Combining optical coherence tomography with visual field data to rapidly detect disease progression in glaucoma: a diagnostic accuracy study

David F Garway-Heath,[1]* Haogang Zhu,[1,2,3] Qian Cheng,[3] Katy Morgan,[4] Chris Frost,[4] David P Crabb,[2] Tuan-Anh Ho[1] and Yannis Agiomyrgiannakis[5]

[1]National Institute for Health Research (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, London, UK
[3]School of Computer Science and Engineering, Beihang University, Beijing, China
[4]Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, UK
[5]Patient representative

*Corresponding author david.garway-heath@moorfields.nhs.uk

Declared competing interests of authors: David Garway-Heath has received consulting fees from Aerie Pharmaceuticals Inc., Alcon, Alimera Sciences, Inc., Allergan, CenterVue Inc., Pfizer Inc., Quark Pharmaceuticals, Quethera Ltd, F Hoffman-La Roche Ltd, Santen Pharmaceutical Co., Ltd, Santhera Pharmaceuticals and Sensimed AG, a grant from Pfizer Inc. and lecture fees from Heidelberg Engineering Ltd, Santen Pharmaceutical Co., Ltd and Topcon Corporation and his institution has received equipment loans from Carl Zeiss Meditec AG, Heidelberg Engineering Ltd and Optovue, Inc. He is also a member of the HTA Clinical Evalution and Trials Board. David P Crabb has received lecture fees from Allergan, F Hoffman-La Roche Ltd and Santen Pharmaceutical Co., Ltd and consulting fees from Allergan and his institution has received unrestricted research funds from Allergan, CenterVue Inc., Novartis UK, F Hoffman-La Roche Ltd and Santen Pharmaceutical Co., Ltd. He has also provided expert testimony for the Driving and Vehicle Licensing Agency. Tuan-Anh Ho has received salary from the National Institute for Health Research (NIHR) Biomedical Research Centre at Moorfields Eye Hospital and UCL Institute of Ophthalmology and consultancy fees from Allergan. David Garway-Heath, David P Crabb, Qian Cheng and Haogang Zhu have a patent application filed for ANSWERS (a method of data analysis evaluated in this work).

Published January 2018
DOI: 10.3310/hta22040

Scientific summary

Imaging technology to detect glaucoma disease progression

Health Technology Assessment 2018; Vol. 22: No. 4
DOI: 10.3310/hta22040

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Glaucoma is a chronic progressive eye disease that can cause irreversible vision loss. The optic nerve is damaged where it enters the eye, resulting in reduced sensitivity to light in regions of the eye's field of vision. In clinical care and in clinical trials, light sensitivity is measured with the visual field (VF) test. VF measurements are, however, variable and the variability increases as damage worsens, making it difficult to identify glaucoma worsening over time. The damage to the nerve can also be measured with imaging techniques, such as optical coherence tomography (OCT), which measures the thickness of the retinal nerve fibre layer (RNFL); the RNFL contains the retinal ganglion cell axons, which leave the eye through the optic nerve head. OCT RNFL and VF measurements have been shown to correlate over the range of glaucoma damage. Combining VF and OCT RNFL measurements may reduce variability, making it easier to identify glaucoma worsening. To establish the validity of combining VF and OCT RNFL measurements, it should first be demonstrated that treatment slows the rate of RNFL thinning to a similar extent that it slows the rate of VF loss.

Methods

We aimed to compare statistical methods that combine VF and OCT data with the reference standard method [Guided Progression Analysis™ (GPA) software (Carl Zeiss Meditec Inc., Dublin, CA, USA) for the Humphrey Field Analyzer (HFA) instrument™ (Carl Zeiss Meditec Inc., Dublin, CA, USA)], which uses only VF data, to establish whether or not combining OCT and VF allows (1) more rapid identification of glaucoma worsening ('progression') and (2) shorter or smaller clinical trials. We also aimed to explore new statistical methods for combining VF and OCT data.

As there is no 'gold standard' test for glaucoma progression to provide a ground truth, relative measures are required. Instead of criterion sensitivity, the 'hit rate' (proportion of eyes identified as progressing) in eyes at risk of worsening was used as an approximation. Criterion specificity was measured in eyes with a very low probability of worsening (clinically stable patients measured frequently over a space of time too short for clinically relevant deterioration to take place). Time to progression, when specificity was fixed at 95%, was used as another measure of test sensitivity. Other metrics to establish the utility of combining VF and OCT data included the accuracy of the estimated rate of progression. Again, as there is no gold standard measurement, a surrogate outcome was used; the modelled rate of progression over the initial five visits was used to predict the last VF in the series, assuming a linear rate of change, and the prediction error was taken as a measure of the model appropriateness. Finally, the ability of the various models to distinguish the treatment arms in clinical trial data was assessed.

The hit rate, time to progression, prediction accuracy and ability to distinguish treatment status of the various statistical methods was evaluated in the 320 participants, or subsets of them, from the United Kingdom Glaucoma Treatment Study (UKGTS) multicentre randomised placebo-controlled clinical trial who had both VF testing and OCT RNFL imaging. Specificity was evaluated in up to 72 stable glaucoma patients who had between 4 and 14 VF and OCT tests within a 3-month period (the RAPID stable data set). The UKGTS was conducted at 10 teaching and general ophthalmology units. The RAPID data set was collected at a single teaching ophthalmology unit. The UKGTS participants were newly diagnosed patients with mild-to-moderate glaucoma (mean deviation better than –16 dB in the worse eye) randomised to a drop therapy to lower intraocular pressure or placebo. RAPID participants were patients with similar glaucoma severity who were on treatment and who were clinically stable.

Participants underwent VF testing with the HFA 24-2 test pattern and RNFL imaging with time-domain Stratus OCT™ (Carl Zeiss Meditec Inc., Dublin, CA, USA) (TD OCT). The reference test for glaucoma progression was based on the GPA software of the HFA. Index tests were based on previously described methods [Analysis with Non-Stationary Weibull Error Regression and Spatial enhancement (ANSWERS) and permutation analyses of pointwise linear regression (PoPLR); ANSWERS was modified so that the rate of RNFL change could be used as a Bayesian prior for the estimates of the rate of VF change and was termed 'structure-guided ANSWERS' (sANSWERS)]. Other index tests were newly developed methods based on permutation tests, multivariate hierarchical models with multiple imputation for censored values (MaHMIC) and multivariate generalised estimating equations with multiple imputation for censored values (MaGIC).

The main outcome measures were progression criterion specificity, hit rate, time to incident progression, future VF sensitivity prediction error, difference in proportions identified as progressing in the UKGTS treatment groups, hazard ratios (HRs) and study sample size required to measure treatment effects.

Results

The estimated criterion specificity was set at 95% for all tests. The hit rate in the UKGTS cohort for the various statistical methods was 22.2% for GPA, 41.6% for PoPLR, 53.8% for ANSWERS and 61.3% for sANSWERS; all pairs of comparison were significantly different at $p \leq 0.042$. Mean survival time was 93.6 weeks for GPA, 82.5 weeks for PoPLR, 72.0 weeks for ANSWERS and 69.1 weeks for sANSWERS. The trend in VF (±OCT RNFL) measurements over the initial 42.4 [standard deviation (SD) 6.2] weeks was used to predict the VF sensitivity values 49.2 (SD 19.8) weeks later; the median prediction errors were 3.8 (5th to 95th percentile 1.7 to 7.6) dB for PoPLR, 3.0 (5th to 95th percentile 1.5 to 5.7) dB for ANSWERS and 2.3 (5th to 95th percentile 1.3 to 4.5) dB for sANSWERS. In distinguishing the UKGTS treatment groups, the HRs were 0.57 [95% confidence interval (CI) 0.34 to 0.90; $p = 0.016$] for GPA, 0.59 (95% CI 0.42 to 0.83; $p = 0.002$) for PoPLR, 0.76 (95% CI 0.56 to 1.02; $p = 0.065$) for ANSWERS and 0.70 (95% CI 0.53 to 0.93; $p = 0.012$) for sANSWERS. Sample size estimates were not reduced by using methods including OCT data.

Permutation test analysis of UKGTS data resulted in hit rates between 8.3% and 17.4%; treatment effects when data were analysed with MaHMIC and MaGIC were non-significant and statistical significance was altered little by incorporating imaging.

Conclusions

The sANSWERS method combining VF and OCT data had a higher hit rate and identified progression more quickly than the reference GPA method and other VF-only methods, and produced more accurate estimates of the rate of progression. However, methods combining VF and OCT data did not improve trial power to identify a treatment effect. The statistical method providing the greatest difference in time to progression and most statistically significant difference was the PoPLR technique using VF data alone. Current OCT imaging technology is already more precise than that evaluated in this work (TD OCT). It is likely, therefore, that current OCT technology would perform better than TD OCT. The size of the RAPID data set limited the precision of the estimates for criterion specificity; however, 'stable' data sets, in which many tests are obtained over a short period of time, are challenging to collect. Future work should evaluate current OCT technology in the context of clinical treatment trials and refine the statistical methods further.

Trial registration

This trial is registered as ISRCTN96423140.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research (NIHR). Data analysed during the study were from the UKGTS. Funding for the UKGTS was provided through an unrestricted investigator-initiated research grant from Pfizer Inc. (New York, NY, USA), with supplementary funding from the NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK. Imaging equipment loans were made by Heidelberg Engineering, Carl Zeiss Meditec and Optovue (Fremont, CA, USA). Pfizer, Heidelberg Engineering, Carl Zeiss Meditec and Optovue had no input into the design, conduct, analysis or reporting of any of the UKGTS or this work. The sponsor for both the UKGTS and RAPID data collection was Moorfields Eye Hospital NHS Foundation Trust. David F Garway-Heath, Tuan-Anh Ho and Haogang Zhu are partly funded by the NIHR Biomedical Research Centre based at Moorfields Eye Hospital and UCL Institute of Ophthalmology. David F Garway-Heath's chair at University College London (UCL) is supported by funding from the International Glaucoma Association.

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.236

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 11/129/245. The contractual start date was in December 2013. The draft report began editorial review in June 2017 and was accepted for publication in October 2017. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).