



# City Research Online

## City, University of London Institutional Repository

---

**Citation:** Garway-Heath, D. F., Quartilho, A., Prah, P., Crabb, D. P. ORCID: 0000-0001-8754-3902, Cheng, Q. and Zhu, H. (2017). Evaluation of Visual Field and Imaging Outcomes for Glaucoma Clinical Trials (An American Ophthalmological Society Thesis). Transactions of the American Ophthalmological Society, 115, T4.

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

---

**Permanent repository link:** <http://openaccess.city.ac.uk/20642/>

**Link to published version:**

**Copyright and reuse:** City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

---

City Research Online:

<http://openaccess.city.ac.uk/>

[publications@city.ac.uk](mailto:publications@city.ac.uk)

---

# Evaluation of Visual Field and Imaging Outcomes for Glaucoma Clinical Trials

## Authors:

David F Garway-Heath, BSc, MB BS, MD, FRCOphth<sup>1</sup>

Ana Quartilho, MSc<sup>1</sup>

Philip Prah, MSc<sup>1</sup>

David P Crabb, PhD<sup>2</sup>

Qian Cheng, BSc<sup>3</sup>

Haogang Zhu, MSc, PhD<sup>1,3</sup>

## Affiliations:

1. NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK.
2. Division of Optometry and Visual Science, School of Health Sciences, City, University of London, UK
3. School of Computer Science and Engineering, Beihang University, Beijing, China

## Corresponding author:

David F Garway-Heath, UCL Institute of Ophthalmology, 11-43 Bath Street, London, EC1V 9EL, UK

Telephone: +44 20 7608 6800

E-mail: david.garway-heath@ Moorfields.nhs.uk

22	<b>TABLE OF CONTENTS</b>	
23	Abstract.....	3
24	Introduction .....	4
25	Methods .....	6
26	Data sources.....	6
27	UKGTS data set.....	6
28	RAPID data set .....	6
29	Participant demographics.....	7
30	Visual field testing .....	7
31	Optical Coherence Tomography imaging.....	7
32	Data analysis methods .....	7
33	Growth curve models.....	8
34	Association of RNFLT change with VF survival .....	8
35	Evaluation of 3 statistical models .....	8
36	Survival analyses .....	9
37	Sample size calculations .....	9
38	Results .....	11
39	Growth curve model .....	11
40	Visual field analysis.....	11
41	OCT analysis .....	11
42	Association of RNFLT change with VF survival .....	11
43	Evaluation of 3 statistical models .....	11
44	Survival analyses .....	11
45	Sample size calculations .....	12
46	Discussion.....	14
47	Limitations and further work .....	16
48	References .....	17
49	Acknowledgements .....	22
50	Tables .....	23
51	Figures and legends .....	30
52		
53		
54		

55 **ABSTRACT**

56

57

58 Purpose: to evaluate the ability of various visual field (VF) analysis methods to discriminate treatment groups in  
59 glaucoma clinical trials and establish the value of optical coherence tomography (OCT) imaging as an additional  
60 outcome.

61 Methods: VFs and retinal nerve fibre layer thickness (RNFLT) measurements (acquired by time-domain OCT)  
62 from 373 glaucoma patients in the UK Glaucoma Treatment Study (UKGTS) at up to 11 scheduled visits over a  
63 2 year interval formed the cohort to assess the sensitivity of progression analysis methods. Specificity was  
64 assessed in 78 glaucoma patients with up to 11 repeated VF and OCT RNFLT measurements over a 3 month  
65 interval. Growth curve models assessed the difference in VF and RNFLT rate of change between treatment  
66 groups. Incident progression was identified by 3 VF-based methods: Guided Progression Analysis (GPA),  
67 'ANSWERS' and 'PoPLR', and one based on VFs and RNFLT: 'sANSWERS'. Sensitivity, specificity and  
68 discrimination between treatment groups was evaluated.

69 Results: the rate of VF change was significantly faster in the placebo, compared to active treatment, group (-  
70 0.29 vs +0.03 dB/year,  $P < .001$ ); the rate of RNFLT change was not different (-1.7 vs -1.1 dB/year,  $P = .14$ ).

71 After 18 months and at 95% specificity, the sensitivity of ANSWERS and PoPLR was similar (35%);  
72 sANSWERS achieved a sensitivity of 70%. GPA, ANSWERS and PoPLR discriminated treatment groups with  
73 similar statistical significance; sANSWERS did not discriminate treatment groups.

74 Conclusions: although the VF progression-detection method including VF and RNFLT measurements is more  
75 sensitive, it does not improve discrimination between treatment arms.

76

77

78

79 **INTRODUCTION**

80 There has been considerable interest over the last decade in improving the design of clinical trials for glaucoma  
81 interventions and, in particular, assessing the potential for imaging measurements of optic nerve structure to be  
82 surrogate outcomes for clinical trials. This is motivated by the perception that that visual field (VF)  
83 measurements of optic nerve function are too insensitive or imprecise, or both, to be able to measure treatment  
84 effects in clinical trials over a short duration.

85 Visual field loss deterioration is a recognised outcome for glaucoma clinical trials,<sup>1</sup> however, VF measurements  
86 are variable and the variability becomes greater as the VF deteriorates.<sup>2-4</sup> Mitigation of the effects of variability,  
87 to accurately detect true disease deterioration ('progression'), requires frequent VF testing and/or a long period  
88 of time.<sup>5,6</sup> In clinical trials with a VF outcome, variability results in the requirement for large numbers of  
89 patients over long observation periods. Historically, the observation periods for trials with a VF outcome have  
90 been 4 years or longer,<sup>7-10</sup> with the shortest being 30 months,<sup>11</sup> until the recently-reported United Kingdom  
91 Glaucoma Treatment Study (UKGTS).<sup>12</sup> The UKGTS was designed with more frequent VF testing, and with  
92 short between-test intervals at the baseline, 18-month and 24-month visits ('clustering'),<sup>13</sup> to establish whether  
93 frequent and clustered tests enable shorter observation periods. The primary outcome analysis was for a  
94 difference in time to a VF progression event at the 24-month follow-up time point between latanoprost-treated  
95 and placebo treated participants. A highly statistically significant difference was evident at 24 months ( $P=.0003$ )  
96 and the difference was even significant at 12 months ( $P=.035$ ).

97 The UKGTS was also designed to enable the evaluation of optic nerve imaging measurements as potential  
98 clinical trial outcomes (VF surrogates), using imaging devices available at the initiation of the trial: scanning  
99 laser ophthalmoscopy,<sup>14,15</sup> scanning laser polarimetry<sup>16</sup> and time-domain (TD) optical coherence tomography  
100 (OCT).<sup>17</sup> For a surrogate, or biomarker, to be suitable as an alternative outcome, it must be strongly associated  
101 with the outcome of greatest relevance to the patient – in the case of glaucoma, this is visual function. The  
102 accepted measure of glaucomatous damage to visual function is standard automated perimetry (SAP),  
103 colloquially, the VF test. Candidates as surrogate outcomes include intraocular pressure (IOP) and  
104 measurements of optic nerve structure derived from ocular imaging.

105 The effect of therapeutic interventions on the IOP has long been used as an outcome in clinical trials of  
106 glaucoma treatments. However, whilst the association between the level of IOP and rate of glaucoma  
107 deterioration is statistically highly significant, IOP is a poor predictor of deterioration because many other  
108 ('non-IOP') factors affect glaucoma susceptibility so that patients deteriorate at all levels of IOP.<sup>18</sup> Furthermore,  
109 IOP is unsuitable as an outcome of a disease-modifying treatment which has no effect on IOP (so-called  
110 'neuroprotective' treatments).

111 The rationale for the use of imaging outcomes as surrogates for VF loss is more obvious. The loss of vision in  
112 glaucoma is a consequence of damage to, and death of, retinal ganglion cells (RGCs). The quantitative and  
113 spatial relationship between image-based measurements of the neural rim at the ONH and RNFL loss and VF  
114 damage is well-recognised<sup>19-25</sup> and imaging-based quantitative measurements have diagnostic utility.<sup>26-32</sup>  
115 Numerous publications support the ability of imaging-based measurements to identify glaucoma  
116 deterioration<sup>14,33-42</sup> and progressive structural change has been shown to be useful as a predictor of subsequent  
117 VF loss.<sup>43,44</sup>

118 The ability of imaging to detect progression has been compared to that of VF testing, controlling for the false-  
119 positive rate of the chosen progression criteria; with criteria matched for specificity, studies have found similar  
120 detection sensitivity for imaging compared to VF testing.<sup>14,36</sup> However, agreement on the eyes demonstrating  
121 glaucomatous progression was poor (for the most part, different eyes were identified as progressing by structure  
122 and function). Measurement variability prevents deterioration from being identified in a proportion of eyes.  
123 Because the source of measurement variability is different in VF testing and imaging, the eyes in which  
124 deterioration is missed are different for the two techniques. It makes sense, therefore, to make use of imaging  
125 data to compensate for the failure of VF testing to identify some of the deteriorating eyes.

126 At present, regulatory authorities recognise VF test outcomes for trials evaluating therapeutic interventions for  
127 glaucoma, but not yet structural outcomes based on imaging.<sup>1,45</sup> Surrogate outcomes, such as structural  
128 measurements based on imaging, need to be strongly correlated with the clinically relevant outcome, in this case  
129 VF loss, and capture the effect of a treatment intervention on that clinically relevant outcome.<sup>46,47</sup> The  
130 correlation between structural and VF measurements has been established<sup>22,23,43,44</sup> and the potential for structural  
131 measurements (scanning laser ophthalmoscopy measurements of the ONH) to capture treatment effects has been  
132 demonstrated.<sup>48</sup> However, no clinical trial data demonstrating that structural outcomes capture treatments effects  
133 on the VF have been published.

134 Making use of imaging measurements does not necessarily require that the measurements be used directly as a  
135 surrogate outcome, as an alternative to VF deterioration. Instead, the imaging measurements can be combined in  
136 Bayesian statistical models with VF data, to provide a background (prior) probability that the visual function of  
137 an eye might be deteriorating. This allows the additional information on the deterioration status of the eye

138 provided from imaging to be utilized, but VF loss remains the primary outcome. Establishing whether a new  
139 model of deterioration better describes the true underlying disease behaviour is not straight-forward, because  
140 there is no external ‘gold standard’ measurement of glaucoma deterioration. An approach to evaluate a model is  
141 to apply it to initial data in a series and use it to predict observed data later in the series;<sup>49-53</sup> the model with  
142 smaller prediction errors can be assumed to be a better representation of the underlying data than the model with  
143 greater prediction errors. Russell demonstrated that the prediction of future visual function states, based on  
144 linear regression of observed VF series, improved when the analysis included the rate of neural rim loss,  
145 measured with the scanning laser ophthalmoscope, as a Bayesian prior.<sup>54</sup> Applying a different statistical  
146 approach, Medeiros also used a Bayesian method to jointly model structural and functional progression and  
147 found that prediction accuracy was greater when structural data were included.<sup>51</sup> Other methods to combine  
148 imaging and VF data are emerging in the literature.<sup>55-57</sup>

149 Validation of any approach to identify glaucoma deterioration is challenging because, as mentioned, there is no  
150 ‘gold standard’ arbiter of the ‘truth’. Various methods have been used in the past to compare different  
151 approaches, all of which make certain assumptions. A general method is to match the false positive frequency  
152 for criteria so that technologies/approaches being compared have similar criterion specificity; it is then assumed  
153 that the technology with the higher ‘hit’ frequency (identified deterioration) is the more sensitive. An indicator  
154 of a test criterion false positive frequency is the number of eyes with stable glaucoma which are flagged as  
155 deteriorating. Defining ‘stable glaucoma’ with a progression criterion becomes a circular argument, so typically  
156 patient cohorts are selected which are at low risk for progression and tested sequentially over a sufficiently short  
157 period of time that measureable change would not occur.<sup>58,59</sup> The main assumption with this approach is that the  
158 variability characteristics for the tests are the same over the short period as they would be over typical clinical  
159 time scales.

160 The variability in VF measurements is well known and often regarded as a consequence of the subjective,  
161 psychophysical nature of the test. On the other hand, imaging devices are regarded as acquiring measurements  
162 objectively, with an expectation that measurement variability would be low. There is, however, appreciable  
163 imprecision in structural measurements. A discernible change in RNFL thickness can be described by ‘tolerance  
164 limits’ for test retest variability ( $1.645 \times \sqrt{2} \times \text{test retest standard deviation}$ ).<sup>60</sup> For a widely-used commercial  
165 spectral-domain OCT, the Cirrus OCT, the tolerance limit for average RNFL thickness measurement is  $3.9\mu\text{m}$ .  
166 The dynamic range of RNFL thickness measurements varies between commercial devices; for the Cirrus OCT, a  
167 value of  $35.5\mu\text{m}$  has been reported.<sup>61</sup> The number of steps of discernible change across the dynamic range is,  
168 therefore, about 9. Measurement imprecision is greater for TD OCT, with tolerance limits reported of between  
169  $6.4$  to  $8\mu\text{m}$ .<sup>62</sup> It is, therefore, by no means clear that imaging provides a more precise estimate of glaucoma  
170 deterioration than VF testing. A recent study showed that deterioration may be identified by either VF testing or  
171 OCT imaging across the spectrum of glaucoma severity, but estimated that deterioration is more likely to be  
172 identified with spectral-domain OCT imaging of the RNFL than VF testing in the earlier stages of glaucoma (up  
173 to around a VF mean deviation [MD] of  $-10\text{dB}$ ) and is more likely with VF testing in the later stages of  
174 glaucoma.<sup>42</sup>

175 The purpose of this study was to evaluate various statistical methods to identify VF deterioration and to  
176 establish whether progression models which include TD OCT measurements of the RNFL are more sensitive in  
177 identifying deterioration and enable better discrimination between treatment arms of a clinical trial.

178 The analyses were undertaken in the UKGTS data sets.<sup>12</sup>

179 Specifically, in evaluating the TD OCT data, we ask the following questions:

- 180 1. Does the rate of RNFL loss differ in the two treatment arms of the UKGTS?
- 181 2. Is the rate of RNFL loss a significant predictor of VF loss in the UKGTS?
- 182 3. Does a composite RNFL/VF outcome provide:
  - 183 a. more sensitive identification of progression?
  - 184 b. more accurate predictions of future VF loss?
  - 185 c. better discrimination between the treatment arms of the trial?

186 The main hypothesis being tested is whether a composite RNFL/VF outcome provides better discrimination  
187 between the treatment arms of a clinical trial of IOP-lowering medication. For reference, we provide sample size  
188 calculations for various clinical trial scenarios based of the analysis providing the best separation between  
189 treatment groups.

190

191 **METHODS**

192

193 **DATA SOURCES**

194 Two data sources were employed. One was a data set from the UKGTS placebo-controlled clinical trial,<sup>12</sup> in  
195 which with VF and OCT imaging data were acquired over an observation period of up to 2 years; OCT imaging  
196 was undertaken on participants from seven of the 10 study sites. This is termed the ‘UKGTS data set’. The  
197 second data set was a test retest data set of glaucoma patients attending a single study site with up to 11 VFs and  
198 OCT images acquired within a 3-month interval. This is termed the ‘RAPID data set’.

199

200

**UKGTS data set**

201 The UKGTS design, participant characteristics and main outcomes are described in detail elsewhere.<sup>12,63,64</sup> The  
202 UKGTS was a multicentre randomized controlled trial conducted at ten centres across the UK. Centres were  
203 district general hospitals, teaching hospitals and tertiary referral centres. The UKGTS was an RCT that  
204 compared the effects of latanoprost, a topical treatment to lower IOP, with placebo on survival from VF  
205 deterioration. 516 patients with newly diagnosed open-angle glaucoma were enrolled, with 777 eyes eligible for  
206 entry into the study.

207 Patients were followed up every 2-3 months after eye drop therapy was initiated, for up to 11 scheduled visits  
208 (Table 1). Participants attended for additional visits, at which VF testing and imaging were repeated, if tentative  
209 VF deterioration was identified according to certain pre-set criteria. Visual function was monitored by VF  
210 testing (detailed below) and ONH structured was monitored with the Heidelberg retina tomograph at all study  
211 locations and with the Stratus OCT (detailed below) and GDxECC Nerve Fiber Analyzer at locations with those  
212 devices. The subset of UKGTS participants with both VF testing and OCT imaging was used in this work.

213 The primary outcome for the trial was glaucomatous VF deterioration (progression) within 24 months. Details  
214 of the method for determining progression in the VFs has been published.<sup>12,63</sup> Progression analysis was  
215 performed in the Humphrey Field Analyzer II-i Guided Progression Analysis (GPA) software. The criterion for  
216 tentative progression was three locations worse than baseline in two consecutive VFs (3 half-shaded locations  
217 [up to two of which could be fully-shaded]). If tentative deterioration was identified, participants returned for  
218 confirmation tests within 1 month. At this confirmation visit, 2 VF tests were performed; if the same criterion of  
219 three half-shaded (or full-shaded) locations was satisfied in these confirmation tests, then the patient was  
220 considered to have progressed. Patients deemed to have progressed left the trial and treatment was adjusted as  
221 deemed appropriate by the treating clinician. Patients leaving the trial were invited to an ‘exit visit’ before  
222 treatment adjustment. If a patient was found to not be progressing at the confirmation visit, then (s)he returned  
223 to the standard visit schedule (Table 1).

224 The study was undertaken in accordance with good clinical practice guidelines and adhered to the Declaration of  
225 Helsinki. The trial was approved by the Moorfields and Whittington Research Ethics Committee on June 1,  
226 2006 (reference 09/H0721/56). All patients provided written informed consent before screening investigations.  
227 An independent Data and Safety Monitoring Committee (DSMC) was appointed by the trial steering committee.  
228 The trial manager monitored adverse events, which were reported immediately to the operational DSMC at  
229 Moorfields Eye Hospital. Serious adverse events were reported to the Medicines and Healthcare Products  
230 Regulatory Agency. This trial registration number is ISRCTN96423140.

231

232

**RAPID data set**

233 The Rapid data set was acquired from volunteer patients attending the glaucoma clinics at Moorfields Eye  
234 Hospital NHS Foundation Trust, which functions as a district general and teaching hospital and a tertiary  
235 referral centre; VF testing and imaging was undertaken in the National Institute for Health Research Clinical  
236 Research Facility.

237 The study ‘Assessing the effectiveness of imaging technology to rapidly detect disease progression in glaucoma:  
238 ‘stable data’ collection’ was undertaken in accordance with good clinical practice guidelines and adhered to the  
239 Declaration of Helsinki. The trial was approved by the North of Scotland National Research Ethics Service  
240 committee on September 27, 2013 (reference 13/NS/0132) and NHS Permissions for Research was granted by  
241 the Joint Research Office at University College Hospitals NHS Foundation Trust on December 3, 2013. All  
242 patients provided written informed consent before screening investigations.

243 The recruitment criteria for the ‘Stable Glaucoma’ Cohort were similar to those of the UKGTS clinical trial and  
244 the number of repeat tests approximated the number acquired during the UKGTS.

245 Inclusion Criteria:

- 246 • Open angle glaucoma (OAG; including primary OAG, normal tension glaucoma and pseudoexfoliation  
247 glaucoma) in either eye according to the definition for entry to the UKGTS.<sup>63</sup>

- 248
- Age over 18 years
- 249
- Snellen visual acuity equal to or better than 6/12
- 250
- Able to give informed consent and attend at the required frequency for the duration of the study.

251

252 Exclusion criteria:

- 253
- Visual field loss worse than -16 dB or paracentral points with sensitivity < 10dB in both the upper and lower hemifields in either eye
- 254
- IOP > 30mmHg in either eye
- 255
- Unable to perform reliable visual field testing (false positive rate > 15%)
- 256
- Poor quality OCT (quality score < 15 for FD-OCT and < 7 for SD-OCT)
- 257
- Refractive error outside the range - 8 to +8 diopters
- 258
- Previous intraocular surgery (other than uncomplicated cataract extraction with posterior chamber lens implantation or uncomplicated Trabeculectomy)
- 259
- Cataract extraction with posterior chamber lens implantation within the last year
- 260
- Diabetic retinopathy
- 261
- 262
- 263

264 Study schedule: participants attended approximately once a week and underwent VF testing and TD OCT  
265 imaging as outlined below. Two sets of tests from each device were acquired at the first visit and one from each  
266 at subsequent visits to give a total of 11 tests for each device, in total. In addition to the VF tests and TD OCT  
267 imaging, participants were also imaged with the Spectralis OCT (Heidelberg Engineering, Heidelberg,  
268 Germany) and the DRI OCT-1 Atlantis (Topcon, Japan).

269 The sample size for the ‘specificity’ data set was determined as a pragmatic solution to balance precision of  
270 estimates and feasibility. A sample of 80 subjects was deemed sufficient to approximate between individual  
271 differences in test-retest variability.

272

## 273 PARTICIPANT DEMOGRAPHICS

274 Table 2 gives the principal demographic data for the subset of UKGTS participants with OCT images.<sup>63</sup> The  
275 participant characteristics in the subset of UKGTS patients with OCT images are very similar to those of the full  
276 UKGTS data set.

277 The principal demographic data for participants in the RAPID test retest study are given in Table 3. The data are  
278 similar; RAPID participants have slightly more advanced glaucoma (VF MD -4.17 compared to -2.65 dB) and  
279 lower IOP (14.0 compared to 19.0 mmHg); there was a lower proportion of white participants in the RAPID  
280 study (67% compared to 88%).

281

282

### 283 Visual field testing

284 SAP visual fields were tested with the Swedish interactive threshold algorithm (SITA) standard 24-2 program  
285 (Humphrey Field Analyzer, HFA; Carl Zeiss Meditec, Dublin CA). Reliable VF tests were included (<15% false  
286 positives and <20% fixation losses). Unreliable tests were repeated on the same day (with a break of at least 30  
287 minutes). All patients had undergone a minimum of two visual field tests before the study started. At the first  
288 visit, patients underwent 2 VF tests and the mean of these was used as the baseline in the GPA analysis; if the  
289 GPA software rejected a baseline VF on the basis of ‘learning’, the next VF in the series was used as a baseline.  
290 VFs rejected by the GPA software were not included in the analyses by other methods.

291 A glaucomatous VF defect, for study inclusion, was defined as a reproducible (in at least 2 consecutive reliable  
292 VFs) reduction in sensitivity at two or more contiguous points with  $P < .01$  loss or greater, or three or more  
293 contiguous points with  $P < .05$  loss or greater, or a 10-dB difference across the nasal horizontal midline at two or  
294 more adjacent points in the total deviation plot. A reliable VF is one with <15% false positives.

295

### 296 Optical Coherence Tomography imaging

297 OCT imaging was performed through dilated pupils with the Stratus OCT (software version 5.0; Carl Zeiss  
298 Meditec) using the ‘landmark’ function. Each patient underwent RNFL scanning with the fast RNFL (3.4mm;  
299 256 A-scans) protocol. The average RNFLT was used for this analysis.

300

301

## 302 DATA ANALYSIS METHODS



303

### **Growth curve models**

304 The aim of this analysis is to identify whether the rate of progression (slope), based on MD or mean RNFLT  
305 values over time, is different between the latanoprost and placebo groups.

306 Subject selection: This analysis considered the subset of UKGTS participants who had OCT imaging available.  
307 If both eyes had glaucoma at baseline (eligible for inclusion in the main UKGTS study), the eye with worse  
308 baseline VF MD was selected for analysis, as determined by the UKGTS statistical analysis plan. Data were  
309 included provided the tests met predetermined quality criteria (VF <15% false positive responses or  
310 measurements outside the range +4 to -30dB; OCT quality score  $\geq 7$ , absence of an image warning message or  
311 measurements outside the range 20 to 135 microns RNFLT). Figure 1 details the selection flow chart for the  
312 analysis. The OCT data set comprises 284 participants; 3 of these did not qualify for the VF analysis, so that the  
313 VF data set comprised 281 participants.

314 A growth curve model is a type of multilevel random slope model where the predictor of interest is a  
315 measurement of time. When data are longitudinal and measurements are repeated within patients, time is used as  
316 an explanatory variable to describe the rate of change in the outcome. Longitudinal models were used in  
317 UKGTS to compare whether the rates of change in a particular outcome differ by intervention group. Thus  
318 interaction terms were used to estimate whether the rates are significantly different. Details of the model are  
319 given in the appendix.

320

321 In addition to the growth curve models, the raw rates of change were plotted to allow assessment of the  
322 distribution of rates of measurement change of the two treatment groups. A crude analysis comparing the VF  
323 MD and OCT RNFLT slope for each participant across treatments groups was made (Mann-Whitney test for  
324 independent samples); this does not take account of the variance in the individual slope estimates.

325

326

### **Association of RNFLT change with VF survival**

327 Progression-free survival was assessed with a Kaplan-Meier survival analysis to illustrate the frequency of  
328 progression and the difference between treatment groups. The progression criterion applied was the GPA  
329 criterion used in the UKGTS outcome report; the participants analysed are the sub-set with OCT images. To  
330 identify whether the rate of OCT RNFLT change was associated with VF progression, a Cox proportional  
331 hazards model was fitted to the data with factors potentially associated with survival failure (treatment  
332 allocation, age, baseline IOP, baseline VF MD and the slope of RNFLT change). Calculations were performed  
333 with MedCalc Statistical Software version 17.1 (MedCalc Software bvba, Ostend, Belgium;  
334 <https://www.medcalc.org>; 2017)

335

336

### **Evaluation of 3 statistical models**

337

338 Progression detection sensitivity

339 The purpose in this section was to evaluate the relative sensitivity of three methods for identifying progression.  
340 These methods were: analysis with non-stationary Weibull error regression and Spatial Enhancement  
341 (ANSWERS),<sup>53,65</sup> permutation analyses of pointwise linear regression (PoPLR)<sup>66</sup> and a modification of  
342 ANSWERS to incorporate the RNFLT slope as a prior: structure-guided ANSWERS (sANSWERS).

343 Subject selection: in this section, 445 eyes of 353 UKGTS participants with at least three follow-up visits and  
344 available OCT images, irrespective of image quality, were included. 107 eyes of 70 RAPID participants with 10  
345 or more VF tests and OCT images were included.

346 ANSWERS: this method is a linear regression technique which formally takes into account the increasing  
347 variability of VF sensitivity estimates as sensitivity declines. It also takes into account the spatial correlation  
348 between sensitivity values at each location within a VF. Application of ordinary least squares linear regression  
349 (OLSLR) makes the assumption that the residuals from the regression are normally distributed. In reality, there  
350 is heteroscedasticity, with more dispersed residuals as sensitivity declines. ANSWERS models this  
351 heteroscedasticity with a mixture of Weibull distributions. Spatial correlation of measurements is also included  
352 into the model using a Bayesian framework. We have previously shown that this technique is more sensitive in  
353 identifying VF progression, and provides more accurate predictions of future VF states, than OLSLR of MD  
354 over time and PoPLR.<sup>53</sup>

355 PoPLR: this is a non-parametric approach based on randomly permuting the observed VF series to identify  
356 whether negative change identified in the observed (un-permuted) series is significant, based on the distribution  
357 of change identified in the permuted series. The slope of VF sensitivity change is determined by OLSLR and the  
358 statistical significance (*P* value) from each location across the VF is combined into a statistic 'S' by using the

359 Truncated Product Method. The statistical significance of  $S$  in the observed series is calculated by comparing it  
360 with a null distribution of  $S$ , derived from permuted sequences of the series.

361 sANSWERS: this method is a modification of ANSWERS in which there is a 2-layered hierarchical Bayesian  
362 model; the prior distribution of the VF progression rate at each VF location is set by the slopes and variance of  
363 the rate of change in the RNFLT; this is similar to the approach described previously to incorporate scanning  
364 laser ophthalmoscope rim area measurement slopes into VF progression analysis.<sup>54</sup> As the spatial  
365 correspondence of peripapillary circle sectors and VF locations is known,<sup>25</sup> each VF location was mapped to one  
366 of 12 peripapillary RNFLT sector measurements; the slope and variance of RNFLT over time formed the  
367 Bayesian prior for the VF slope.

368

369 The specificity of various criteria to ‘call’ progression was evaluated in the RAPID test retest data set and the  
370 ‘hit’ rate (a surrogate for criterion ‘sensitivity’ which includes true change and the false positive change allowed  
371 by the criterion specificity) was determined from the UKGTS data set for each criterion evaluated.

372 Criterion specificity was determined for the seven, 13, 18 and 22 month time point. When data were permuted,  
373 the VF tests and OCT images for the same day were tied (permuted together); when there was no OCT image  
374 associated with a VF test, the VF was permuted alone. 100 permutations were performed for each eye and each  
375 time point. The test schedule of the UKGTS was mimicked (Table 1), so that 2 VF tests and equivalent OCT  
376 RNFLT measurements were taken at visits 1, 2, 7, and 8 and the time interval between tests was assumed to be  
377 as for the UKGTS schedule. In this analysis, the RAPID data series comprise series lengths between 10 and 14  
378 tests. The 18 and 22 month time points require 12 and 14 tests, respectively. Where fewer than these numbers  
379 were available in a RAPID series, the available data were taken and the series randomly re-sampled to make up  
380 the required series length.

381

#### 382 a) Prediction of future VF state

383 The purpose in this section was to evaluate how well the three analysis methods (detailed above) model the true  
384 rate of VF loss. As there is no ‘gold standard’ for the true rate, a surrogate indicator was investigated. This  
385 surrogate is the accuracy for predicting the final VF (sensitivity at each location) in a series based on the initial 5  
386 visits in the series and the rate of loss estimated by the analysis method.

387 This analysis was performed on 445 eyes in the dataset with sufficiently long follow-up and both VF tests and  
388 OCT images (irrespective of image quality). A trend line fitted to the tests in the first 5 visits by OLSLR (as in  
389 PoPLR) and with the ANSWERS and sANSWERS techniques. The per-subject error for a method is the average  
390 absolute difference between the measured sensitivity and the predicted sensitivity across the 52 non-blind spot  
391 locations in the VF. The absolute difference is the square root of the squared error.

392

393

#### Survival analyses

394 The purpose of this section is to evaluate the 3 methods (detailed above) for their ability to distinguish the  
395 treatment arms of the UKGTS in the subset of participants with OCT images (irrespective of image quality).

396 The criterion selected for each method was that which gives a 5% false positive rate when applied at any  
397 particular time point in the series. The GPA criterion applied in the UKGTS is presented for comparison.

398 This analysis was performed on 353 UKGTS participant with OCT data, with the first eye showing progression  
399 labelling the participant has having progressed (failed); this mirrors the clinical trial scenario where the unit of  
400 analysis is the participant. The Hazard Ratio (HR) and associated  $P$  value are given as a measure of treatment  
401 group separation. Calculations were performed with MedCalc Statistical Software version 17.1 (MedCalc  
402 Software bvba, Ostend, Belgium; <https://www.medcalc.org>; 2017)

403 The criterion 5% false positive rate for the 3 methods does not control for the serial application of the criterion  
404 over time (at each test the participant performs), so that the false positive rate for the test series is likely higher  
405 (lower specificity). To offset this higher false positive rate, the combination of two criteria, ANSWERS AND  
406 PoPLR, was evaluated.

407 The agreement between methods in identifying progression in the UKGTS participants with OCT data was also  
408 assessed.

409

410

#### Sample size calculations

411 The purpose of this section was to estimate the required sample size for various clinical trial scenarios for  
412 observation periods of 12 and 18 months per participant and equal allocation of participants between study arms.

413 The trial scenarios were comparing:

414

1. placebo with an intervention with an effect size of that observed for latanoprost in the UKGTS

- 415 2. an intervention half as effective as latanoprost with an intervention with an effect size equivalent to
- 416 latanoprost
- 417 3. an intervention 75% as effective as latanoprost with an intervention with an effect size equivalent to
- 418 latanoprost
- 419 4. an intervention with an effect size equivalent to latanoprost with a combination treatment with an effect
- 420 size equivalent to 2\*latanoprost (latanoprost plus latanoprost)
- 421 5. an intervention with an effect size equivalent to latanoprost with a combination treatment with an effect
- 422 size equivalent to 1.5\*latanoprost (latanoprost plus ½ latanoprost)

423

424 The sample size calculations were based on survival curves of UKGTS data and the ‘ANSWERS AND PoPLR’  
425 criterion for VF deterioration. The hazard ratio (HR) for the Latanoprost group compared to the Placebo group  
426 was 0.472; a HR of 0.500 was taken for the calculations. In the UKGTS data, progression (deterioration) events  
427 were observed from 10 weeks onwards (one sufficient data had been collected for analysis), so the event rate  
428 was calculated over the 10 to 78 week (18 month) = 68 week interval (Figure 4). The event rate for the Placebo  
429 group was approximately 52% over 68 weeks = 0.76%/week; for the Latanoprost group, the rate was  
430 approximately 28% over 68 weeks = 0.41%/week. For each scenario, the calculations were made for the 42 and  
431 68 week periods over which deterioration events could be identified and then the initial 10-week data collection  
432 period was added back to give the total observation period.

433 The observed attrition rate (loss to follow-up) over the 68 week period was approximately 0.5% per week. In  
434 addition, approximately 10% of UKGTS participants were lost to follow-up before the 10 week time point.  
435 These attrition rates were assumed for the sample size calculations.

436 Samples sizes were estimated for definitively-powered studies (Type I error rate of 0.05 and Type II error rate of  
437 0.10) and pilot studies (Type I error rate of 0.10 and Type II error rate of 0.20) for various study scenarios.

438 The sample size calculations were made with an on-line calculator.<sup>67,68</sup>

439

440

441

442 **RESULTS**

443

444 **GROWTH CURVE MODEL**

445 **Visual field analysis**

446 There was a significant interaction between rate of change and intervention, so that latanoprost-treated eyes had  
447 a more positive rate of VF MD change than the placebo-treated eyes ( $P=.001$ ; Tables 4 and 5).

448 The distribution of rates of change is shown in Figure 5. It can be seen clearly in the histogram that the placebo  
449 group has faster rates of deterioration than the latanoprost group (data shifted to the left). The d'Agostino-  
450 Pearson test for Normal distribution rejected normality ( $P<.0001$ ). A Mann-Whitney two-tailed test  
451 (independent samples) identified that the distribution of slopes was significantly different  $P=.0015$ .

452

453 **OCT analysis**

454 There was no difference in average RNFLT at baseline between intervention groups. Overall, average RNFLT  
455 changes at a rate of  $-1.39$  ( $-1.79$  to  $-0.99$ ) microns per year (data not shown); there was there a significant  
456 interaction showing that this rate of change was statistically significant (Table 6). There was, however, no  
457 significant difference in the rate of RNFLT change between the placebo- and latanoprost-treated groups. Table 7  
458 give the average slope values for each group,  $-1.7$  microns/year for the placebo group and  $-1.1$  microns/year in  
459 the latanoprost group ( $P=.14$ ).

460 The distribution of rates of change is shown in Figure 6. Similarly to the VF data, the placebo group has faster  
461 rates of deterioration than the latanoprost group (data shifted to the left). The d'Agostino-Pearson test  
462 for Normal distribution rejected normality ( $P=.0026$ ). A Mann-Whitney two-tailed test (independent samples)  
463 identified that the distribution of slopes approached statistical significance  $P=.0799$ .

464

465

466 **ASSOCIATION OF RNFLT CHANGE WITH VF SURVIVAL**

467 The VF progression-free survival is presented in Figure 7 for the participants in the UKGTS with Oct data.

468 The significance of the association of various factors with progression-free survival is given in Table 8. Only  
469 treatment allocation was significantly associated with survival ( $P=.0094$ ), however, baseline (pre-treatment)  
470 IOP, baseline (visit 1) VF MD and the rate of OCT RNFLT change approached statistical significance ( $P$   
471 between  $.07$  and  $.08$ ).

472

473

474 **EVALUATION OF 3 STATISTICAL MODELS**

475 a) Progression detection sensitivity

476 Figure 8 illustrates the 'hit rate' (true positives plus false positives with the 5% criterion in the UKGTS data set)  
477 plotted against the false positive rate (subjects identified as deteriorating in the 'stable' test retest data set) as the  
478 criterion for flagging an eye as deteriorating is varied.

479 At the 5% false positive rate and after 22 months observation, the hit rate for the ANSWERS and PoPLR  
480 methods was very similar, at about 38%. For comparison, the hit rate with the GPA criterion applied in the  
481 UKGTS in this subset of eyes with OCT data was 87/394 eligible eyes (22%). The hit rate for sANSWERS was  
482 considerably greater at about 72%, suggesting that, for the same false positive, sANSWERS is much more  
483 sensitive at identifying a progressing eye. A similar pattern is seen for shorter follow-up durations, but with  
484 ANSWERS showing greater sensitivity than PoPLR for short follow-up durations.

485

486 b) Prediction of future VF state

487 The period over which the initial trend line was fitted was a mean (standard deviation) 43.7 (6.6) weeks and the  
488 interval from the initial period to the predicted VF was 54.0 (19.7) weeks. The median (5<sup>th</sup> to 95<sup>th</sup> centile)  
489 prediction error across subjects was 3.9 (1.9 to 8.2) dB for OLSLR, 3.1 (1.6 to 6.0) dB for ANSWERS and 2.5  
490 (1.4 to 4.9) dB for sANSWERS. The difference between methods was evaluated with the Wilcoxon signed-rank  
491 test; all pairs of comparisons were significantly different at the  $P<.0001$  level.

492

493

494 **SURVIVAL ANALYSES**

495 The following analyses apply to 353 UKGTS participants with OCT data, with the participant the unit of  
496 analysis (either eye, if eligible, showing progression).

- 497 a) GPA analysis  
 498 For reference, the survival analysis according to the GPA survival criterion applied in the UKGTS is  
 499 shown in Figure 7. The HR is 0.543 (95% CI 0.312 – 0.838); Logrank test to compare the survival  
 500 curves was significant at  $P=0.006$   
 501 Four of 70 participants in the RAPID data set demonstrated progression by this criterion. Therefore, the  
 502 false positive estimate for the VF series (when this criterion is applied to each VF test in the series) in  
 503 the RAPID data was  $= 4/70 = 5.7\%$  (95% CI 1.6% - 14.6%)  
 504
- 505 b) ANSWERS  
 506 The survival analysis according to the ANSWERS criterion is shown in Figure 9.  
 507 The HR is 0.602 (95% CI 0.441 – 0.821); Logrank test to compare the survival curves was significant  
 508 at  $P=0.0012$   
 509
- 510 c) PoPLR  
 511 The survival analysis according to the PoPLR criterion is shown in Figure 10.  
 512 The HR is 0.590 (95% CI 0.435 to 0.800); Logrank test to compare the survival curves was significant  
 513 at  $P=0.0006$   
 514
- 515 d) sANSWERS  
 516 The survival analysis according to the PoPLR criterion is shown in Figure 11.  
 517 The HR is 0.834 (95% CI 0.655 – 1.066); Logrank test to compare the survival curves was not  
 518 significant ( $P=0.13$ )  
 519
- 520 e) Combined ‘ANSWERS AND PoPLR’  
 521 The survival analysis according to the ‘ANSWERS AND PoPLR’ criterion is shown in Figure 12.  
 522 The HR is 0.472 (95% CI 0.333 – 0.668); Logrank test to compare the survival curves was significant  
 523 at  $P<0.0001$   
 524
- 525 f) The agreement between the GPA, ANSWERS and PoPLR criteria in identifying progression is shown  
 526 in Figure 13. The agreement was ‘fair’ to ‘moderate’, with the following weighted Kappa values: GPA  
 527 vs ANSWERS 0.34 (95% CI 0.25 to 0.42), GPA vs PoPLR 0.34 (95% CI 0.25 to 0.42) and ANSWERS  
 528 vs PoPLR 0.58 (95% CI 0.50 to 0.67).  
 529  
 530

### 531 **SAMPLE SIZE CALCULATIONS**

532 Sample size calculations have been calculated for studies of 12 and 18 months per participant and for a  
 533 definitive study (Type I error rate of 0.05, Type II error rate of 0.10) and a pilot study (Type I error rate of 0.10,  
 534 Type II error rate of 0.20). The numbers given are for the total sample (both arms).  
 535

- 536 1. Sample size for a placebo-controlled study, with an effect size of that observed for latanoprost in the  
 537 UKGTS (Table 9); assumed HR 0.50 and event rate in Placebo group of 0.76%/week (0.395 events/year).  
 538
- 539 2. Sample size comparing an intervention half as effective as latanoprost (group 0) with an intervention  
 540 with an effect size equivalent to latanoprost (Table 10); assumed HR 0.50 and event rate in group 0 of  
 541 0.58%/week (0.304 events/year).  
 542
- 543 3. Sample size comparing an intervention 75% as effective as latanoprost (group 0) with an intervention  
 544 with an effect size equivalent to latanoprost (Table 11); assumed HR 0.75 and event rate in group 0 of  
 545 0.50%/week (0.259 events/year).  
 546
- 547 4. Sample size comparing an intervention with an effect size equivalent to latanoprost (group 0) with a  
 548 combination treatment with an effect size equivalent to 2\*latanoprost (latanoprost plus latanoprost) (Table 12);  
 549 assumed HR 0.50 and event rate in group 0 of 0.41%/week (0.213 events/year).  
 550
- 551 5. Sample size comparing an intervention with an effect size equivalent to latanoprost (group 0) with a  
 552 combination treatment with an effect size equivalent to 1.5\*latanoprost (latanoprost plus ½ latanoprost) (Table  
 553 13); assumed HR 0.75 and event rate in group 0 of 0.41%/week (0.213 events/year).  
 554

555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572

573 **DISCUSSION**

574 The results of this study show that, whereas the rate of RNFLT loss was faster in the placebo-treated eyes, the  
575 difference from the latanoprost-treated eyes did not reach statistical significance. However, the association of  
576 the rate of RNFLT change with incident VF loss approached significance and adding the rate of RNFLT change  
577 as a Bayesian prior in a model of VF progression made the model considerably more sensitive at identifying  
578 progression (for the same false positive rate) and more accurate in modelling the rate of progression. Despite  
579 this, adding the OCT structural data to the vision function data from VF testing did not provide greater  
580 separation between the treatment groups in the UKGTS.

581 Identifying the best model for analysing times series of repeated data is challenging. We chose growth curve  
582 models as the most suitable. This analysis identified a highly statistically significant difference ( $P=0.001$ )  
583 between treatment groups based on the rate of VF MD change, but did not identify a difference ( $P=0.14$ ) between  
584 treatment groups based on the rate of OCT RNFLT change. It is obvious that the signal compared to the ‘noise’  
585 (variability) is lower in the OCT data than in the VF. The growth curve models assume a Normal distribution of  
586 the rate of change data. Figures 5 and 6 show that the data are not normally distributed. There are likely two  
587 underlying distributions – the noise, which may be approximately normally distributed and the signal (true rates  
588 of change) which may have a distribution approximating a Weibull probability density function ( $\kappa=0.5, \lambda=1$ ;  
589 Figure 14), with many subjects changing slowly and fewer changing more rapidly. The effect of treatment on  
590 these slopes of change may be greatest on those changing the fastest, so that a parametric approach fails to  
591 identify that signal. A Mann-Whitney test identified that the distribution of RNFLT slopes approached statistical  
592 significance ( $P=0.08$ ), however, this analysis does not take account of the variance in the measurements giving  
593 rise to the slope estimates. It may be that non-parametric multilevel models may better detect the signal in the  
594 data. {Rights, 2016 #2331} That said, the principal problem is that the signal-to-noise ratio in the TD OCT data  
595 is low relative to that of the VF data. The variability characteristics of measurements from spectral-domain (SD)  
596 OCT images are much better, with the variability of SD OCT RNFLT measurements being about half that of TD  
597 OCT.<sup>69</sup>

598 The Cox proportional hazards analysis, with OCT RNFLT as a predictor variable, demonstrated that the rate of  
599 RNFLT change approached significance as a predictor of incident VF loss ( $P=0.0722$ ). Thus, the data in this  
600 study support that the treatment effect on RNFLT measurements is in the same direction as that on VF  
601 measurements and that the structural outcomes are associated with the VF loss, but the signal-to-noise ratio of  
602 the TD OCT measurements is insufficient for the measurements to have much utility in the context of study  
603 power. SD OCT, because of its better signal-to-noise characteristics, may be more useful.

604 When the RNFLT rate of change is included as a Bayesian prior in the ANSWERS technique (structure-guided  
605 ANSWERS; sANSWERS), the accuracy of modelling the rate of VF loss, as estimated by the prediction of  
606 future VF loss, is improved over that of ANSWERS without the structural prior and the PoPLR technique. This  
607 implies that the RNFLT data contain information relevant to VF loss. Furthermore, when the false-positive rate  
608 was equated between techniques, sANSWERS had considerably greater sensitivity to identify progression than  
609 ANSWERS and PoPLR.

610 The optimal outcome measure for a clinical trial should distinguish the treatment groups (the HR should indicate  
611 a large difference) and the proportion of participants with an outcome should be high, so the number of  
612 participants required for the trial is low and/or the duration of observation is short. However, the proportion of  
613 participants with an outcome should not be so high that the identification of a difference between treatments  
614 groups is precluded. The GPA criterion applied in the UKGTS was designed to have greater sensitivity in the  
615 24-2 VF than the conventional GPA criterion (three locations different from baseline at the 5% level on three  
616 consecutive occasions), which was designed for the 30-2 VF tests used in the Early Manifest Glaucoma Trial  
617 (EMGT)<sup>70</sup>; the 30-2 test has 40% more test locations than the 24-2, so the opportunity to detect progression is  
618 greater for a 30-2 VF. The false-positive rate of the UKGTS criterion in the RAPID data set was 5.7% (95% CI  
619 1.6% - 14.6%). This compares with an estimated false-positive rate of 2.6% over the course of 10 follow-up  
620 visits for the EMGT GPA criterion in the 24-2 VF.<sup>59</sup> The UKGTS GPA criterion distinguished between the  
621 treatment groups well (the HR in the subset of UKGTS participants with OCT images was 0.543 (95% CI 0.312  
622 – 0.838),  $P=0.006$ ). The ANSWERS and PoPLR techniques distinguished similarly well, but with a greater  
623 number of events (Figure 13), which is a positive attribute. The false-positive rate for the ANSWERS, PoPLR  
624 and sANSWERS was set at 5% for each application. In clinical practice, as well as in clinical trials, such  
625 progression analyses are applied at each visit. Thus, the serial application of the analysis is likely to inflate the  
626 false-positive rate. The approach taken in this work to mitigate this effect was to evaluate a criterion for  
627 progression that required change by both ANSWERS and PoPLR. This resulted in very good separation  
628 between treatment groups (HR 0.472 (95% CI 0.333 – 0.668);  $P<0.0001$ ) and a moderately high proportion of  
629 participants with progression.

630 The sANSWERS technique, as shown by the estimate of sensitivity at a 5% false-positive rate, is considerably  
631 more sensitive than the other techniques. The consequence of this in the survival analysis is that so many  
632 participants are identified as progressing that the opportunity to distinguish the treatment groups is reduced.  
633 The sample size estimates show that a placebo-controlled trial of an intervention as effective as latanoprost can  
634 be undertaken with an observation period of only 12 months and as few participants as 502. However, sample  
635 sizes need to be much larger for studies comparing the impact of the addition of a treatment to latanoprost. For  
636 example, identifying the treatment benefit of an intervention half as effective as latanoprost when added to  
637 latanoprost requires 3029 participants observed over a period of 18 months.

638 The sample size estimates are conservative, including both an initial drop-out rate of 10% and an additional rate  
639 of 25% per year over the duration of follow-up. These figures are based on the UKGTS, which had an especially  
640 onerous follow-up regime with many investigations and questionnaires at initial visits, as well as frequent visits.  
641 Although the frequency of visits would need to be maintained in future trials, the burden of tests could be  
642 reduced, with an anticipated beneficial impact on the loss to follow-up rate.

643 Naturally, these sample size estimates relate to cohorts similar to the UKGTS cohort; that is newly-diagnosed  
644 subjects with early glaucoma and relatively low IOP. Including newly-diagnosed patients has advantages and  
645 disadvantages. An important advantage is that such patients have not had any previous disease-modifying  
646 treatment, so the placebo arm fairly reflects the natural history of untreated glaucoma and the treatment arm  
647 provides information on the disease modifying effect of a single intervention. However, even though the  
648 UKGTS protocol included steps to minimize the inclusion of subjects still learning the VF test,<sup>63</sup> the mean MD  
649 slope in the treatment arm was slightly positive (0.03 dB/year), despite approximately 20% of latanoprost-  
650 treated subjects being identified as having VF deterioration in the first year (by the 'ANSWERS AND PoPLR'  
651 criterion). This net slight improvement in VF MD suggests either that treatment induces visual field  
652 improvement in a proportion of patients or that VF learning effects are causing progressively more positive MD  
653 measurements over time. The former hypothesis was tested recently in the EMGT data and found not to be the  
654 case.<sup>71</sup> If the latter hypothesis is the case, then the measured rates of VF likely underestimate the true rate of  
655 glaucoma-related VF loss. Thus the -0.29dB/year average rate of MD loss in the placebo-treated arm may be an  
656 under-estimate. Although the average IOP in the UKGTS cohort, at approximately 20mmHg,<sup>12</sup> was less than  
657 1mmHg lower than the average IOP in the EMGT, the rate of MD loss in the untreated arm was half that in the  
658 EMGT (-0.29 dB/year in the UKGTS and -0.6 dB/year in the EMGT,<sup>7</sup> later revised to -1.03 dB/year for a longer  
659 observation period<sup>72</sup>). The rate of VF loss was measured over a longer period in the EMGT, so the impact of VF  
660 learning (if occurring mostly over the initial part of the observation period) may be less than that on the UKGTS  
661 data.

662 Quigley evaluated samples sizes for trials in glaucoma based on assumed rates of MD deterioration.<sup>73</sup> The rates  
663 considered for the (treated) control group were all more than 50% greater than the observed mean rate in  
664 untreated patients in the UKGTS. Thus, the calculations *may* be over-optimistic, although the caveats stated  
665 above apply. Also, Quigley's model assessed the mean and standard deviations of rates of change, whereas it is  
666 known that rate-of-change VF data are not normally distributed.<sup>72</sup> His sample size estimate for a treatment  
667 reducing the rate of progression by 50% over that of a treated control group was 294 (323 adding a 10% initial  
668 loss to follow-up), although Type I and II error rates weren't stated and the duration of observation was not  
669 defined. In the placebo group of the UKGTS, the mean rate of MD change was -0.29 dB/year (median -0.15  
670 dB/year), with a standard deviation of 1.94 dB/year. An observation period longer than the 2 years in the  
671 UKGTS would be required to reduce the standard deviation of the rate of change to the 1.04 dB/year assumed  
672 by Quigley. Our sample size estimate for the same scenario (50% reduction in the rate of progression over that  
673 of a treated control group), based on UKGTS trial data, for an observation period of 18 months, was 601  
674 participants (including the 10% initial loss to follow up).

675 Because the IOP level was not a recruitment criterion, the UKGTS cohort is probably fairly representative of an  
676 unselected clinical glaucoma population and the results of the trial can, therefore, be generalized to patients in  
677 the clinic. A caveat is that no data were obtained on the IOP and degree of VF loss of subjects declining to  
678 participate in the UKGTS. If there had been a tendency for individuals with higher IOP and greater degrees of  
679 VF loss to decline participation, then the UKGTS cohort may have 'milder' disease than the unselected clinical  
680 glaucoma population. Study power is strongly influenced by the event rate (in this case, VF deterioration) and,  
681 therefore, study power may be increased (and the required sample size and observation duration may be  
682 reduced) by enriching the study population with patients more likely to achieve a deterioration event. This can  
683 be done by selecting patients on the basis of risk factors for deterioration, such as higher IOP or the presence of  
684 optic disc haemorrhages. Whereas doing this may reduce the required sample size or observation duration, there  
685 are potential disadvantages. The outcome of such studies can only be generalized to similar patients and there is  
686 a risk that a treatment effect may be incorrectly estimated if the treatment is more, or less, effective in the trial  
687 cohort compared to the target clinic population. Disc haemorrhages, for example, are well known to be a risk  
688 factor for glaucoma deterioration,<sup>74,75</sup> and, although IOP-lowering may be beneficial in these eyes,<sup>76</sup> the



689 incidence of disc haemorrhages does not seem to be affected by IOP-lowering treatment.<sup>77</sup> If disc haemorrhages  
690 represent, at least in part, a non-IOP related risk, then enriching a population with patients with a history of disc  
691 haemorrhages in a study assessing the effect of IOP-lowering may not increase study power and may, in fact,  
692 have the opposite effect.

### 693 **LIMITATIONS AND FURTHER WORK**

694 The major limitation in these data is the imaging technology that was available at the time. The finding of little  
695 benefit to trial power may relate to the low signal-to-noise ratio of the TD OCT RNFLT measurements. Future  
696 trials assessing the potential of SD OCT are warranted.

697 The ANSWERS, PoPLR and sANSWERS progression criteria were not adjusted to account for the impact of  
698 multiple testing in time on the false-positive rate. Further work will explore the adjusting of the significance  
699 criterion on the separation between treatment groups and the proportion of subjects identified as progression. An  
700 additional ‘rate of change’ threshold criterion may also be beneficial.

701 In searching for the appropriate statistical techniques to evaluate the difference in repeated measures over time,  
702 non-parametric approaches may be helpful. {Rights, 2016 #2331} The joint modelling of incident VF loss with  
703 the rate of change in structural measurements, as suggested by Medeiros,<sup>48</sup> may be helpful and non-parametric  
704 approached need to be explored.<sup>78,79</sup>

705 A limitation that is hard to address when evaluating alternative progression criteria in real-world trial data is that  
706 the data are censored as a consequence of the progression criterion that were applied in the trial – once a  
707 participant is identified as progressing (s)he exits the study and the data series is curtailed. If an alternative  
708 progression criterion fails to identify progression in a censored series, it is not possible to know whether that  
709 criterion may have identified progression in that participant had the data not been censored. The only way  
710 around this problem is to build virtual models of progressing patients.

711 The estimate of specificity for the UKGTS GPA criterion was made in 70 RAPID study participants, so the  
712 estimate is fairly imprecise. Permuting the VF series from these 70 participants may increase the precision.  
713 However, it is presently not possible to permute VF data and analyse GPA progression with the GPA software.

714  
715

716 APPENDIX

717 The equation for a longitudinal model allowing for the interaction between rate of change and intervention  
718 group is shown below:

719

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 rand_j + \beta_3 t_{ij} rand_j + u_{0j} + u_{1j} t_{ij} + \varepsilon_{ij}$$

720

721  $i$  = occasion of repeated measure (level 1 indicator)

722  $j$  = participant (level 2 indicator)

723  $y_{ij}$  = Response of outcome at occasion  $i$  for participant  $j$

724  $t_{ij}$  = time of occasion  $i$  for participant  $j$

725  $rand_j$  = Randomisation group for participant  $j$

726  $\beta_0$  = Overall intercept, expected value of  $y$  when  $t_{ij}=0$  and  $rand=0$

727  $\beta_1$  = Average regression coefficient of time for patients in the placebo group ( $rand=0$ )

728  $\beta_2$  = Treatment effect/difference between treatments when  $t_{ij}=0$

729  $\beta_3$  = Interaction coefficient between time and intervention group

730  $u_{0j}$  = Individual-specific (between participants) random effect of the intercept (allows each patient to have their  
731 own intercept)

732  $u_{1j}$  = Individual-specific (between participants) random effect of the time coefficient (random slope: allows  
733 each patient to have their own slope)

734  $\varepsilon_{ij}$  = occasion-specific (within participant) residual

735

736 In Stata, the VF model specified was:

737 **xtmixed** md i.rand##c.ytime || studyno: ytime, **cov(uns)**

738 md = mean deviation; rand = randomised treatment (reference group = placebo); ytime = continuous time in  
739 years between visual field measurements

740 The OCT model specified was:

741 **xtmixed** mean\_avg\_thickness i.rand##c.ytime || studyno: ytime, **cov(uns)**

742 mean\_avg\_thickness = average RNFL thickness from repeats within visit; rand = randomised treatment  
743 (reference group = placebo); ytime = continuous time in years between OCT measurements;

744

745 VF measurements were repeated at several visits (1, 2, 7, 8 and 11); the intended purpose was to obtain a more  
746 precise estimate of the slope. This resulted in a 3-level structure of the data; tests at level 1, nested within visits  
747 at level 2, nested within participants at level 3 (Figure 2a).

748 In a longitudinal model, the measurement occasion and therefore its indicator (e.g. time) form level 1 units,  
749 however, available VF data indicated only the day of follow-up visit (level 2) rather than the exact time of each  
750 test, so that the time of the two measurements could not be distinguished at level 1. Therefore, we estimated the  
751 time tests were taken, based on knowledge of the study protocol (on average there was likely to be 2.5 hours  
752 between VF tests that were taken on the same day). We used the variable VF\_id to order these repeat visual field  
753 tests within a visit and added 2.5 hours of time between visual field tests. Thus the data could now be  
754 restructured to 2-levels (Figure 2b).

755 OCT scans were taken at repeated follow-up visits. Within each visit, typically 3 scans were taken (5 at baseline  
756 and last visit), with three repeat instances within scans (fast RNFL protocol). Leading to a 4-level structure;  
757 instances at level 1, nested within scans at level 2, nested within visits at level 3, nested within participants at  
758 level 4 (Figure 3a). The three repeat instances within scans were averaged to provide a single scan result  
759 (mimicking the OCT software output). The time of each scan was recorded in the data, so we were able to  
760 restructure the data into two levels (Figure 3b) according to actual scan time.

761

762

763 REFERENCES

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

- 767 2. Henson DB, Chaudry S, Artes PH, Faragher EB, Ansons A. Response variability in the visual field:  
768 comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Invest Ophthalmol Vis*  
769 *Sci.* 2000;41(2):417-421.
- 770 3. Artes PH, Iwase A, Ohno Y, Kitazawa Y, Chauhan BC. Properties of perimetric threshold estimates  
771 from Full Threshold, SITA Standard, and SITA Fast strategies. *Invest Ophthalmol Vis Sci.*  
772 2002;43(8):2654-2659.
- 773 4. Russell RA, Crabb DP, Malik R, Garway-Heath DF. The relationship between variability and  
774 sensitivity in large-scale longitudinal visual field data. *Invest Ophthalmol Vis Sci.* 2012;53(10):5985-  
775 5990.
- 776 5. Chauhan BC, Garway-Heath DF, Goni FJ, et al. Practical recommendations for measuring rates of  
777 visual field change in glaucoma. *Br J Ophthalmol.* 2008;92(4):569-573.
- 778 6. Jansonius NM. On the accuracy of measuring rates of visual field change in glaucoma. *Br J*  
779 *Ophthalmol.* 2010;94(10):1404-1405.
- 780 7. Heijl A, Leske MC, Bengtsson B, et al. Reduction of intraocular pressure and glaucoma progression:  
781 results from the Early Manifest Glaucoma Trial. *Arch Ophthalmol.* 2002;120(10):1268-1279.
- 782 8. Group CN-TGS. Comparison of glaucomatous progression between untreated patients with normal-  
783 tension glaucoma and patients with therapeutically reduced intraocular pressures. Collaborative  
784 Normal-Tension Glaucoma Study Group. *Am J Ophthalmol.* 1998;126(4):487-497.
- 785 9. Musch DC, Gillespie BW, Lichter PR, Niziol LM, Janz NK, Investigators CS. Visual field progression  
786 in the Collaborative Initial Glaucoma Treatment Study the impact of treatment and other baseline  
787 factors. *Ophthalmology.* 2009;116(2):200-207.
- 788 10. Investigators TA. The Advanced Glaucoma Intervention Study (AGIS): 7. The relationship between  
789 control of intraocular pressure and visual field deterioration. The AGIS Investigators. *Am J Ophthalmol.*  
790 2000;130(4):429-440.
- 791 11. Krupin T, Liebmann JM, Greenfield DS, Ritch R, Gardiner S, Low-Pressure Glaucoma Study G. A  
792 randomized trial of brimonidine versus timolol in preserving visual function: results from the Low-  
793 Pressure Glaucoma Treatment Study. *Am J Ophthalmol.* 2011;151(4):671-681.
- 794 12. Garway-Heath DF, Crabb DP, Bunce C, et al. Latanoprost for open-angle glaucoma (UKGTS): a  
795 randomised, multicentre, placebo-controlled trial. *Lancet.* 2015;385(9975):1295-1304.
- 796 13. Crabb DP, Garway-Heath DF. Intervals between visual field tests when monitoring the glaucomatous  
797 patient: wait-and-see approach. *Invest Ophthalmol Vis Sci.* 2012;53(6):2770-2776.
- 798 14. Strouthidis NG, Scott A, Peter NM, Garway-Heath DF. Optic disc and visual field progression in  
799 ocular hypertensive subjects: detection rates, specificity, and agreement. *Invest Ophthalmol Vis Sci.*  
800 2006;47(7):2904-2910.
- 801 15. Poli A, Strouthidis NG, Ho TA, Garway-Heath DF. Analysis of HRT images: comparison of reference  
802 planes. *Invest Ophthalmol Vis Sci.* 2008;49(9):3970-3975.
- 803 16. Medeiros FA, Leite MT, Zangwill LM, Weinreb RN. Combining structural and functional  
804 measurements to improve detection of glaucoma progression using Bayesian hierarchical models.  
805 *Invest Ophthalmol Vis Sci.* 2011;52(8):5794-5803.
- 806 17. Leung CK, Chiu V, Weinreb RN, et al. Evaluation of retinal nerve fiber layer progression in glaucoma:  
807 a comparison between spectral-domain and time-domain optical coherence tomography.  
808 *Ophthalmology.* 2011;118(8):1558-1562.
- 809 18. Leske MC, Heijl A, Hyman L, et al. Predictors of long-term progression in the early manifest glaucoma  
810 trial. *Ophthalmology.* 2007;114(11):1965-1972.
- 811 19. Airaksinen PJ, Drance SM, Douglas GR, Schulzer M. Neuroretinal rim areas and visual field indices in  
812 glaucoma. *Am J Ophthalmol.* 1985;99(2):107-110.
- 813 20. Jonas JB, Grondler AE. Correlation between mean visual field loss and morphometric optic disk  
814 variables in the open-angle glaucomas. *Am J Ophthalmol.* 1997;124(4):488-497.
- 815 21. Bartz-Schmidt KU, Thumann G, Jonescu-Cuypers CP, Krieglstein GK. Quantitative morphologic and  
816 functional evaluation of the optic nerve head in chronic open-angle glaucoma. *Surv Ophthalmol.*  
817 1999;44 Suppl 1:S41-53.
- 818 22. Garway-Heath DF, Holder GE, Fitzke FW, Hitchings RA. Relationship between electrophysiological,  
819 psychophysical, and anatomical measurements in glaucoma. *Invest Ophthalmol Vis Sci.*  
820 2002;43(7):2213-2220.
- 821 23. Ajtony C, Balla Z, Somoskeoy S, Kovacs B. Relationship between visual field sensitivity and retinal  
822 nerve fiber layer thickness as measured by optical coherence tomography. *Invest Ophthalmol Vis Sci.*  
823 2007;48(1):258-263.

- 824 24. Read RM, Spaeth GL. The practical clinical appraisal of the optic disc in glaucoma: the natural history  
825 of cup progression and some specific disc-field correlations. *Trans Am Acad Ophthalmol Otolaryngol.*  
826 1974;78(2):OP255-274.
- 827 25. Garway-Heath DF, Poinoosawmy D, Fitzke FW, Hitchings RA. Mapping the visual field to the optic  
828 disc in normal tension glaucoma eyes. *Ophthalmology.* 2000;107(10):1809-1815.
- 829 26. Garway-Heath DF, Hitchings RA. Quantitative evaluation of the optic nerve head in early glaucoma.  
830 *Br J Ophthalmol.* 1998;82(4):352-361.
- 831 27. Wollstein G, Garway-Heath DF, Hitchings RA. Identification of early glaucoma cases with the  
832 scanning laser ophthalmoscope. *Ophthalmology.* 1998;105(8):1557-1563.
- 833 28. Deleon-Ortega JE, Arthur SN, McGwin G, Jr., Xie A, Monheit BE, Girkin CA. Discrimination  
834 between glaucomatous and nonglaucomatous eyes using quantitative imaging devices and subjective  
835 optic nerve head assessment. *Invest Ophthalmol Vis Sci.* 2006;47(8):3374-3380.
- 836 29. Izatt JA, Hee MR, Swanson EA, et al. Micrometer-scale resolution imaging of the anterior eye in vivo  
837 with optical coherence tomography. *Arch Ophthalmol.* 1994;112(12):1584-1589.
- 838 30. Schuman JS, Hee MR, Puliafito CA, et al. Quantification of nerve fiber layer thickness in normal and  
839 glaucomatous eyes using optical coherence tomography. *Arch Ophthalmol.* 1995;113(5):586-596.
- 840 31. Schuman JS, Hee MR, Arya AV, et al. Optical coherence tomography: a new tool for glaucoma  
841 diagnosis. *Curr Opin Ophthalmol.* 1995;6(2):89-95.
- 842 32. Akashi A, Kanamori A, Nakamura M, Fujihara M, Yamada Y, Negi A. Comparative assessment for the  
843 ability of Cirrus, RTVue, and 3D-OCT to diagnose glaucoma. *Invest Ophthalmol Vis Sci.*  
844 2013;54(7):4478-4484.
- 845 33. Chauhan BC, McCormick TA, Nicolela MT, LeBlanc RP. Optic disc and visual field changes in a  
846 prospective longitudinal study of patients with glaucoma: comparison of scanning laser tomography  
847 with conventional perimetry and optic disc photography. *Arch Ophthalmol.* 2001;119(10):1492-1499.
- 848 34. Wollstein G, Schuman JS, Price LL, et al. Optical coherence tomography longitudinal evaluation of  
849 retinal nerve fiber layer thickness in glaucoma. *Arch Ophthalmol.* 2005;123(4):464-470.
- 850 35. Artes PH, Chauhan BC. Longitudinal changes in the visual field and optic disc in glaucoma. *Prog Retin*  
851 *Eye Res.* 2005;24(3):333-354.
- 852 36. Leung CK, Cheung CY, Weinreb RN, et al. Evaluation of retinal nerve fiber layer progression in  
853 glaucoma: a study on optical coherence tomography guided progression analysis. *Invest Ophthalmol*  
854 *Vis Sci.* 2010;51(1):217-222.
- 855 37. Mansouri K, Leite MT, Medeiros FA, Leung CK, Weinreb RN. Assessment of rates of structural  
856 change in glaucoma using imaging technologies. *Eye (Lond).* 2011;25(3):269-277.
- 857 38. Xin D, Greenstein VC, Ritch R, Liebmann JM, De Moraes CG, Hood DC. A comparison of functional  
858 and structural measures for identifying progression of glaucoma. *Invest Ophthalmol Vis Sci.*  
859 2011;52(1):519-526.
- 860 39. Leung CK, Yu M, Weinreb RN, Lai G, Xu G, Lam DS. Retinal nerve fiber layer imaging with spectral-  
861 domain optical coherence tomography: patterns of retinal nerve fiber layer progression.  
862 *Ophthalmology.* 2012;119(9):1858-1866.
- 863 40. Leung CK, Ye C, Weinreb RN, Yu M, Lai G, Lam DS. Impact of age-related change of retinal nerve  
864 fiber layer and macular thicknesses on evaluation of glaucoma progression. *Ophthalmology.*  
865 2013;120(12):2485-2492.
- 866 41. Leung CK. Diagnosing glaucoma progression with optical coherence tomography. *Curr Opin*  
867 *Ophthalmol.* 2014;25(2):104-111.
- 868 42. Abe RY, Diniz-Filho A, Zangwill LM, et al. The Relative Odds of Progressing by Structural and  
869 Functional Tests in Glaucoma. *Invest Ophthalmol Vis Sci.* 2016;57(9):OCT421-428.
- 870 43. Chauhan BC, Nicolela MT, Artes PH. Incidence and rates of visual field progression after  
871 longitudinally measured optic disc change in glaucoma. *Ophthalmology.* 2009;116(11):2110-2118.
- 872 44. Medeiros FA, Alencar LM, Zangwill LM, Bowd C, Sample PA, Weinreb RN. Prediction of functional  
873 loss in glaucoma from progressive optic disc damage. *Arch Ophthalmol.* 2009;127(10):1250-1256.
- 874 45. Weinreb RN, Kaufman PL. Glaucoma research community and FDA look to the future, II: NEI/FDA  
875 Glaucoma Clinical Trial Design and Endpoints Symposium: measures of structural change and visual  
876 function. *Invest Ophthalmol Vis Sci.* 2011;52(11):7842-7851.
- 877 46. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med.*  
878 1989;8(4):431-440.

- 879 47. Medeiros FA. Biomarkers and surrogate endpoints in glaucoma clinical trials. *Br J Ophthalmol*.  
880 2015;99(5):599-603.
- 881 48. Medeiros FA, Lisboa R, Zangwill LM, et al. Evaluation of progressive neuroretinal rim loss as a  
882 surrogate end point for development of visual field loss in glaucoma. *Ophthalmology*.  
883 2014;121(1):100-109.
- 884 49. McNaught AI, Crabb DP, Fitzke FW, Hitchings RA. Modelling series of visual fields to detect  
885 progression in normal-tension glaucoma. *Graefes Arch Clin Exp Ophthalmol*. 1995;233(12):750-755.
- 886 50. Medeiros FA, Zangwill LM, Weinreb RN. Improved prediction of rates of visual field loss in glaucoma  
887 using empirical Bayes estimates of slopes of change. *J Glaucoma*. 2012;21(3):147-154.
- 888 51. Medeiros FA, Zangwill LM, Girkin CA, Liebmann JM, Weinreb RN. Combining structural and  
889 functional measurements to improve estimates of rates of glaucomatous progression. *Am J Ophthalmol*.  
890 2012;153(6):1197-1205 e1191.
- 891 52. Pathak M, Demirel S, Gardiner SK. Nonlinear, multilevel mixed-effects approach for modeling  
892 longitudinal standard automated perimetry data in glaucoma. *Invest Ophthalmol Vis Sci*.  
893 2013;54(8):5505-5513.
- 894 53. Zhu H, Crabb DP, Ho T, Garway-Heath DF. More Accurate Modeling of Visual Field Progression in  
895 Glaucoma: ANSWERS. *Invest Ophthalmol Vis Sci*. 2015;56(10):6077-6083.
- 896 54. Russell RA, Malik R, Chauhan BC, Crabb DP, Garway-Heath DF. Improved estimates of visual field  
897 progression using bayesian linear regression to integrate structural information in patients with ocular  
898 hypertension. *Invest Ophthalmol Vis Sci*. 2012;53(6):2760-2769.
- 899 55. Bizios D, Heijl A, Bengtsson B. Integration and fusion of standard automated perimetry and optical  
900 coherence tomography data for improved automated glaucoma diagnostics. *BMC Ophthalmol*.  
901 2011;11:20.
- 902 56. Raza AS, Zhang X, De Moraes CG, et al. Improving glaucoma detection using spatially correspondent  
903 clusters of damage and by combining standard automated perimetry and optical coherence tomography.  
904 *Invest Ophthalmol Vis Sci*. 2014;55(1):612-624.
- 905 57. Tatham AJ, Weinreb RN, Medeiros FA. Strategies for improving early detection of glaucoma: the  
906 combined structure-function index. *Clin Ophthalmol*. 2014;8:611-621.
- 907 58. Zhu H, Crabb DP, Fredette MJ, Anderson DR, Garway-Heath DF. Quantifying discordance between  
908 structure and function measurements in the clinical assessment of glaucoma. *Arch Ophthalmol*.  
909 2011;129(9):1167-1174.
- 910 59. Artes PH, O'Leary N, Nicoleta MT, Chauhan BC, Crabb DP. Visual field progression in glaucoma:  
911 what is the specificity of the Guided Progression Analysis? *Ophthalmology*. 2014;121(10):2023-2027.
- 912 60. Mwanza JC, Chang RT, Budenz DL, et al. Reproducibility of peripapillary retinal nerve fiber layer  
913 thickness and optic nerve head parameters measured with cirrus HD-OCT in glaucomatous eyes. *Invest*  
914 *Ophthalmol Vis Sci*. 2010;51(11):5724-5730.
- 915 61. Mwanza JC, Budenz DL, Warren JL, et al. Retinal nerve fibre layer thickness floor and corresponding  
916 functional loss in glaucoma. *Br J Ophthalmol*. 2015;99(6):732-737.
- 917 62. Kotowski J, Wollstein G, Folio LS, Ishikawa H, Schuman JS. Clinical use of OCT in assessing  
918 glaucoma progression. *Ophthalmic Surg Lasers Imaging*. 2011;42 Suppl:S6-S14.
- 919 63. Garway-Heath DF, Lascaratos G, Bunce C, Crabb DP, Russell RA, Shah A. The United Kingdom  
920 Glaucoma Treatment Study: a multicenter, randomized, placebo-controlled clinical trial: design and  
921 methodology. *Ophthalmology*. 2013;120(1):68-76.
- 922 64. Lascaratos G, Garway-Heath DF, Burton R, et al. The United Kingdom Glaucoma Treatment Study: a  
923 multicenter, randomized, double-masked, placebo-controlled trial: baseline characteristics.  
924 *Ophthalmology*. 2013;120(12):2540-2545.
- 925 65. Zhu H, Russell RA, Saunders LJ, Ceccon S, Garway-Heath DF, Crabb DP. Detecting changes in retinal  
926 function: Analysis with Non-Stationary Weibull Error Regression and Spatial enhancement  
927 (ANSWERS). *PLoS One*. 2014;9(1):e85654.
- 928 66. O'Leary N, Chauhan BC, Artes PH. Visual field progression in glaucoma: estimating the overall  
929 significance of deterioration with permutation analyses of pointwise linear regression (PoPLR). *Invest*  
930 *Ophthalmol Vis Sci*. 2012;53(11):6776-6784.
- 931 67. Kohn MA, Jarrett MS, Senyak J. Sample Size Calculators. 2016; <http://www.sample-size.net/sample-size-survival-analysis/>. Accessed 04 Feb 2017.
- 932  
933 68. Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics*.  
934 1983;39(2):499-503.

- 935 69. Leung CK, Cheung CY, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain  
936 optical coherence tomography: a variability and diagnostic performance study. *Ophthalmology*.  
937 2009;116(7):1257-1263, 1263 e1251-1252.
- 938 70. Leske MC, Heijl A, Hyman L, Bengtsson B. Early Manifest Glaucoma Trial: design and baseline data.  
939 *Ophthalmology*. 1999;106(11):2144-2153.
- 940 71. Bengtsson B, Heijl A. Lack of Visual Field Improvement After Initiation of Intraocular Pressure  
941 Reducing Treatment in the Early Manifest Glaucoma Trial. *Invest Ophthalmol Vis Sci*.  
942 2016;57(13):5611-5615.
- 943 72. Heijl A, Bengtsson B, Hyman L, Leske MC, Early Manifest Glaucoma Trial G. Natural history of  
944 open-angle glaucoma. *Ophthalmology*. 2009;116(12):2271-2276.
- 945 73. Quigley HA. Clinical trials for glaucoma neuroprotection are not impossible. *Curr Opin Ophthalmol*.  
946 2012;23(2):144-154.
- 947 74. Budenz DL, Anderson DR, Feuer WJ, et al. Detection and prognostic significance of optic disc  
948 hemorrhages during the Ocular Hypertension Treatment Study. *Ophthalmology*. 2006;113(12):2137-  
949 2143.
- 950 75. Leske MC, Heijl A, Hussein M, et al. Factors for glaucoma progression and the effect of treatment: the  
951 early manifest glaucoma trial. *Arch Ophthalmol*. 2003;121(1):48-56.
- 952 76. Medeiros FA, Alencar LM, Sample PA, Zangwill LM, Susanna R, Jr., Weinreb RN. The relationship  
953 between intraocular pressure reduction and rates of progressive visual field loss in eyes with optic disc  
954 hemorrhage. *Ophthalmology*. 2010;117(11):2061-2066.
- 955 77. Bengtsson B, Leske MC, Yang Z, Heijl A, Group E. Disc hemorrhages and treatment in the early  
956 manifest glaucoma trial. *Ophthalmology*. 2008;115(11):2044-2048.
- 957 78. Ding J, Wang JL. Modeling longitudinal data with nonparametric multiplicative random effects jointly  
958 with survival data. *Biometrics*. 2008;64(2):546-556.
- 959 79. Song X, Davidian M, Tsiatis AA. A semiparametric likelihood approach to joint modeling of  
960 longitudinal and time-to-event data. *Biometrics*. 2002;58(4):742-753.
- 961  
962  
963

964 **ACKNOWLEDGEMENTS**

965

966 **Funding**

967 The sponsor for both the UKGTS and RAPID data collection was Moorfields Eye Hospital NHS Foundation  
968 Trust. The Sponsor was responsible for ensuring the IRB approval and NHS Permissions were in place before  
969 the initiation of the studies and research governance. The Sponsor is the employer of two statisticians  
970 contributing to the analysis of the data (AQ and PP), but had no influence on the choice of analysis or  
971 interpretation of the data.

972 The principal funding for this work was the United Kingdom's National Institute for Health Research Health  
973 Technology Assessment (HTA) Project Funding: 11/129/245 - Assessing the Effectiveness of Imaging  
974 Technology to Rapidly Detect Disease Progression in Glaucoma. Additional unrestricted funding was obtained  
975 from Pfizer Inc to support the statistical analyses.

976 Funding for the UKGTS was through an unrestricted investigator-initiated research grant from Pfizer, with  
977 supplementary funding from the UK's NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS  
978 Foundation Trust and UCL Institute of Ophthalmology, London, UK. Equipment loans were made by  
979 Heidelberg Engineering, Carl Zeiss Meditec and Optovue (Optovue, Fremont, CA, USA).

980 DFG-H, AQ, PP and HZ are partly funded by the NIHR Biomedical Research Centre based at Moorfields Eye  
981 Hospital and UCL Institute of Ophthalmology.

982 DFG-H's chair at UCL is supported by funding from the International Glaucoma Association.

983 The views expressed are those of the authors and not necessarily those of the National Health Service, the  
984 National Institute for Health Research, or the Department of Health.

985

986 **Contributions of authors:**

987 Design and conduct of study (DGH, DPC, HZ); analysis and interpretation (DGH, AQ, PP, QC, HZ); writing  
988 the article (DGH); critical revision and approval of manuscript (DGH, AQ, PP, DPC, QC, HZ); data collection  
989 (DGH, AQ); statistical expertise (DGH, AQ, PP, DPC, QC, HZ); obtaining funding (DGH, DPC, HZ); literature  
990 search (DGH)

991 The authors would like to thank Dr Tuan Ho for his administrative support for the study.

992 **Disclosures**

993 Funding support: NIHR (DGH, AQ, PP, DPC, HZ), Industry (Pfizer) through employer (AQ, PP), Industry  
994 (Alcon, Pfizer, Santen) through employer (DGH),

995 Financial disclosures: DGH (consulting fees Aerie, Alcon, Alimera, Allergan, CenterVue, Pfizer, Quark,  
996 Quethera, Roche, Santen, Santhera, Sensimed; Lecture fees Santen, Topcon); DPC (Lecture fees Allergan)

997 Pending patent: ANSWERS (DGH, DPC, HZ)

998





	Visit 1 Mont h0	Visit 2 Mont h2	Visit 3 Mont h4	Visit 4 Mont h7	Visit 5 Month 10	Visit 6 Mont h13	Visit 7 Mont h16	Visit 8 Month 18	Visit 9 Mont h20	Visit 10 Month 22	Visit 11 Month 24
Visual Fields	2	2	1	1	1	1	2	2	1	1	2
HRT	3	2	1	1	1	1	2	3	1	1	1
Optic disc photography	1	1	1	1	1	1	1	1	1	1	1
GDxVCC	3	2	1	1	1	1	2	3	1	1	1
OCT	5	3	3	3	3	3	3	5	3	3	5

Table 1: Schedule of visual field testing and imaging; number of tests/images at each visit (HRT: Heidelberg retina tomography, VCC: variable cornea compensation, OCT: optical coherence tomography)

	Placebo (n = 178 participants; 264 eyes)		Latanoprost (n = 183 participants; 264 eyes)	
	Median	5 <sup>th</sup> to 95 <sup>th</sup> percentile	Median	5 <sup>th</sup> to 95 <sup>th</sup> percentile
Age (years)	66.3	47.3 – 81.1	65.7	44.7 – 79.6
IOP (mmHg)	19.0	12.0 – 28.0	19.0	12.5 – 27.0
SAP MD (dB)	-2.73	-10.60 – -0.17	-2.57	-10.98 – -0.02
RNFL thickness ( $\mu$ )	75.3	48.2 – 106.6	77.2	56.1 – 101.3
Visual acuity (Snellen)	6/6	6/5 – 6/9	6/6	6/5 – 6/12
Refractive error (D)	0.00	-6.85 – 3.13	-0.13	-6.13 – 2.29
	Number	%	Number	%
Sex (female)	86	48	79	43
Ethnic origin				
White	153	86	165	90
Black	15	8	8	4
Indian subcontinent	4	2	8	4
Other/unknown	6	3	2	1

Table 2. Principal baseline characteristics for the subset of the UK Glaucoma Treatment Study cohort with OCT images

Age, sex and ethnic origin are subject variables; IOP and SAP MD and RNFL thickness are eye variables. Data are provided for eligible eyes.

D = diopters; dB = decibel; mmHg = millimetres of mercury; IOP = baseline (pre-treatment) intraocular pressure; MD = baseline (visit 1) mean deviation; SAP = standard automated perimetry

	(n = 72 participants; 114 eyes)	
	Median	5 <sup>th</sup> to 95 <sup>th</sup> percentile
Age (years)	70.3	50.0 – 85.6
IOP (mmHg)	14	8.0 – 21.0
SAP MD (dB)	-4.17	-14.22 – 0.88
RNFL thickness ( $\mu$ )	69.0	45.1 – 95.6
Visual acuity (Snellen)	6/6	6/4 – 6/12
Refractive error (D)	-0.13	-7.48 – 2.95
	Number	%
Sex (female)	42	58
Ethnic origin		
White	48	67
Black	16	22
Indian subcontinent	4	6
Other/unknown	4	6

Table 3. Principal baseline characteristics for the ‘RAPID’ test retest cohort

Age, sex and ethnic origin are subject variables; IOP and SAP MD and RNFL thickness are eye variables. Data are provided for eligible eyes.

D = diopters; dB = decibel; mmHg = millimetres of mercury; IOP = intraocular pressure; MD = mean deviation; SAP = standard automated perimetry

Parameter	Estimate	95% confidence interval	p-value
Constant	-4.33	(-4.87 to -3.8)	<0.001
time	-0.34	(-0.5 to -0.18)	<0.001
latanoprost	0.61	(-0.16 to 1.37)	0.12
time x latanoprost	0.38	(0.16 to 0.61)	0.001
intercept variance	10.39	(8.77 to 12.31)	
time variance	0.54	(0.41 to 0.72)	
intercept-time covariance	0.59	(0.22 to 0.95)	
Within individual variance	1.33	(1.26 to 1.39)	

Table 4: Estimates of rate of change in visual field mean deviation allowing interaction with intervention groups, for patients eligible for the OCT analysis

Parameter	Estimate	95% confidence interval
Placebo intercept	-4.33	(-4.87 to -3.8)
Placebo slope	-0.34	(-0.5 to -0.18)
Latanoprost intercept	-3.73	(-4.27 to -3.19)
Latanoprost slope	0.05	(-0.11 to 0.2)

Table 5: Visual field mean deviation intercept and slope by intervention

Parameter	Estimate	95% confidence interval	p- value
Constant	75.19	(72.8 to 77.58)	<0.001
time	-1.70	(-2.27 to -1.12)	<0.001
latanoprost	1.58	(-1.81 to 4.97)	0.36
time x latanoprost	0.60	(-0.2 to 1.4)	0.14
intercept variance	210.00	(177.83 to 247.99)	
time variance	8.18	(6.41 to 10.43)	
intercept-time covariance	2.38	(-3.43 to 8.2)	
Within individual variance	16.89	(16.32 to 17.49)	

Table 6: Estimates of rate of change in average retinal nerve fiber layer thickness allowing interaction with intervention groups

Parameter	Estimate	95% confidence interval
Placebo intercept	75.19	(72.8 to 77.58)
Placebo slope	-1.7	(-2.27 to -1.12)
Intervention intercept	76.77	(74.36 to 79.17)
intervention slope	-1.1	(-1.65 to -0.54)

Table 7: Retinal nerve fiber layer thickness intercept and slope by intervention

Covariate	b	SE	Wald	P	Exp(b)	95% CI of Exp(b)
Age	0.01885	0.01357	1.9309	0.1647	1.0190	0.9923 to 1.0465
Allocation	-0.7446	0.2865	6.7547	0.0094	0.4749	0.2709 to 0.8327
IOP	0.05189	0.02872	3.2655	0.0708	1.0533	0.9956 to 1.1142
mean_MD	0.08614	0.04930	3.0533	0.0806	1.0900	0.9896 to 1.2005
OCT_RNFL_slope	-0.07104	0.03952	3.2315	0.0722	0.9314	0.8620 to 1.0064

Table 8: Cox proportional hazards model for progression-free survival

Observation period	Definitive trial	Pilot study
<b>18 months</b>	353	207
<b>12 months</b>	502	294

Table 9. Sample size calculation for a placebo-controlled study, with an effect size of that observed for latanoprost in the UK Glaucoma Treatment Study (includes 10% initial loss to follow-up and additional participant attrition of 0.5% per week)

Observation period	Definitive trial	Pilot study
<b>18 months</b>	440	257
<b>12 months</b>	633	371

Table 10. Sample size calculation for a study comparing an intervention half as effective as latanoprost with an intervention with an effect size equivalent to latanoprost (includes 10% initial loss to follow-up and additional participant attrition of 0.5% per week)

Observation period	Definitive trial	Pilot study
<b>18 months</b>	2552	1502
<b>12 months</b>	3689	2171

Table 11. Sample size calculation for a study comparing an intervention 75% as effective as latanoprost (group 0) with an intervention with an effect size equivalent to latanoprost (includes 10% initial loss to follow-up and additional participant attrition of 0.5% per week)

Observation period	Definitive trial	Pilot study
<b>18 months</b>	601	352
<b>12 months</b>	878	515

Table 12. Sample size calculation for a study comparing an intervention with an effect size equivalent to latanoprost with a combination treatment with an effect size equivalent to 2\*latanoprost (includes 10% initial loss to follow-up and additional participant attrition of 0.5% per week)

Observation period	Definitive trial	Pilot study
<b>18 months</b>	3029	1783
<b>12 months</b>	4417	2599

Table 13. Sample size calculation for a study comparing an intervention with an effect size equivalent to latanoprost with a combination treatment with a combination treatment with an effect size equivalent to 1.5\*latanoprost (includes 10% initial loss to follow-up and additional participant attrition of 0.5% per week)

## FIGURES AND LEGENDS

Figure 1: Flow chart for subject and test data selection. Each OCT scan is comprised of 3 peripapillary sweeps; for the purpose of this analysis, each sweep is counted as an image.

Figure 2: Visual field data structure for the growth curve models

Figure 3: OCT data structure for the growth curve models

Figure 5: Distribution of the rates of visual field mean deviation change for the subset of UK Glaucoma Treatment Study participants with OCT images (placebo, 143 participants; latanoprost, 141 participants)

Figure 6: Distribution of the rates of optical coherence tomography retinal nerve fiber layer thickness change for the subset of UK Glaucoma Treatment Study participants with OCT images (placebo, 143 participants; latanoprost, 141 participants)

Figure 7: Kaplan-Meier survival curves for the subset of UK Glaucoma Treatment Study participants with OCT images applying the Guided Progression Analysis criterion for progression.

Figure 8: The ‘hit rate’ is the proportion of UK Glaucoma Treatment Study participants identified as deteriorating at criterion false positive rates between 0 and 15%. Analyses are shown for ANSWERS, PoPLR and sANSWERS models. Data are shown for series intervals (baseline to final observation) of up to 7, 13, 18 and 22 months. The shorter series are a subset of the longer series, so that an eye identified as ‘progressed’ earlier in the series is carried forward as ‘progressed’ in the later series. Data are shown for 445 eyes of 353 participants.

Figure 9: Kaplan-Meier survival curves for the subset of UK Glaucoma Treatment Study participants with OCT images applying the ANSWERS criterion for progression.

Figure 10: Kaplan-Meier survival curves for the subset of UK Glaucoma Treatment Study participants with OCT images applying the PoPLR criterion for progression.

Figure 11: Kaplan-Meier survival curves for the subset of UK Glaucoma Treatment Study participants with OCT images applying the structure-guided ANSWERS (sANSWERS) criterion for progression.

Figure 12: Kaplan-Meier survival curves for the subset of UK Glaucoma Treatment Study participants with OCT images applying the ‘ANSWERS AND PoPLR’ criterion for progression.

Figure 13: Venn diagram illustrating the agreement for UK Glaucoma Treatment Study participants identified as progressing by Guided Progression Analysis, ANSWERS and PoPLR criteria for progression. The numbers represent the number of participants in each category.

Figure 14: illustration of a Weibull probability density function ( $\kappa=0.5$ ,  $\lambda=1$ )

Figure 1

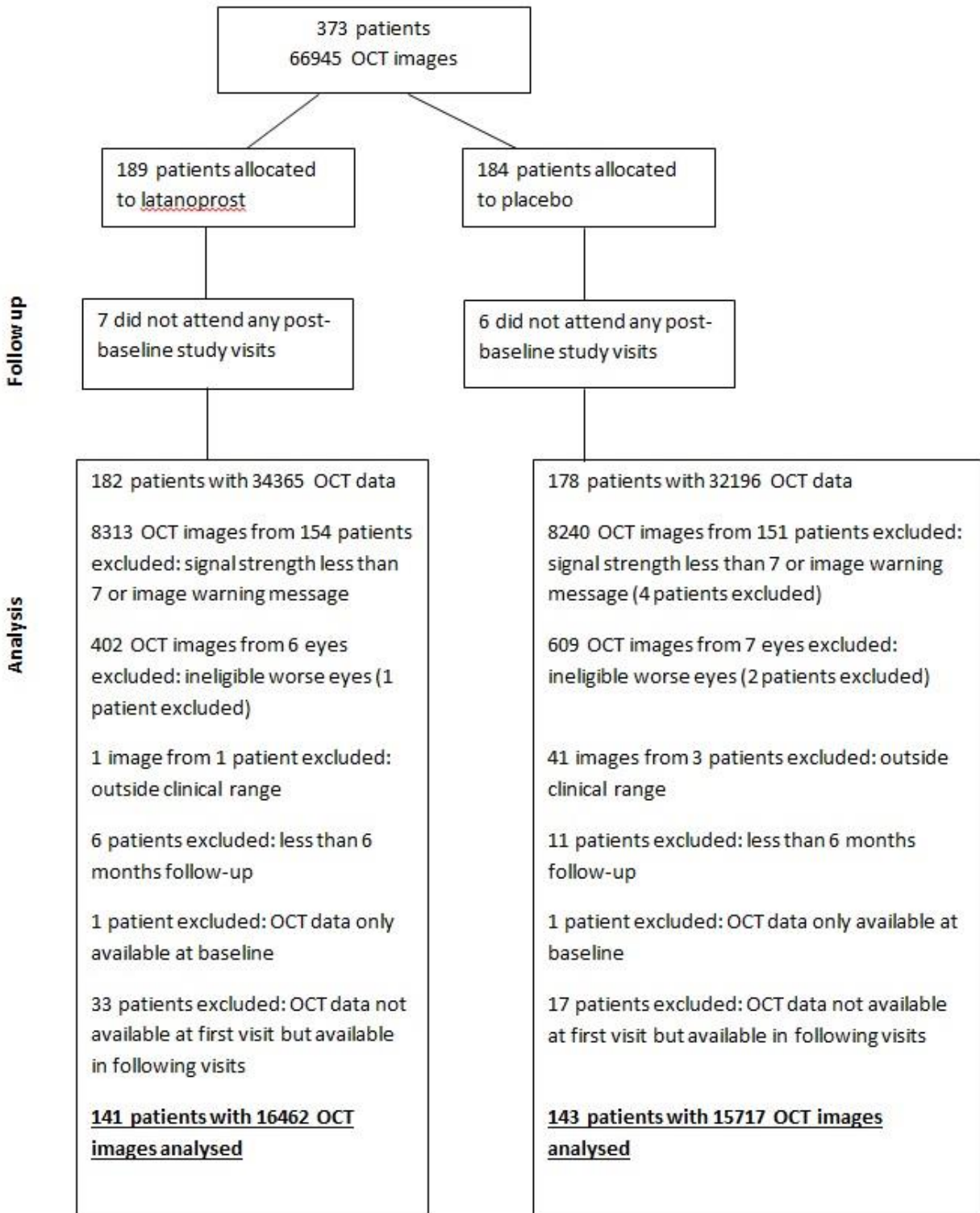
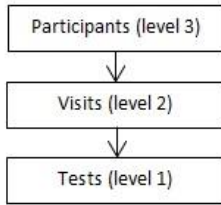




Figure 2

a: original data structure



b: Data restructured

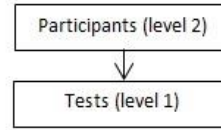
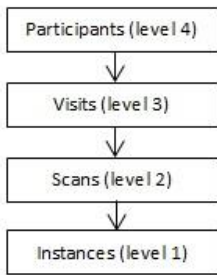


Figure 3

a: original data structure



b: Data restructured

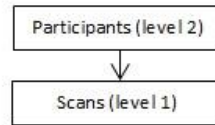


Figure 4

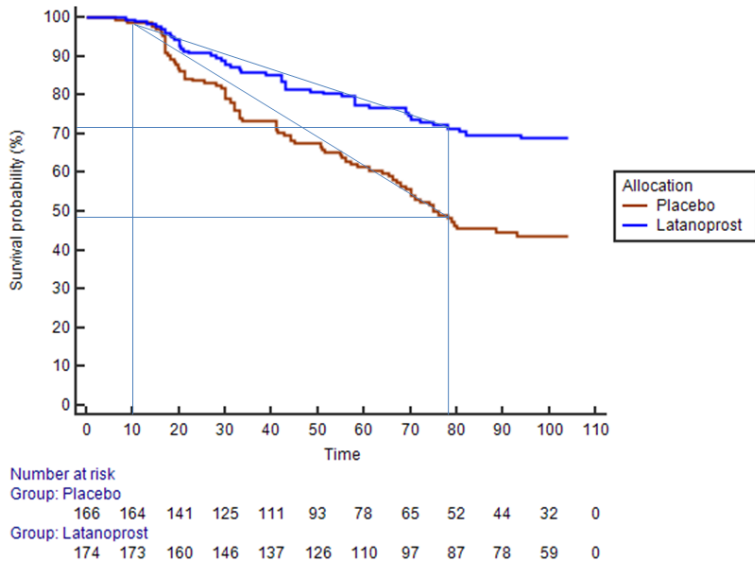


Figure 5

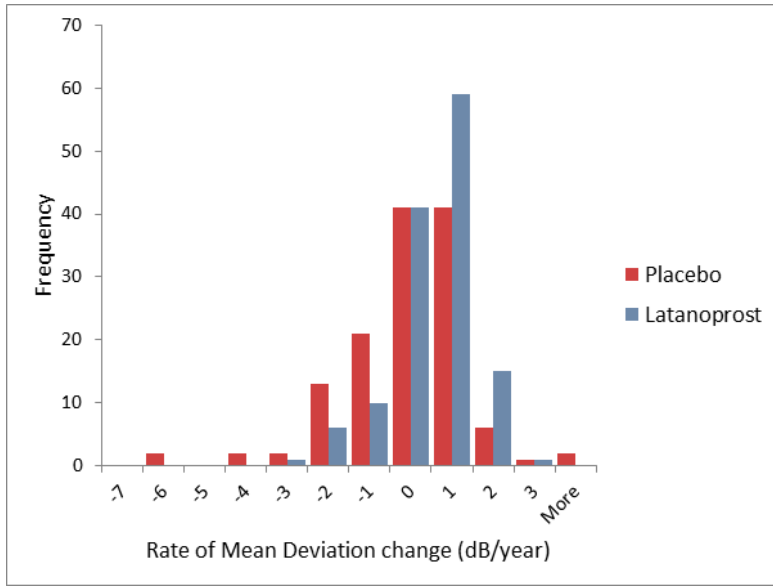


Figure 6

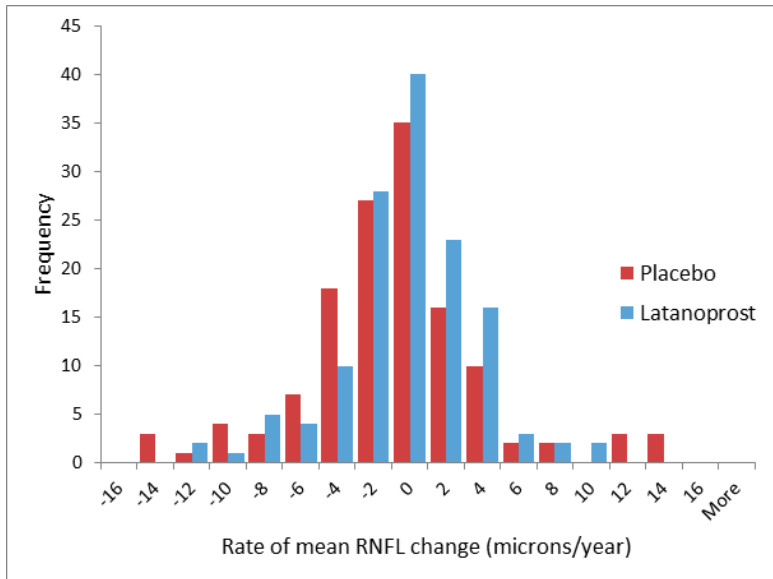
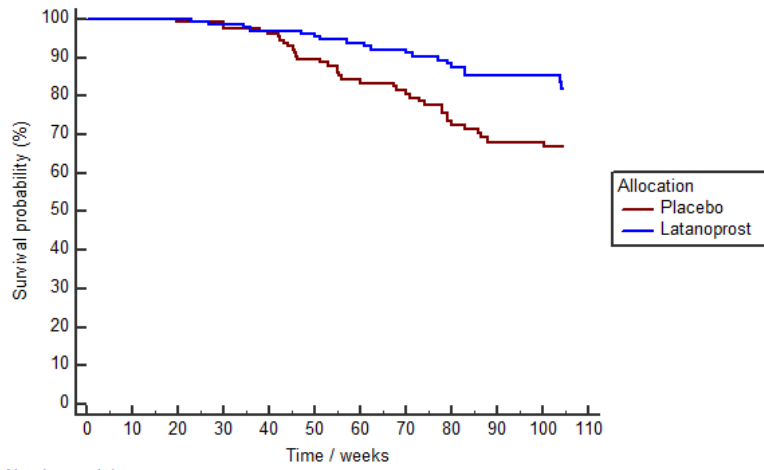


Figure 7



Number at risk

Time (weeks)	0	10	20	30	40	50	60	70	80	90	100	110
Group: Placebo	132	132	131	127	122	105	91	85	71	59	52	0
Group: Latanoprost	134	134	134	131	126	120	112	102	90	84	68	0

Figure 8

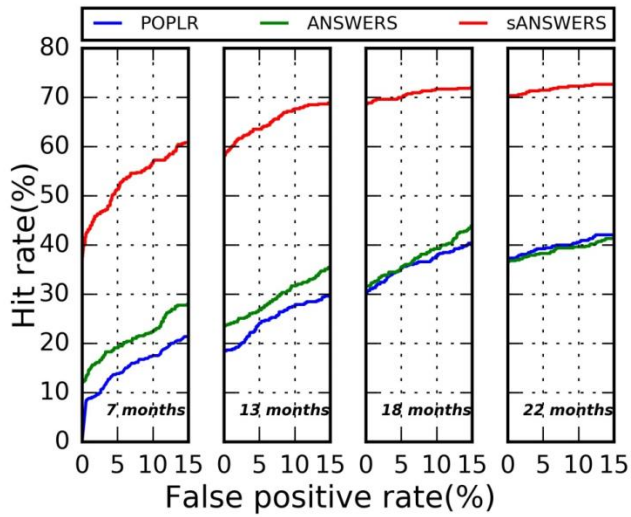
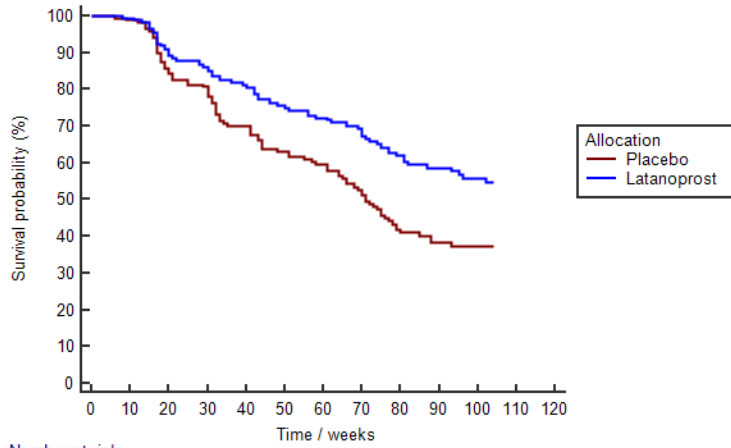


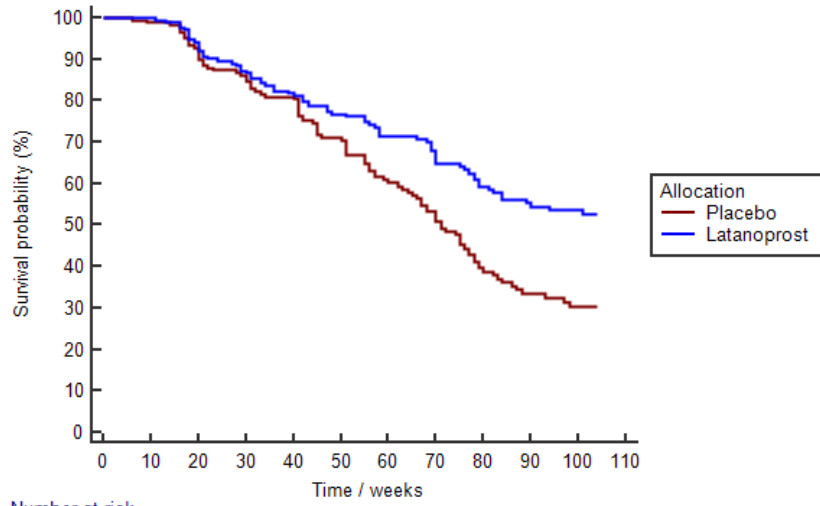
Figure 9



Number at risk

Time / weeks	0	10	20	30	40	50	60	70	80	90	100	110	120
Group: Placebo	166	164	139	126	109	91	81	66	50	42	30	0	0
Group: Latanoprost	174	173	153	140	129	117	105	91	77	67	49	0	0

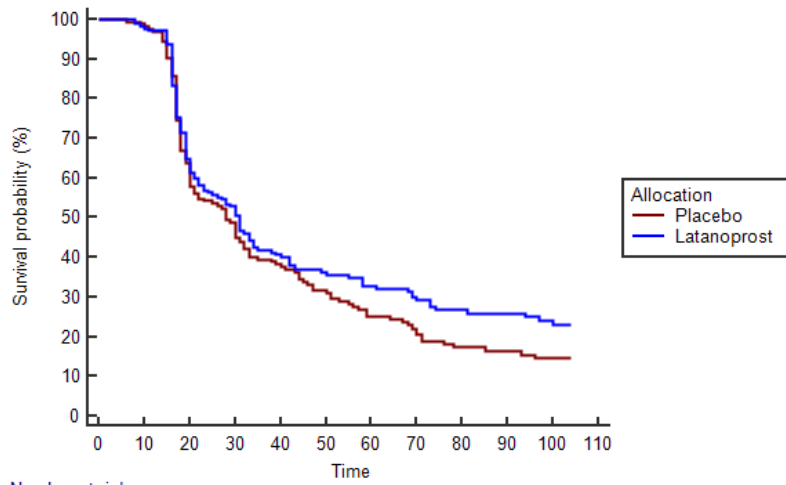
Figure 10



Number at risk

Time / weeks	0	10	20	30	40	50	60	70	80	90	100	110
Group: Placebo	166	164	147	134	123	98	79	64	47	35	25	0
Group: Latanoprost	174	174	158	144	132	121	102	87	75	65	51	0

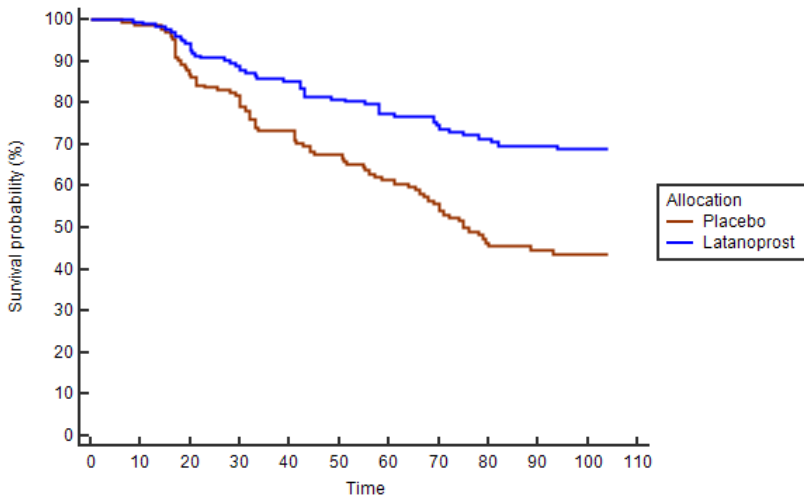
Figure 11



Number at risk

Time	0	10	20	30	40	50	60	70	80	90	100	110
Group: Placebo	166	163	96	73	59	44	33	26	20	18	13	0
Group: Latanoprost	174	170	107	83	64	56	48	37	31	29	24	0

Figure 12



Number at risk

Time	0	10	20	30	40	50	60	70	80	90	100	110
Group: Placebo	166	164	141	125	111	93	78	65	52	44	32	0
Group: Latanoprost	174	173	160	146	137	126	110	97	87	78	59	0

Figure 13

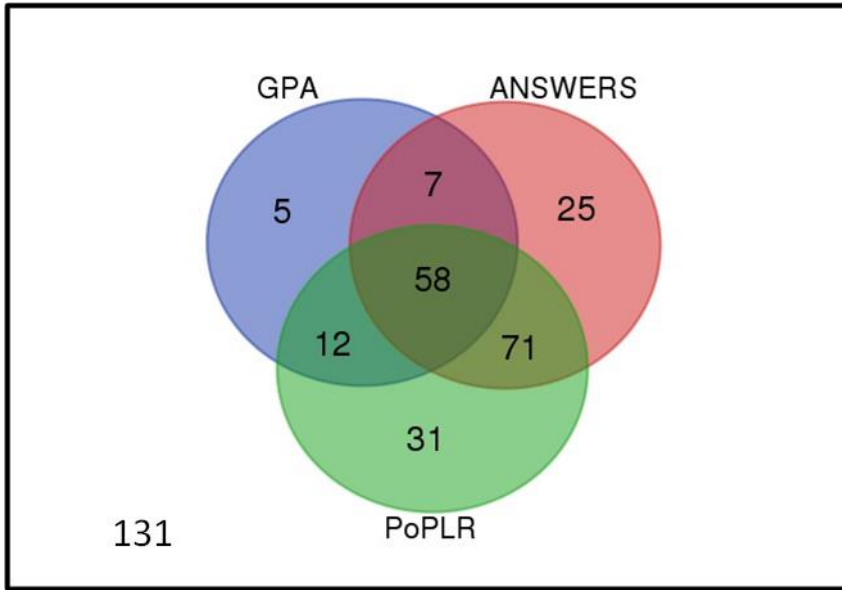


Figure 14

