



Published in final edited form as:

*Ophthalmology*. 2019 June ; 126(6): 822–828. doi:10.1016/j.ophtha.2019.01.029.

## Agreement and Predictors of Discordance of Six Visual Field Progression Algorithms

Osamah J. Saeedi, MD, MS<sup>\*,1</sup>, Tobias Elze, PhD<sup>2</sup>, Loris D'Acunto, MS<sup>3</sup>, Ramya Swamy, MD, MS<sup>1</sup>, Vikram Hegde, MS<sup>3</sup>, Surabhi Gupta, MS<sup>3</sup>, Amin Venjara, MS<sup>4</sup>, Joby Tsai, BS<sup>1</sup>, Jonathan S Myers, MD<sup>5</sup>, Sarah R. Wellik, MD<sup>6</sup>, Carlos Gustavo De Moraes, MD, MPH<sup>7</sup>, Louis R. Pasquale, MD<sup>8,9</sup>, Lucy Q. Shen, MD<sup>8</sup>, Michael V. Boland, MD, PhD<sup>10</sup>

<sup>1</sup>University of Maryland Department of Ophthalmology and Visual Sciences, Baltimore, MD

<sup>2</sup>Schepens Eye Research Institute, Harvard Medical School, Boston, MA

<sup>3</sup>San Francisco, CA

<sup>4</sup>Princeton, NJ

<sup>5</sup>Wills Eye Hospital, Philadelphia, PA

<sup>6</sup>Bascom Palmer Eye Institute, University of Miami School of Medicine, Miami, FL

<sup>7</sup>Bernard and Shirlee Brown Glaucoma Research Laboratory, Edward S. Harkness Eye Institute, Columbia University Medical Center, New York, NY, USA.

<sup>8</sup>Massachusetts Eye and Ear, Harvard Medical School, Boston, MA

<sup>9</sup>Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

<sup>10</sup>Wilmer Eye Institute, Johns Hopkins University, Baltimore, MD

### Abstract

**Purpose:** To determine the agreement of six established visual field progression algorithms in a large dataset of visual fields from multiple institutions and to determine predictors of discordance amongst these algorithms.

**Design:** Retrospective Longitudinal Cohort

**Subjects, Participants, and/or Controls:** Visual fields from five major eye care institutions in the United States. This analysis included a subset of eyes with at least five SITA-Standard 24-2 visual fields that met our reliability criteria. Of a total of 831,240 fields, a subset of 90,713 visual fields of 13,156 eyes of 8,499 patients met the inclusion criteria.

\*Corresponding Author: Osamah Saeedi, MD, Department of Ophthalmology and Visual Sciences, University of Maryland School of Medicine, 419 W Redwood Street, Suite 470, Baltimore, MD 21201, USA (osaeedi@som.umaryland.edu).

Address for Reprints: 419 W. Redwood St, Suite 470, Baltimore, MD 21201

**Conflict of Interest:** No conflicting relationship exists for any author

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Methods:** Six commonly used visual field progression algorithms (mean deviation slope, visual field index slope, Advanced Glaucoma Intervention Study, Collaborative Initial Glaucoma Treatment Study, pointwise linear regression, and permutation of pointwise linear regression) were applied to this cohort and each eye was determined to be stable or progressing using each measure. Agreement between individual algorithms was tested using Cohen's Kappa coefficient. Bivariate and multivariable analyses were used to determine predictors of *discordance* (3 algorithms progressing and 3 algorithms stable).

**Main Outcome Measures:** Agreement and *discordance* between algorithms

**Results:** Individual algorithms showed poor to moderate agreement with each other when compared directly (Kappa range: 0.12 – 0.52). 11.7% of eyes progressed based on at least four algorithms. Major predictors of *discordance*, or lack of agreement among algorithms were more depressed initial MD ( $P < 0.01$ ) and older age at first available visual field ( $P < 0.01$ ). A greater number of visual fields ( $P < 0.01$ ), more years of follow up ( $P < 0.01$ ) and eye care institution ( $P = 0.03$ ) were also associated with *discordance*.

**Conclusions:** This extremely large comparative series demonstrates that existing algorithms have limited agreement, and that agreement varies with clinical parameters including institution. These issues underscore the challenges to the clinical use and application of progression algorithms and of applying “big data” results to individual practices.

## Précis

Six major visual field progression algorithms show poor to moderate agreement applied to a large multi-center dataset. Variability in visual field associated with worse mean deviation and older age increase the likelihood that visual field progression algorithms disagree.

## Introduction:

Despite advances in imaging, visual fields remain the gold standard to determine glaucoma progression and a cornerstone of glaucoma management.<sup>1</sup> On average, each glaucoma or glaucoma suspect patient has 1 to 2 visual fields per year depending on disease stage.<sup>2,3</sup> Visual fields are primarily used to determine the stage of glaucoma and to monitor for worsening over time. If a patient's disease is determined to be progressing based on visual field, the clinician typically opts to escalate therapy by altering the medication regimen, or intervening with laser or incisional surgery.

Strategies for determination of visual field progression, or worsening, vary and include event- and trend-based analyses.<sup>4</sup> Algorithms can assess progression using global indices (Mean Deviation [MD] and Visual Field Index [VFI]), on a point by point basis (Glaucoma Progression Analysis [GPA], Permutation of Pointwise Linear Regression [PoPLR]), or by analyzing clusters of points (Advanced Glaucoma Intervention Study [AGIS], Collaborative Initial Glaucoma Treatment Study [CIGTS]). While numerous algorithms have been developed to determine visual field worsening, only a handful are commonly used, and ultimate determination is subjective and made on a patient-by-patient basis.<sup>5</sup> These algorithms were generally developed for clinical trials with a specific study population in mind, strict inclusion and exclusion criteria, and ultimately may not be generally applicable.

This is one reason that visual field progression algorithms vary in their sensitivity and specificity and often have varying levels of agreement depending on which population they are applied to.<sup>6-9</sup> Assessment of agreement between different strategies and specific algorithms in a large real-world population is valuable to both clinicians and researchers, particularly in the interpretation and application of clinical trial results to an individual practice population. The purpose of this study was to determine the agreement and *discordance* (lack of agreement) between six visual field algorithms in a large collection of visual fields from five major academic institutions, and to evaluate risk factors for *discordance*.

## Methods:

### Data Source: Glaucoma Research Network Visual Field Database

We utilized the visual field database of the Glaucoma Research Network (GRN) which includes visual fields from the Wilmer Eye Institute (Baltimore, MD), Massachusetts Eye and Ear (Boston, MA), the Wills Eye Hospital (Philadelphia, PA), the Columbia University Medical Center (New York, NY), and the Bascom Palmer Eye Institute (Miami, Florida). This dataset has been used in prior work characterizing visual fields.<sup>10, 11</sup> This retrospective study was approved by the Institutional Review Boards of the participating institutions and adhered to the tenets of the Declaration of Helsinki. Identifying information from the visual fields was removed, but all other information from each test was retained. No clinical or diagnostic information was available for any of the subjects.

### Inclusion/Exclusion Criteria

Of the 831,240 visual fields from 177,172 patients in the parent dataset, we included visual fields that were SITA Standard 24-2 with a white size III stimulus on a white background. Only tests from patients older than 18 years were included. The same inclusion and exclusion criteria were applied for all algorithms to compare a broad sample of reliable visual fields that a practitioner would encounter. We excluded any tests with 20% or greater fixation losses, false positives greater than or equal to 15%, or where the false negative rate was NA. A false negative rating of NA indicates that there was an insufficient number of test points eligible for presentation of FN catch trials. This happens because either fewer than 6 false negative questions are asked or more than 7% of the test points have a threshold < 0 dB (Carl Zeiss Meditec, personal communication, November 26, 2018). We also excluded fields in which the Glaucoma Hemifield Test (GHT) noted “Abnormally High Sensitivity”, “General Reduction of Sensitivity” or “Borderline/General Reduction.” After these exclusion criteria were applied, only eyes with at least five eligible studies were included in the analysis. A subanalysis was completed with the above inclusion and exclusion criteria as well as including only patients with at least five years follow up and mean deviation greater than or equal to -12 decibels.

### Programming Existing Algorithms

We applied automated algorithms for six commonly used visual field algorithms: AGIS algorithm, CIGTS algorithm, Mean Deviation (MD) Slope, Visual Field Index (VFI) slope, pointwise linear regression (PLR), and PoPLR. Visual field algorithms were all implemented

in R (R Development Core Team Vienna, Austria). The VFI values, PLR algorithm, and PoPLR algorithm were obtained from an available open-source visual field package.<sup>12</sup> The AGIS and CIGTS algorithms were programmed specifically for this purpose using the methodology described in the literature.<sup>13,14</sup> Each automated algorithm was tested prior to its use by comparing its results to an expert's (OJS) implementation of each algorithm.

For both the MD and VFI, we calculated the slope of the value over time. If the slope was  $-1$  dB/year<sup>15</sup> or  $-1\%$  per year or worse<sup>16</sup>, we classified the patient as "progressing." Otherwise these cases were considered "stable". For both the AGIS and CIGTS scores, we counted the number of subsequent visual fields with scores at least 4 points greater than the score of the initial test (current score  $-$  initial score  $+ 4$ ). If we found that at least 3 subsequent scores were worse, then the eye was considered to be progressing. These three fields had to be consecutive, but if a field in the subsequent series had a score similar to baseline (i.e. within a delta of the baseline, with delta being 4 for AGIS and 3 for CIGTS) then the counter was reset to zero. Otherwise the eye was considered to be stable.<sup>13, 14</sup> For PLR, an eye was determined to be progressing if at least three points had a slope of  $-1$  ( $P < 0.01$ ) or worse.<sup>6</sup> For PoPLR, 5000 unique permutations were randomly selected. Simple linear regression was used to derive P value for change over time at individual locations.  $P < 0.05$  was the criteria for progression, otherwise the eye was determined to be stable.

## Analysis:

Basic demographics were calculated for each eye and each individual including: practice location, MD of first visual field, years of follow up, age at initial presentation, number of visual fields, right or left eye, and fields per year. We compared the rate of progression for all six algorithms individually as well as the rate of eyes progressing in four out of six algorithms (defined here as *majority progression*), five out of six algorithms, and six out of six algorithms. Agreement between each algorithm was determined using Cohen's kappa coefficient. Kappa values were classified as fair, good, or excellent as specified by Fleiss et al.<sup>17</sup>

To assess overall disagreement between all algorithms, we assessed overall *discordance* amongst algorithms, defined as when three algorithms determined that the set of visual fields was worsening and three algorithms determined that the fields were stable. We completed bivariate and multivariable analysis to determine predictors of discordance. Predictive variables assessed included: Institution, MD of first visual field, length of follow up, age at initial presentation, number of visual fields, right or left eye, and number of visual fields per year. Mean deviation was categorized into three categories based on the first three categories of the Hodapp-Anderson-Parrish criteria<sup>18</sup>: MD  $\geq -6$ ,  $-6 < \text{MD} < -12$ , and MD  $< -12$ . Years of follow up were categorized as  $\leq 5$  years,  $5 < \text{years} < 10$  years, and  $> 10$  years. Age was categorized as 18-40 years of age, 40-60 years, 60-80 years and  $> 80$  years. The number of visual fields was categorized as 5, 6-10 and  $> 10$ . *Discordance* was the outcome variable. We used the chi square statistic for bivariate analysis of categorical predictors. A generalized estimating equation using binary logistic regression was used for multivariable analysis and the predictors that were significant on bivariate analysis. Predictors with odds ratios greater than or equal to 1.5 were considered major predictors of *discordance*, as per Cohen's rules of

thumb. The rationale for this was to identify the associations with larger effect sizes, and to avoid weak associations with smaller effect sizes that are common in a large dataset such as this.<sup>19</sup> We then assessed significant differences in location by stratifying all variables by site.

We conducted a separate subanalysis to focus on patients with longer follow up and moderate visual field deficits, so as to avoid the potential difficulty in assessing visual field progression at worse MD values or shorter follow up times. The subanalysis repeated the same analyses described above, but excluded patients with less than 5 years of follow up and patients with mean deviation worse than  $-12$ . We modified categories for years of follow up ( $5 < \text{years} \leq 10$  years, and  $> 10$  years) and MD ( $\text{MD} \leq -6$ ,  $-12 < \text{MD} \leq -6$ ). SPSS (IBM, Chicago, IL) was used for statistical analysis.

## Results:

A total of 90,713 visual fields of 13,156 eyes of 8,499 patients met the inclusion criteria. The average age was  $67.1 \pm 12.3$  years (range 18 – 94 years). The data set included 6,479 (49.2%) right eyes and 6,677 (50.8%) left eyes. Overall, the average baseline MD was  $-4.9 \pm 5.8$  dB and average final MD was  $-6.3 \pm 6.8$  dB. The average baseline PSD was  $4.6 \pm 3.9$  and average final PSD was  $5.2 \pm 4.1$ . The average number of fields per eye was  $6.90 \pm 2.42$ . Annually,  $1.25 \pm 1.35$  fields per eye were performed. Table 1 shows the available demographics of this sample by eye and by subject.

Table 2 shows the number of eyes that were progressing using each algorithm and which were progressing using *majority progression* (progressing in at least four of six algorithms), progression in at least five of six algorithms, and the number progressing in all six algorithms. Nearly 50% of eyes progressed using PLR as compared to 5.6% of eyes that progressed using the AGIS criteria. Using majority progression (four of six algorithms), 11.7% of eyes progressed. Only 2.5% of eyes progressed by all six algorithms. *Discordance* was noted in 9.3% of eyes. Table 3 shows the agreement (kappa) between each algorithm. We found moderate agreement between MD and VFI, MD and AGIS, VFI and PLR, VFI and PoPLR, and AGIS and CIGTS. There was poor agreement between all other algorithms.

Bivariate analysis (Table 4) showed that institution, worse initial MD, greater follow up, older age, and a greater number of visual fields were predictors of *discordance*.

Multivariable analysis (Table 5) showed that the major predictors of *discordance* were MD worse than  $-6$  and age 60 and older at first visual field and lesser predictors were institution 4, greater than five years of follow up, and six visual fields or more. Institution 2 had a slightly lower rate of discordance.

There were significant differences in predictive variables for each institution. Institution 4 had the worst baseline mean deviation ( $-5.99 \pm 6.49$  dB,  $P < 0.01$ ) and the oldest age ( $65.4 \pm 11.6$  years,  $P < 0.01$ ). Institution 5 had the highest number of fields per year ( $1.4 \pm 0.9$ ,  $p < 0.01$ ). Institution 3 had the greatest length of follow up ( $6.8 \pm 2.7$  years,  $P < 0.01$ ), the highest average number of fields ( $7.4 \pm 2.8$ ,  $P < 0.01$ ), and the highest rate of *majority progression* (15.2%,  $P < 0.01$ ).

The subanalysis included 7,814 eyes of 5,239 patients. 948 (12.1%) eyes showed *majority progression* and 246 (3.1%) eyes progressed in all six algorithms. Discordance was noted in 654 (8.4%) eyes. Moderate agreement was noted between MD and AGIS, VFI and CIGTS, VFI and PLR, VFI and PoPLR, AGIS and CIGTS, and PLR and PoPLR. The remainder of comparisons showed poor agreement with one another. Major predictors of discordance MD worse than  $-6$  (OR: 3.19, 95% CI(2.68, 3.80)), age 60 - 80 (OR: 2.63, 95% CI(1.40, 4.92)) and age  $> 80$  (OR: 3.33, 95% CI(1.67, 6.63)), and more than 10 visual fields (OR: 1.61, 95% CI(1.20, 2.16)). Lesser predictors of discordance were Institution 3 (OR: 1.27, 95% CI(1.05, 1.54)) and Institution 4 (OR: 1.48, 95% CI(1.08, 2.03)) and 6-10 visual fields (OR: 1.30, 95% CI(1.04, 1.61)).

## Discussion:

Using a large dataset of visual fields from multiple institutions, we found a high degree of variation in the rate of progression using different algorithms, ranging from 5.8% using the AGIS criteria to 49.7% using PLR as we implemented it. Furthermore, algorithms showed poor-to-moderate agreement with each other when applied to the full dataset. We found that older age and worse initial mean deviation were strong significant predictors of *discordance*. The sub-analysis, which only included eyes with at least five years of follow up and MD of  $-12$  or better similarly found poor-to-moderate agreement between algorithms, a comparable rate of *majority progression* and *discordance*, and also found major predictors for progression to be MD worse than  $-6$  and age greater than 60. A notable difference was that greater than 10 visual fields was a major predictor of discordance.

The poor-to-moderate agreement between all algorithms potentially indicates the need for the development of a consensus algorithm, which we have attempted to do by defining *majority progression*. This allows for the use of multiple different approaches to determine progression with one metric. It also ensures that the criteria for progression is not overly stringent, as would be the case with setting the standard for progression at five out of six algorithms (5.9% rate of progression) or six out of six algorithms (2.5% rate of progression).

Prior studies comparing progression rates using different algorithms have also noted poor-to-moderate agreement between algorithms. Vesti *et al.* used a computerized model of visual field progression to compare CIGTS, PLR, and Glaucoma Change Probability (GCP) analysis and found that all three methods agreed in 22.4-35.5% depending on simulated variability of visual fields. Similar to our study, they also found that PLR had the highest rate of progression.<sup>6</sup> Artes *et al.* compared progression using MD to VFI and found moderate agreement ( $K = 0.69$ ) between the two indices in a large sample from a single institution.<sup>20</sup> While this was a higher interrater agreement than we found between these two groups, it is comparable using Fleiss's criteria.<sup>17</sup>

Major predictors of *discordance* between algorithms included worse initial MD, specifically worse than  $-6$  dB and age older than 60. This is consistent with prior work finding that reduction in sensitivity of visual field<sup>21</sup> and older age<sup>22</sup> are associated with greater variability in visual fields. The additional variability may explain *discordance*: why some visual field progression algorithms would determine that a given set of visual fields showed

evidence of progression and other algorithms would determine the same set of fields was stable. Greater number of visual fields was a major predictor of discordance in the subanalysis but not in the primary analysis. More follow up, and institution were also significantly associated with discordance between algorithms to a lesser degree. Patients with greater variability in visual fields likely had more tests to confirm visual field worsening, which may explain the association between discordance and a greater number of visual fields. Similarly, those with more advanced disease and associated variability in visual fields may be more likely to follow up long-term. The difference between institutions may be due to unmeasured confounders such as type of glaucoma, differences in demographics such as race, adherence to medication, or other practitioner or patient-specific factors. This difference between institutions also indicates the potential limitations in applicability of “big data” studies. A study such as ours with a large population and tests may not always apply well to a given local population.<sup>23-25</sup> Individual practices will need to evaluate their own local conditions, such as baseline glaucoma severity and patient age, before they implement results of such studies.

While this study included a large population of patients and multiple institutions, it incorporated only information available in the visual field test itself and did not include clinical information such as intraocular pressure, optic nerve assessment, biometric properties such as central corneal thickness or axial length, type of glaucoma, or history of ocular and systemic diseases. We included six major algorithms that utilized different strategies to determine visual field progression but could not include all available progression algorithms. The Glaucoma Progression Algorithm (GPA)<sup>26</sup> could not be included in this analysis as it is proprietary and the information was not included in the dataset. In order to reliably compare progression rates across all algorithms, we applied the same reliability criteria to all six algorithms. However, the AGIS and CIGTS algorithms had specific reliability criteria that were not used. While we have emphasized agreement for the purposes of determining visual field progression, stability of visual field is also important clinically. As our outcome measure is when 3 algorithms agreed and 3 disagreed (showing 3 stable and 3 progressing), our findings regarding discordance apply to both progression and stability. Finally, the current analysis reveals the general conditions when disagreement is high, but does not reveal the underlying cause as to why disagreement occurs.

Determination of visual field progression, while critical to glaucoma management, is challenging and often relies on subjective judgment. Available algorithms for visual field progression have poor to moderate agreement with one another when applied to a real-world dataset, particularly when applied to more advanced glaucoma. Algorithms are more likely to disagree when the initial mean deviation is worse or in older patients. Composite measures of progression, such as *majority progression* as we have defined it here may be one way to address this lack of consistency, but is not feasible with currently available clinical tools. New approaches to visual field interpretation such as the application of machine learning algorithms to large datasets of visual fields and other diagnostic modalities ultimately may help clinicians to assess progression of glaucoma.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Funding:

Dr. Saeedi is funded by an NIH Career Development Award (K23EY025014). Dr. Elze was funded by the BrightFocus Foundation, the Lions Foundation, the Grimshaw-Gudewicz Foundation, Research to Prevent Blindness and the Alice Adler Fellowship. Dr. De Moraes was funded by Research to Prevent Blindness and NIH/NEI grant R01 EY025253. Dr. Shen is supported by the Harvard Glaucoma Center of Excellence. Dr. Pasquale is supported by NIH/NEI R01 EY015473.

## References

1. Camp AS, Weinreb RN. Will Perimetry Be Performed to Monitor Glaucoma in 2025? *Ophthalmology*. 2017;124(12S):S71–S75. [PubMed: 28865878]
2. Traverso CE, Walt JG, Kelly SP, et al. Direct costs of glaucoma and severity of the disease: a multinational long term study of resource utilisation in Europe. *Br J Ophthalmol*. 2005;89(10):1245–9 [PubMed: 16170109]
3. Lee PP, Walt JG, Doyle JJ, et al. A multicenter, retrospective pilot study of resource use and costs associated with severity of disease in glaucoma. *Arch Ophthalmol*. 2006;124(1):12–9. [PubMed: 16401779]
4. Aref AA, Budenz DL. Detecting Visual Field Progression. *Ophthalmology*. 2017;124(12S):S51–S56. [PubMed: 29157362]
5. Tanna AP, Bandi JR, Budenz DL, et al. Interobserver agreement and intraobserver reproducibility of the subjective determination of glaucomatous visual field progression. *Ophthalmology*. 2011;118(1):60. [PubMed: 20723992]
6. Vesti E, Johnson CA, Chauhan BC. Comparison of different methods for detecting glaucomatous visual field progression. *Invest Ophthalmol Vis Sci*. 2003;44(9):3873–9. [PubMed: 12939303]
7. Heijl A, Bengtsson B, Chauhan BC, et al. A comparison of visual field progression criteria of 3 major glaucoma trials in early manifest glaucoma trial patients. *Ophthalmology*. 2008;115(9):1557–65. [PubMed: 18378317]
8. Birch MK, Wishart PK, O'Donnell NP. Determining progressive visual field loss in serial Humphrey visual fields. *Ophthalmology*. 1995;102(8):1227–34. [PubMed: 9097752]
9. Goldbaum MH, Sample PA, Chan K, et al. Comparing machine learning classifiers for diagnosing glaucoma from standard automated perimetry. *Invest Ophthalmol Vis Sci*. 2002; 43(1):162–9. [PubMed: 11773027]
10. Wang M, Pasquale LR, Shen LQ, et al. Reversal of Glaucoma Hemifield Test Results and Visual Field Features in Glaucoma. *Ophthalmology*. 2018 3;125(3):352–360. [PubMed: 29103791]
11. Wang M, Shen LQ, Boland MV, et al. Impact of Natural Blind Spot Location on Perimetry. *Sci Rep*. 2017 7 21;7(1):6143. [PubMed: 28733615]
12. Marín-Franch I, Swanson WH. The visualFields package: A tool for analysis and visualization of visual fields. *J Vision*. 2013;13(4):10. doi:10.1167/13.4.10.
13. AGIS investigators Advanced Glaucoma Intervention Study. 2. Visual field test scoring and reliability. *Ophthalmology*. 1994 8;101(8):1445–55. [PubMed: 7741836]
14. Gillespie BW, Musch DC, Guire KE, Mills RP, Lichter PR, Janz NK, Wren PA; 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 6;44(6):2613–20. [PubMed: 12766064]
15. Chan TCW, Bala C, Siu A, et al. Risk Factors for Rapid Glaucoma Disease Progression. *Am J Ophthalmol*. 2017 8;180:151–157. [PubMed: 28624324]



16. Rao HL, Kumbar T, Kumar AU, et al. Agreement between event-based and trend-based glaucoma progression analyses. *Eye (Lond)*. 2013 7;27(7):803–8. [PubMed: 23598668]
17. Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions In: Statistical Methods for Rates and Proportions*. 2nd ed. New York: Wiley; 1981:222–223.
18. Hodapp E, Parrish RK II, Anderson DR. *Clinical decisions in glaucoma*. St Louis: The CV Mosby Co; 1993 pp. 52–61.
19. Chen H, Cohen P, Chen S. How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. *Communications in Statistics: Simulation & Computation*. 4 2010, Vol. 39 Issue 4, p860–864.
20. Artes PH, O'Leary N, Hutchison DM, Heckler L, Sharpe GP, Nicolela MT, Chauhan BC. Properties of the statpac visual field index. *Invest Ophthalmol Vis Sci*. 2011 6 8;52(7):4030–8. [PubMed: 21467169]
21. Russell RA, Crabb DP, Malik R, Garway-Heath DF. The relationship between variability and sensitivity in large-scale longitudinal visual field data.
22. Katz J, Sommer A. A longitudinal study of the age-adjusted variability of automated visual fields. *Arch Ophthalmol*. 1987 8;105(8):1083–6. [PubMed: 3632417]
23. Shah ND, Steyerberg EW, Kent DM. Big Data and Predictive Analytics: Recalibrating Expectations. *JAMA*. 2018 7 3;320(1):27–28 [PubMed: 29813156]
24. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. *JAMA*. 2018 4 3;319(13):1317–1318 [PubMed: 29532063]
25. Gonzales JA, Lietman TM, Acharya NR. The Draw(back)s of Big Data. *JAMA Ophthalmol*. 2017 5 1;135(5):422–423. [PubMed: 28418450]
26. Heijl A, Leske MC, Bengtsson B, et al. Reduction of intraocular pressure and glaucoma progression: results from the Early Manifest Glaucoma Trial. *Arch Ophthalmol*. 2002 10;120(10):1268–79. [PubMed: 12365904]

**Table 1:**

## Characteristics of Included Eyes and Subjects

	<b>N(%)Eyes</b>	<b>N(%) Subjects</b>
<b>Institution</b>		
1	6232 (47.4%)	3934 (46.3%)
2	862 (6.6%)	612 (7.2%)
3	3760 (28.6%)	2395 (28.2%)
4	906 (6.9%)	613 (7.2%)
5	1396 (10.6%)	945 (11.1%)
<b>Eye</b>		
Right	6479 (49.2%)	
Left	6677 (50.8%)	
<b>Mean Deviation (MD)</b>		
-6	9282 (70.6%)	
-6 < MD < -12	2146 (16.3%)	
-12	1728 (13.1%)	
<b>Follow up (years)</b>		
5	4268 (32.4%)	2972 (35.0%)
5-10	7526 (57.2%)	4741 (55.8%)
>10	1362 (10.4%)	786 (9.2%)
<b>Age (years)</b>		
18 – 40	501 (3.8%)	326 (3.8%)
40 - 60	4110 (31.2%)	2567 (30.2%)
60 – 80	7589 (57.7%)	4927 (58.0%)
>80	956 (7.3%)	679 (8.0%)
<b>Number of Fields (<math>\mu</math> +/- <math>\sigma</math>)</b>	6.90 +/- 2.42	
<b>Fields per year (<math>\mu</math> +/- <math>\sigma</math>)</b>	1.25 +/- 1.35	
<b>Number of Fields</b>		
5	4669 (35.5%)	
6-10	7351 (55.9%)	
>10	1136 (8.6%)	
<b>Fields per year</b>		
1	5122 (38.9%)	
1-2	7109 (54.0%)	
>2	925 (7.0%)	

MD = Mean Deviation

**Table 2:**

## Rate of Progression for Each Algorithm

Algorithm	N (%)
MD	1188 (9.0%)
VFI	3514 (26.7%)
AGIS	767 (5.8%)
CIGTS	1358 (10.3%)
PLR	6538 (49.7%)
PoPLR	4245 (32.3%)
Majority Progression (at least 4 of 6 progressing)	1535 (11.7%)
0 of 6 progressing (6 stable)	5469 (41.5%)
1 of 6 progressing (5 stable)	3173 (24.1%)
2 of 6 progressing (4 stable)	1752 (13.3%)
3 of 6 progressing (3 stable)	1227 (9.3%)
4 of 6 progressing (2 stable)	756 (5.7%)
5 of 6 progressing (1 stable)	446 (3.3%)
6 of 6 progressing (0 stable)	333 (2.5%)

MD = Mean Deviation, VFI = Visual Field Index, AGIS = Advanced Glaucoma Intervention Study, CIGTS = Collaborative Initial Glaucoma Treatment Study, PLR = Pointwise Linear Regression, PoPLR = Permutation of Pointwise Linear Regression

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3:**

## Agreement of Individual Progression Algorithms

	MD	VFI	AGIS	CIGTS	PLR
MD					
VFI	<b>0.42</b>				
AGIS	<b>0.42</b>	0.27			
CIGTS	0.37	0.33	<b>0.51</b>		
PLR	0.18	<b>0.52</b>	0.12	0.18	
POPLR	0.25	<b>0.42</b>	0.19	0.31	0.32

Bolded ICC values are those of moderate agreement as per Fleiss's criteria

MD = Mean Deviation, VFI = Visual Field Index, AGIS = Advanced Glaucoma Intervention Study, CIGTS = Collaborative Initial Glaucoma Treatment Study, PLR = Pointwise Linear Regression, PoPLR = Permutation of Pointwise Linear Regression

**Table 4:**

Predictors of Discordance of Algorithms – Bivariate Analysis

Variable	VF either progressing or stable in 4 of 6 algorithms (n = 11929)	VF progressing in 3 algorithms and stable in 3 algorithms (n = 1227)	P-value
<b>Institution</b>			
1	5668 (90.9%)	564 (9.1%)	P < 0.01
2	815 (94.5%)	47 (5.5%)	
3	3369 (89.6%)	391 (10.4%)	
4	795 (87.7%)	111 (12.3%)	
5	1282 (91.8%)	114 (8.2%)	
<b>Eye</b>			
Right	5903 (91.1%)	576 (8.9%)	0.09
Left	6026 (90.3%)	651 (9.7%)	
<b>Initial Mean Deviation (mean +/- SD)</b>	-4.5 +/- 5.7	-8.0 +/- 6.2	P < 0.01
<b>Initial Mean Deviation (categorical)</b>			
-6	8713 (93.9%)	569 (6.1%)	P < 0.01
-6 < MD < -12	1800 (83.9%)	346 (16.1%)	
-12	1416 (81.9%)	312 (18.1%)	
<b>Follow up (years)</b>	6.3 +/- 2.5	6.6 +/- 2.5	P < 0.01
<b>Follow up (years - categorical)</b>			
5	3910 (91.6%)	358 (8.4%)	P = 0.03
6-10	6797 (90.3%)	729 (9.7%)	
>10	1222 (89.7%)	140 (10.3%)	
<b>Age (years)</b>	63.4 +/- 12.2	68.1 +/- 11.3	P < 0.01
<b>Age (years - categorical)</b>			
18 – 40	476 (95.0%)	25 (5.0%)	P < 0.01
40 – 60	3854 (93.8%)	256 (6.2%)	
60 – 80	6785 (89.4%)	804 (10.6%)	
>80	814 (85.1%)	142 (14.9%)	
<b>Number of Fields (mean +/- SD)</b>	6.9 +/- 2.4	7.3 +/- 2.7	P < 0.01
<b>Fields per year (mean +/- SD)</b>	1.2 +/- 1.2	1.3 +/- 2.2	P = 0.11
<b>Number of Fields (categorical)</b>			
5	4301 (92.1%)	368 (7.9%)	P < 0.01
6-10	6634 (90.2%)	717 (9.8%)	
>10	994 (87.5%)	142 (12.5%)	
<b>Fields per year (categorical)</b>			
1	4662 (91.0%)	460 (9.0%)	P = 0.53
1-2	6433 (90.5%)	676 (9.5%)	

Variable	VF either progressing or stable in 4 of 6 algorithms (n = 11929)	VF progressing in 3 algorithms and stable in 3 algorithms (n = 1227)	P-value
>2	834 (90.2%)	91 (9.8%)	

VF = Visual Field, SD = Standard Deviation

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 5:**Predictors of *Discordance* of Visual Field Progression Algorithms - Multivariable Analysis

Variable	Odds ratio (Exp (B))	95% CI	
		Lower	Upper
<b>Location</b>			
1	REF		
2 *	0.72	0.53	0.99
3 *	1.16	1.01	1.34
4 *	1.33	1.06	1.68
5	1.00	0.80	1.25
<b>Initial Mean Deviation</b>			
MD -6	REF		
-6 < MD < -12 *	2.75	2.37	3.19
MD -12 *	3.20	2.75	3.73
<b>Follow up (years)</b>			
5	REF		
5-10 *	1.17	1.01	1.35
>10 *	1.32	1.04	1.67
<b>Age (years)</b>			
18 - 40	REF		
40 - 60	1.30	0.83	2.02
60 - 80 *	2.15	1.40	3.31
>80 *	2.77	1.73	4.42
<b>Number of Fields</b>			
5	REF		
6-10 *	1.17	1.01	1.34
>10 *	1.47	1.17	1.85

\* indicates significance to  $P < 0.05$