candidates. Although previous studies have primarily been on automated analyses, diagnostics, and outcome predictions, the potential application of AI in clinical trials is particularly attractive.[13] In this study, we explore whether AI could identify patients who will progress rapidly using baseline data and analyze how that would impact the required follow up period for clinical trials. We hypothesized that AI could identify high-risk patients thus resulting in clinical trials that require significantly shorter follow up periods to meet the clinical endpoints.

METHODS

This study was exempted from the University of Washington Institutional Review board due to de-identified data usage. All research in this study conformed to the Declaration of Helsinki. The University of Washington Department of Ophthalmology VF database was used as the data source for this study. The VF dataset has been previously detailed[14] and can be found publicly online at https://github.com/uw-biomedical-ml/uwhvf. In short, all patients in the dataset underwent Humphrey Field Analyzer (HFA) II 24-2 stimulus size III VFs, performed with either a Swedish Interactive Thresholding Algorithm (Standard or Fast) or a full-threshold strategy.

Every VF for each eye was temporally aligned and the first VF was treated as the baseline VF for each eye. If there was more than 2.25 years between VFs then the eye was deemed to have been lost to follow-up. The criteria for progression based on at least 5 points of at least 7 dB decrease was measured against the baseline VF and only if two consecutive follow-up VFs had ≥ 7 dB loss in total deviation (TD) in the same ≥ 5 visual field locations compared to the baseline VF. Kaplan Meier survival curves were constructed. Eyes lost to follow up were censored. Additional survival curves were analysed based on age in decades and initial mean total deviation (MD). Of note, the HFA Mean Deviation makes use of a weighting system based on location-specific variability estimates, which cannot be extracted from the HFA device. Therefore, we reported mean total deviation, which is simply the arithmetic average of the TD values and whose interpretation is essentially equivalent.[15,16] The log rank test was used to compare survival curves. A p value of less than 0.05 was used for significance testing.

The whole dataset was then split at the patient level 50% for training and 50% for held-out test-set (Supplementary Figure). In the training set, a separate elastic-net Cox regression model was optimized for each time point (1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, and 5 years) using patient level 10-fold cross validation and the values for alpha, gamma, and lambda were optimized for

the best Harrell's C-index. For input, the model had total deviation point-wise values as a vector of 54 points, age, gender, and laterality. After the parameters for the best fitting elastic-net Cox model were fitted, the training set was again run through the model and the upper 75th quartile of the risk probability was calculated. This value was set as the threshold for determining inclusion at inference time.

For the held-out test set, three different scenarios for selecting patients for simulating inclusion in a clinical trial were tested. The first was no selection and using all the patients in the test set; the second was a selection of the fastest progressing cohort, identified based on the results of Kaplan Meier survival curves built with the training set and stratified by baseline decade of age and initial MD at fixed cut-offs; the third was a selection based on the elastic-net Cox regression model for patients who were above the risk threshold determined from the training set. For each of these, the event rates were calculated at different time points for simulating trial length and a sample size calculation was performed by simulating a percent reduction in the event rate and using the two proportion, equal sample size estimation by Cohen et al (R, pwr package, version 1.3-0) for a power of 80% and an alpha of 0.05. The bootstrap method was performed to calculate robust means and 95% confidence intervals for the sample sizes.

Python (https://www.python.org/, version 3.8.3) was used to perform preprocessing and R (https://www.r-project.org/, version 4.0.3) with the elastic-net library (glmnet, version 4.1-2) was used for fitting the model. The code for this paper was open sourced under the BSD license along with this paper (https://github.com/uw-biomedical-ml/uwhvf-endpoint).

RESULTS

The baseline demographics of this dataset have been previously described. All 7,428 eyes of 3,871 patients were included in this analysis. The mean age was 61.8 (std: 14.8) years and 46% were male. Initial MD of all the VFs was -5.8 (std: 6.0) dB. The average time to event

interval was 2.2 years (std: 1.8) in eyes that exhibited progression. A total of 3614 eyes of 1,880 patients were used for the training set of the machine learning model and the held-out test set included 3814 eyes of 1991 patients.

A total of 13% of patients met the criteria for progression at five years from the initial visual field (Figure 1). Differences in survival curves by log rank test were observed when stratified by initial MD ($p < 0.0001$) or age ($p < 0.0001$). Faceting by both decade of life and by initial MD revealed that the age group of 60 to 80 with an initial MD of worse than -5.0 progressed the most in the cohort (Figure 2). This criteria was used for the subsequent power calculation as the “subgroup” simulation.

An elastic-net Cox regression model was fitted in the training set for each time-point of interest using the pointwise TD values of the first visual field, age, gender, tested eye and the MD at baseline. The fitted coefficient values of the model are shown in Supplementary Figure 2 with darker regions corresponding to more strongly negative coefficient values.

The trained models were used to select participants from the test set to include in the simulated clinical trial and the resulting sample size calculations are shown in Figure 3. For effect sizes of 10% to 50% and for every time point, the AI driven inclusion of participants gave the best statistical power and was the best strategy for reducing the number of patients needed for a clinical trial. The C-indexes ranged from 0.763 - 0.787. For example, the number of patients required for a three year trial with an effect size of 30% was 1656 [95% confidence interval (CI), 1638–1674] , 903 [95% CI, 884–922], and 636 [95% CI, 625–646] for the entire cohort, subgroup, and model-selected patients, respectively. The C-index for the three year model was 0.765. For a 5 year trial, 30% effect size, the number of patients required was 750 [95% CI, 741–759] , 306 [95% CI, 300–313], and 268 [95% CI, 263–273] for the entire cohort, subgroup, and model-selected patients respectively. The C-index for the five year model was 0.767.

DISCUSSION

In this study we simulated the impact of current FDA criteria of VF progression used for regulatory approvals of new glaucoma treatments using longitudinal data of patients receiving care at a tertiary care academic practice. We found that the majority of our cohort did not meet the VF progression criteria of ≥ 7 dB decrease in threshold sensitivity in ≥ 5 visual field locations even after being followed for 5 years. Due to varying rates of progression in patients, we calculated that the number of patients needed in a two-year clinical trial for this progression criteria would be over 4000 even if we assume a strong effect size of 30% reduction in progression rate. While recognizing the resulting challenges of conducting clinical trials of long duration and large sample size, we demonstrate that using an AI model to identify and recruit high risk patients into clinical trials may shorten clinical trials by six months or substantially reduce the number of patients needed to meet the endpoint at most effect sizes and trial durations. For example, for an effect size of 30%, the AI based selection yielded a sample size of 636 for 80% power at three years, compared to 1656 patients when no selection was applied. Alternatively, the trial with no selection would need to last greater than 5 years to obtain the same power with a similar sample size (Figure 3).

The overall rate of progression was 13% as defined by the FDA criteria, which is lower than that of glaucoma trials such as the Early Manifest Glaucoma Trial (EMGT). The visual field dataset used for this study was curated from all patients who had visual field testing at a single academic center, which would also include glaucoma suspects.[14] Other comparable cohorts also demonstrate a very slow rate of progression (median MD rate of -0.05 dB/yr).[17]

In our cohort, older patients were more likely to meet the criteria for VF progression, specifically patients over the age of 60 years. Age has been found to be associated with faster rates of visual field decline in prior studies.[12,18,19] In this study, patients with worse visual fields at baseline, MD < -5 dB, were also more likely to meet the VF progression criteria. This is consistent with some prior studies.[7,20,21] The same trends for the association of initial sensitivity

on progression rates were observed within each age category. Of note, the AI algorithm was not trained to identify high-risk patients for clinical management.

IOP pressure reduction has been the mainstay of glaucoma treatment and for regulatory approval of new therapies and intervention. While IOP can delay onset and slow progression, it is only a risk factor for glaucoma, and many patients will progress despite treatment. Furthermore, similar IOP reductions can lead to different rates of VF loss.[22] There are several IOP-lowering therapies, but no clinically-available neuroprotective agents for glaucoma.[23] Thus, there are no non-inferiority metrics to compare neuroprotective agents, only what the FDA has set forth as clinically meaningful.[4]

The phase III clinical trials for memantine as a neuroprotective agent for glaucoma were randomized, double-masked, placebo-controlled, 4-year studies that enrolled 2298 patients from 128 centers.[24] Criteria for SAP VF progression consisted of 5 or more visual field locations that demonstrated significant reductions from baseline in a confirmatory visual field obtained within 8 weeks. The study found no difference between the control and treatment arms at the estimated cost of ~100 million dollars.[25] These results cast a shadow of doubt on the feasibility of glaucoma neuroprotection trials. The current FDA criteria would likely require study designs that are resource-intensive and cost-prohibitive. Better VF criteria for trials exist and our work highlights a need for regulatory bodies to systematically review this literature.

Currently there is no reference standard for what constitutes visual field progression. This study used criteria set forth by the FDA for regulatory approval, but other methodologies to evaluate for visual field progression exist. In general, there are the following groups: clinical judgment, defect classification, event-based analysis, and trend-based analysis.[26] A combination of these VF criteria have been used as the VF portion of the main outcome for major glaucoma trials. The EMGT used the presence of at least three test points with a measured sensitivity below the 5th percentile of test-retest variability expected from the baseline at the same locations over three consecutive tests for the VF portion of their main outcome

measure.[27] The Advanced Glaucoma Intervention Study used a scoring system from 0 - 20 based on the defect size, depth, and location.[28] The Collaborative Initial Glaucoma Treatment Study also used a 20 point system, albeit with slightly different criteria.[29] The United Kingdom Glaucoma Treatment Study's VF criteria consisted of three locations worse than baseline at P < 5% in two consecutive VF and three locations worse than baseline at the 5% levels in the two subsequent consecutive reliable visual fields.[30] The locations meeting criteria at baseline did not have to be identical to the locations on subsequent fields.[30] The Guided Progression Analysis (GPA) of the HFA is used as another clinical tool, and it highlights points on a probability plot if the change exceeds the variability of stable glaucoma patients.[31]

There may be alternative methodologies to decrease the time required to detect progression. Though not accepted for regulatory use, trend-based analysis may be able to decrease the time required to detect progression and offers the ability to extrapolate future VF data.[32–34] De Moraes et al. argue that the criteria set forth by the FDA corresponds to a rate of progression ≤ -0.5 dB/year for ≥ 5 locations.[35] As such trend-based analysis could offer the potential to extrapolate future VF behavior to earlier quantify treatment effect. Moreover, simulation studies have shown that trend-based analyses with hierarchical models can greatly improve the power compared to event-based methods, and might therefore be preferable.[36] Unfortunately the FDA has yet to clarify whether progression based on techniques such as trend-based analysis is acceptable. Machine learning has also been utilized to accurately predict future visual field loss.[37] Other options to improve on neuroprotection study design include selecting for patients at risk of faster glaucoma progression, modifying VF testing intervals[38], adding structural measures for glaucoma[39,40], employing futility trials[41], selecting patients that exhibit smaller variability in their VF test results[42] or using the two eyes from the same patient for paired comparisons of treatment and placebo[42].

Limitations of this study include the lack of associated clinical information such as diagnosis, medication regimen, and concurrent surgeries. Additionally, reliability indices and

test-rest limits were not available in this dataset. As such the GPA could not be reliably replicated. A large number of patients were lost to follow-up as would be expected in a non-study clinical setting. In addition, this analysis was performed on a large retrospective database from a single clinical center and may not generalize to other populations. As previously described, this cohort includes all patients at the University of Washington who underwent VF testing. The inclusion of glaucoma suspect patients and unreliable tests would presumably increase the sample size calculations, but the AI driven model was better at selecting visual fields at risk of progression compared to subgroup analysis of patients with known risk factors for VF progression. Furthermore, the VF testing intervals in this retrospective cohort do not reflect those that can be employed by clinical trials, such as multiple baseline tests, the "wait-and-see" approach, multiple tests on the same day, and trend-based endpoints, which can reduce the sample size required.[30,36,38] Future work can include comparison of different criteria for progression such as the GPA and incorporation of clinical information or optic nerve head structural data to improve model prediction.

In conclusion, we demonstrate the promise of AI approaches for identifying populations at higher risk of visual field progression to increase the power of clinical trials and reduce the trial duration or number of patients needed.

ACKNOWLEDGEMENTS

Financial Support: NIH/NEI K23EY029246 (Bethesda, MD), NIH/NIA R01AG060942 (Bethesda, MD), NIH/NIA U19AG066567 (Bethesda, MD), Latham Vision Research Innovation Award (Seattle, WA), and an unrestricted grant from Research to Prevent Blindness (New York, NY). The sponsors / funding organizations had no role in the design or conduct of this research.

Financial Disclosures: Dr. G. Montesano is a consultant for CenterVue SpA; Dr. D P. Crabb reports consultancy agreements with Apellis, Santen, Centervue, Medisoft, grants from Allergan, Apellis, Santen, speaker fees from Allergan, Santen, THEA, all outside of the submitted work. Dr. A. Lee reports support from the US Food and Drug Administration, grants from Santen, Carl Zeiss Meditec, and Novartis, personal fees from Genentech, Topcon, and Verana Health, outside of the submitted work; This article does not reflect the opinions of the Food and Drug Administration.

Table of Contents Statement

Table of Contents Statement: Progression of visual field loss is an important clinical endpoint in glaucoma clinical trials; however, long follow-up times and significant resources are required. An artificial intelligence model can improve the selection for patients at higher risk of progression to decrease the number of patients needed to be enrolled and shorten the required clinical trial time.

REFERENCES

1. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol*. 2006;90(3):262-267.

2. Kass MA, Heuer DK, Higginbotham EJ, et al. The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Arch Ophthalmol*. 2002;120(6):701-713; discussion 829-830.

3. The Advanced Glaucoma Intervention Study (AGIS): 7. The relationship between control of intraocular pressure and visual field deterioration.The AGIS Investigators. *Am J Ophthalmol*. 2000;130(4):429-440.

4. Weinreb RN, Kaufman PL. The glaucoma research community and FDA look to the future: a report from the NEI/FDA CDER Glaucoma Clinical Trial Design and Endpoints Symposium. *Invest Ophthalmol Vis Sci*. 2009;50(4):1497-1505.

5. Richler M, Werner EB, Thomas D. Risk factors for progression of visual field defects in medically treated patients with glaucoma. *Can J Ophthalmol*. 1982;17(6):245-248.

6. Gordon MO, Beiser JA, Brandt JD, et al. The Ocular Hypertension Treatment Study: Baseline Factors That Predict the Onset of Primary Open-Angle Glaucoma. *Arch Ophthalmol*. 2002;120(6):714-720.

7. AGIS Investigators. The Advanced Glaucoma Intervention Study (AGIS): 12. Baseline risk factors for sustained loss of visual field and visual acuity in patients with advanced glaucoma. *Am J Ophthalmol*. 2002;134(4):499-512.

8. Leske MC, Heijl A, Hyman L, et al. Predictors of long-term progression in the early manifest

glaucoma trial. *Ophthalmology*. 2007;114(11):1965-1972.

9. Musch DC, Gillespie BW, Lichter PR, Niziol LM, Janz NK, CIGTS Study Investigators. Visual field progression in the Collaborative Initial Glaucoma Treatment Study the impact of treatment and other baseline factors. *Ophthalmology*. 2009;116(2):200-207.

10. Fujino Y, Asaoka R, Murata H, et al. Evaluation of Glaucoma Progression in Large-Scale Clinical Data: The Japanese Archive of Multicentral Databases in Glaucoma (JAMDIG). *Invest Ophthalmol Vis Sci*. 2016;57(4):2012-2020.

11. Asaoka R, Murata H, Fujino Y, et al. Effects of ocular and systemic factors on the progression of glaucomatous visual field damage in various sectors. *Br J Ophthalmol*. 2017;101(8):1071-1075.

12. Kim JH, Rabiolo A, Morales E, et al. Risk Factors for Fast Visual Field Progression in Glaucoma. *Am J Ophthalmol*. 2019;207:268-278.

13. Lee CS, Lee AY. How Artificial Intelligence Can Transform Randomized Controlled Trials. *Transl Vis Sci Technol*. 2020;9(2):9-9.

14. Montesano G, Chen A, Lu R, Lee CS, Lee AY. UWHVF: A Real-World, Open Source Dataset of Perimetry Tests From the Humphrey Field Analyzer at the University of Washington. *Transl Vis Sci Technol*. 2022;11(1):2.

15. Muthusamy V, Turpin A, Walland MJ, Nguyen BN, McKendrick AM. Increasing the Spatial Resolution of Visual Field Tests Without Increasing Test Duration: An Evaluation of ARREST. *Transl Vis Sci Technol*. 2020;9(13):24.

16. Montesano G, Bryan SR, Crabb DP, et al. A Comparison between the Compass Fundus Perimeter and the Humphrey Field Analyzer. *Ophthalmology*. 2019;126(2):242-251.

17. Chauhan BC, Malik R, Shuba LM, Rafuse PE, Nicolela MT, Artes PH. Rates of glaucomatous visual field change in a large clinical population. *Invest Ophthalmol Vis Sci*. 2014;55(7):4135-4143.

18. Nouri-Mahdavi K, Hoffman D, Coleman AL, et al. Predictive factors for glaucomatous visual field progression in the Advanced Glaucoma Intervention Study. *Ophthalmology*. 2004;111(9):1627-1635.

19. Prata TS, De Moraes CGV, Teng CC, Tello C, Ritch R, Liebmann JM. Factors affecting rates of visual field progression in glaucoma patients with optic disc hemorrhage. *Ophthalmology*. 2010;117(1):24-29.

20. Lee JM, Caprioli J, Nouri-Mahdavi K, et al. Baseline prognostic factors predict rapid visual field deterioration in glaucoma. *Invest Ophthalmol Vis Sci*. 2014;55(4):2228-2236.

21. Boodhna T, Saunders LJ, Crabb DP. Are rates of vision loss in patients in English glaucoma clinics slowing down over time? Trends from a decade of data. *Eye* . 2015;29(12). doi:10.1038/eye.2015.161

22. Krupin T, Liebmann JM, Greenfield DS, Ritch R, Gardiner S, Low-Pressure Glaucoma Study Group. A randomized trial of brimonidine versus timolol in preserving visual function: results from the Low-Pressure Glaucoma Treatment Study. *Am J Ophthalmol*. 2011;151(4):671-681.

23. Sena DF, Lindsley K. Neuroprotection for treatment of glaucoma in adults. *Cochrane Database Syst Rev*. 2017;1:CD006539.

24. Weinreb RN, Liebmann JM, Cioffi GA, et al. Oral Memantine for the Treatment of Glaucoma: Design and Results of 2 Randomized, Placebo-Controlled, Phase 3 Studies. *Ophthalmology*. 2018;125(12):1874-1885.

25. Khatib TZ, Martin KR. Protecting retinal ganglion cells. *Eye* . 2017;31(2):218-224.

26. Spry PGD, Johnson CA. Identification of progressive glaucomatous visual field loss. *Surv Ophthalmol*. 2002;47(2):158-173.

27. Leske MC, Heijl A, Hyman L, Bengtsson B. Early Manifest Glaucoma Trial: design and baseline data. *Ophthalmology*. 1999;106(11):2144-2153.

28. Advanced Glaucoma Intervention Study. 2. Visual field test scoring and reliability. *Ophthalmology*. 1994;101(8):1445-1455.

29. Musch DC, Lichter PR, Guire KE, Standardi CL. The Collaborative Initial Glaucoma Treatment Study: study design, methods, and baseline characteristics of enrolled patients. *Ophthalmology*. 1999;106(4):653-662.

30. Garway-Heath DF, Crabb DP, Bunce C, et al. Latanoprost for open-angle glaucoma (UKGTS): a randomised, multicentre, placebo-controlled trial. *Lancet*. 2015;385(9975):1295-1304.

31. Artes PH, O'Leary N, Nicolela MT, Chauhan BC, Crabb DP. Visual field progression in glaucoma: what is the specificity of the Guided Progression Analysis? *Ophthalmology*. 2014;121(10):2023-2027.

32. de Moraes CG, Song C, Liebmann JM, Simonson JL, Furlanetto RL, Ritch R. Defining 10-2 visual field progression criteria: exploratory and confirmatory factor analysis using pointwise linear regression. *Ophthalmology*. 2014;121(3):741-749.

33. Nouri-Mahdavi K, Hoffman D, Gaasterland D, Caprioli J. Prediction of visual field progression in glaucoma. *Invest Ophthalmol Vis Sci*. 2004;45(12):4346-4351.

34. Chen A, Nouri-Mahdavi K, Otarola FJ, Yu F, Afifi AA, Caprioli J. Models of glaucomatous

visual field loss. *Invest Ophthalmol Vis Sci*. 2014;55(12):7881-7887.

35. De Moraes CG, Liebmann JM, Levin LA. Detection and measurement of clinically meaningful visual field progression in clinical trials for glaucoma. *Prog Retin Eye Res*. 2017;56:107-147.

36. Improving the Feasibility of Glaucoma Clinical Trials Using Trend-Based Visual Field Progression End Points. *Ophthalmology Glaucoma*. 2019;2(2):72-77.

37. Wen JC, Lee CS, Keane PA, et al. Forecasting future Humphrey Visual Fields using deep learning. *PLoS One*. 2019;14(4):e0214875.

38. Crabb DP, Garway-Heath DF. Intervals between visual field tests when monitoring the glaucomatous patient: wait-and-see approach. *Invest Ophthalmol Vis Sci*. 2012;53(6):2770-2776.

39. Weinreb RN, Kaufman PL. Glaucoma research community and FDA look to the future, II: NEI/FDA Glaucoma Clinical Trial Design and Endpoints Symposium: measures of structural change and visual function. *Invest Ophthalmol Vis Sci*. 2011;52(11):7842-7851.

40. Medeiros FA, Lisboa R, Zangwill LM, et al. Evaluation of progressive neuroretinal rim loss as a surrogate end point for development of visual field loss in glaucoma. *Ophthalmology*. 2014;121(1):100-109.

41. Quigley HA. Clinical trials for glaucoma neuroprotection are not impossible. *Curr Opin Ophthalmol*. 2012;23(2):144-154.

42. Montesano G, Quigley HA, Crabb DP. Improving the Power of Glaucoma Neuroprotection Trials Using Existing Visual Field Data. *Am J Ophthalmol*. 2021;229:127-136.

FIGURE LEGENDS

Figure 1. Survival by baseline mean total deviation. Kaplan-meier survival curve for no visual field progression as defined by a ≥ 7 dB decrease in total deviation from baseline in ≥ 5 visual field locations for ≥ 2 consecutive visits stratified by initial mean total deviation. The shading represents the 95% confidence interval for each respective strata. MD = Mean Total Deviation

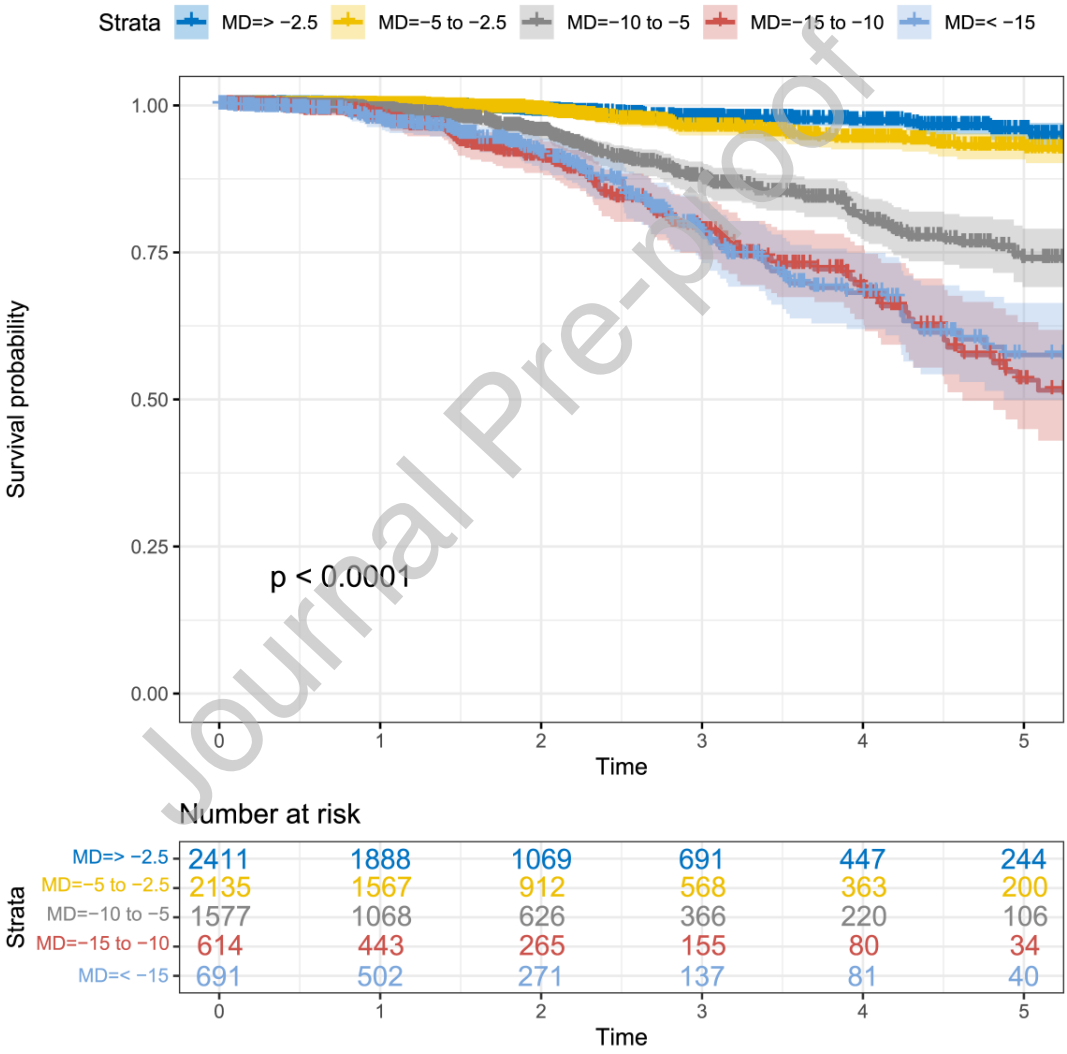

Figure 2. Survival by baseline mean total deviation and by decade of life. Kaplan-meier survival curve for no visual field progression as defined by a ≥ 7 dB decrease in total deviation from baseline in ≥ 5 visual field locations for ≥ 2 consecutive visits stratified by initial mean total deviation within each age group. The shading straddling represents the 95% confidence interval for each respective strata.

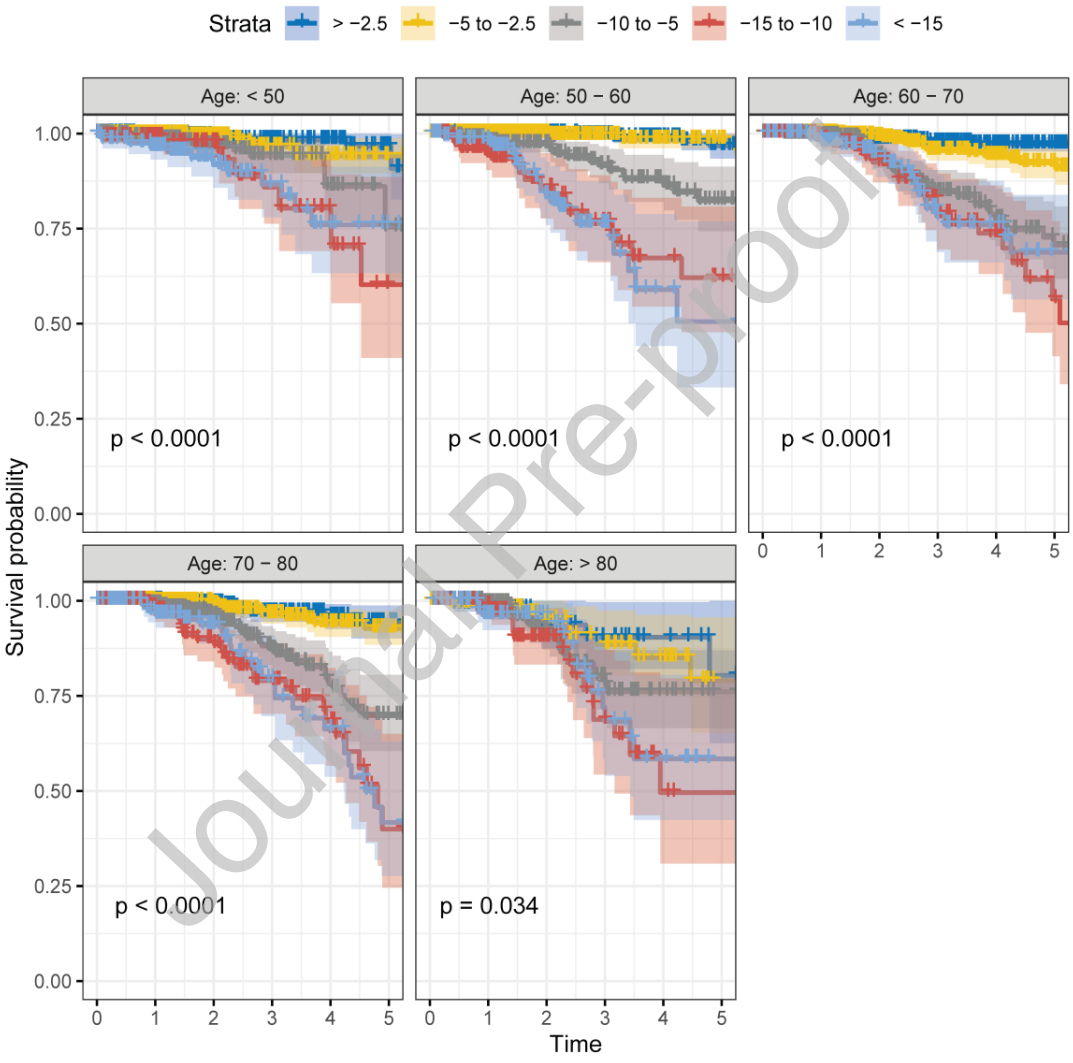

Figure 3. Sample size calculations using data from the held-out test set by three different selection criteria. These graphs plot the number of patients needed to detect a difference in visual field progression for different trial durations and treatment effect sizes. The red line simulates no selection criteria. The blue line represents using the criteria of recruiting patients who are 60 to 80 years of age with a starting mean total deviation of -5.0 or worse. The green line represents using the elastic-net Cox regression models (artificial intelligence model) for patient selection.