Toward a new definition of glaucomatous optic neuropathy for clinical research

Jayant Iyer[a,b], *Jayme R. Vianna*[c], *Balwantray C. Chauhan*[c], *and Harry A. Quigley*[a]

Purpose of review
A process is ongoing to produce a definition of glaucomatous optic neuropathy (GON) using quantitative, objective data from structural and functional tests. At present, a common practice is to define GON by subjective features said to be 'characteristic' as judged by those experienced in glaucoma care.

Recent findings
An objective definition would standardize the comparison of clinical research results across studies, without precluding simultaneous use of idiosyncratic definitions in the same reports. To achieve this goal, expert opinion was solicited to reach optimal agreement on one or more consensus, GON definitions. An interactive period of online discussion by 176 international experts led to 110 responses in an online survey that narrowed possible definitional structures into testable criteria.

Summary
Two approaches to validation of one or more sets of criteria for definite and possible GON are ongoing. The general principles include definition for each eye individually, inclusion of a borderline category, no intraocular pressure criterion, and both structural and functional defects in appropriate physical locations. Each validation approach uses clinician diagnosis as a standard against which objective criteria are compared, with the initial approach using a three-level categorical scale, and the second approach using 0–100 scaling.

Keywords
definition, glaucoma, optical coherence tomography, research, visual field

INTRODUCTION

There is currently no objective definition of glaucomatous optic neuropathy

The development of accurate tonometry and gonioscopy in the mid-20th century enabled a better understanding of the nature of the various conditions described as glaucoma. Vision loss associated with both primary forms of glaucoma, open and closed angle, involves structural remodeling of the optic nerve head and progressive loss of peripheral vision function beginning in the mid-field. Stereophotography of the optic disc and perimetry with the Goldmann instrument gave more concrete features of GON. However, the broad variations in the disc size and appearance, and the lack of a normative database for manual fields remained impediments to an objective definition. Initial population-based studies [1,2] defined glaucoma with subjective disc assessment and kinetic field tests and indicated that open angle glaucoma (OAG) occurs at all levels of intraocular pressure (IOP).

At a 1991 World Health Organization Collaborating Vision Centres, glaucoma was not listed on official publications as a cause of blindness, since, as one expert stated: 'Glaucoma cannot be defined or treated, so it isn't on the list'. To identify the importance of glaucoma for public health planning, data from existing population-based prevalence surveys were analysed in 1996 [3], using standard criteria applied to studies in which an optic disc exam and visual field test had been performed on every person,

[a]Glaucoma Center of Excellence, Wilmer Institute, Johns Hopkins, Baltimore, Maryland, USA, [b]Singapore National Eye Centre, Singapore, Singapore and [c]Department of Ophthalmology and Visual Sciences, Dalhousie University, Halifax, Nova Scotia, Canada

Correspondence to Harry A. Quigley, MD, Wilmer 122, Johns Hopkins Hospital, Baltimore, MD 21287, USA. Tel: +1 410 955 2777; fax: +1 410 955 2542; e-mail: hquigley@jhmi.edu

*Supported in part by contributions to the Glaucoma Center of Excellence, Johns Hopkins.

Curr Opin Ophthalmol 2020, 31:85–90

DOI:10.1097/ICU.0000000000000644

KEY POINTS

- Current definitions of GON are subjective.
- An objective GON definition in OAG will allow standardized comparisons across research studies.
- Initial objective parameters are being validated against the current gold standard: clinician opinion.
- Further validation methods may include use of longitudinal databases and artificial intelligence.

to determine the prevalence of OAG and angle closure glaucoma (ACG). This permitted the first recognition that glaucoma is the second leading cause of world blindness. It applied a defined set of structural and functional alterations, GON, that were characteristic of the damage found in a variety of clinical glaucoma entities, both primary and secondary.

It became clear that better specification of OAG prevalence, incidence [4], and rates of progression and blindness required more research using a definition of GON with objective criteria. A major step in producing objective data for visual field assessment was the automation of threshold perimetry. Focus groups and expert opinion contributed to a definition of OAG and ACG proposed by a consensus panel and designed to be used in prevalence surveys to standardize comparisons in epidemiological research [5▪▪]. Subsequently, many more prevalence surveys were conducted in Europe, Asia, Africa, and among Hispanic persons, using this standardized definition of GON to permit comparable estimates of worldwide OAG and ACG prevalence [6]. The Foster *et al.* paper detailing the definition of glaucoma had been cited 1350 times by 2017, comprising 10% of all papers in which the subject is OAG [7].

A recognized weakness of the Foster classification was the use of the cup/disc ratio, as a surrogate for GON structural damage. This issue was solved by the application of quantitative image analysis methods that included, but moved beyond cup/disc ratio, now most commonly utilizing optical coherence tomography (OCT) for glaucoma diagnosis [8]. Its robust diagnostic ability permits quantification of the structural criteria for GON and could replace definitional use of cup/disc ratio. Yet, GON continues to be defined in publications by subjective descriptions, such as the presence of 'characteristic glaucoma optic disc and field change' despite the proven inadequacy of interobserver and intraobserver grading of the disc or field [9,10]. It is now appropriate to replace clinical examination with the most sensitive and specific objective criteria for GON upon which we can agree. The subsequent discussion and validation are limited to a GON definition for primary OAG, because ACG-produced injury could be sufficiently different to require separate criteria.

THE PRESENT PROCESS

In 2016, investigators at the Wilmer Glaucoma Center of Excellence considered an approach to develop a new GON definition for clinical glaucoma research. Although the previous definition was designed for prevalence surveys often performed in rural areas, where the newest technology was unavailable, modern clinical research is carried out in centers where both OCT and automated perimetry are routinely used. The purpose of this activity was to produce a consensus definition of GON to facilitate comparability among clinical research investigations.

The procedures to gather initial information were based on the Delphi method [11] in which experts are posed questions and respond in two or more rounds. After each round, a facilitator provides a summary of the expert opinions from the previous round and solicits further comments. Experts revise their initial answers in response to other members of the group with the intent to narrow the range of answers to converge towards consensus. The members of all world glaucoma societies were invited to comment, with participants from the American, European, Asia-Pacific, Japanese, Korean, Australian—New Zealand, and Latin American Glaucoma Societies. In all, 176 experts contributed online, with all comments available to each member

SURVEY RESULTS AFTER DISCUSSION

After two rounds of discussion, 18 survey questions were administered online through Qualtrix. The 110 respondents reached consensus on several features of a definition (percentage in agreement given in parenthesis).

(1) The definition will be applied to each eye of a person separately (100%).
(2) The clinical examination of the optic disc and retina is needed to rule out conditions that would simulate glaucoma (81%).
(3) The purpose of the clinical exam is to exclude nonglaucoma disease (48%) or to assure that glaucoma is present (inclusionary) (52%).
(4) The clinical exam should be documented with photography (54%).
(5) IOP should not be a criterion for diagnosis (74%).

(6) The criteria should try to include early GON (78%).
(7) OCT nerve fiber layer (NFL) thickness should be included either alone (40%) or with segmented macular thickness (60%).
(8) OCT segmented macular thickness alone is not a sufficient structure criterion (99%).
(9) OCT rim width cannot replace the need for a clinical exam (98%) because of large variations in disc and rim configurations, particularly in myopic eyes.
(10) Both OCT and field are needed for definite GON (37%).
(11) GON may be present if either OCT or field are abnormal (63%).
(12) The visual field test standard is the 24-2 program (77%), whereas 20% favored including both 24-2 and 10-2 tests.
(13) The OCT defect must be in the appropriate, opposite hemifield from the field defect (64%).
(14) Progressive change in OCT or field is not required for GON definition (78%).
(15) Progressive change, if it can be suitably defined, qualifies an eye for GON even if static criteria are not met (85%).

IMPORTANT ISSUES RAISED AFTER THE SURVEY

Progression

Although it was proposed that progressive change in OCT or field is an important feature of GON, pragmatically, longitudinal data were considered unlikely to be available for many subjects who would otherwise qualify to participate in glaucoma research. Furthermore, this would require a definition of progressive worsening, for which there also are no agreed criteria. Findings at a single point in time could thus be considered the standard, until further analysis of longitudinal databases or in prospective studies.

Optical coherence tomography

Quantitative OCT parameters with comparison to normative data that could be considered are nerve fiber layer and segmented macular thicknesses, disc size, rim width/area as delineated by Bruch's membrane opening, and quality measures of the image. Any definitional use of OCT would need to take into account differences among instruments made by different companies in data acquired, methods of acquisition, normative database variation, and statistical analysis. To be included, each instrument would have to provide a comparable set of data to analyze and validate.

NFL or retinal thickness loss represent important measures of GON damage, but are not specific enough, as many retinal and neurological conditions also share NFL loss. GON is distinguished from other anterior optic neuropathies by laminar deformation leading to deepening and widening of the cup [12,13]. Yet, the past 'gold standard' measure of disc topography – cup to disc ratio – did not take into account variation due to disc diameter and obliquity of the nerve head, particularly problematic in myopic eyes. Minimum rim width has much to offer as a specific feature, especially if classifications of normality are adjusted for Brunch's membrane opening area. The method to determine rim width differs among present OCT instruments.

Nerve fiber layer or segmented thickness data

The discussion compared such variables as thickness at the set distance from the disc compared to using the pattern of loss (e.g. 'wedge'-shaped). The former has much to offer in simplicity, whereas quantitative definitions of 'pattern' would be needed to eliminate subjectivity.

Optical coherence tomography parameters

The superior and inferior temporal disc/NFL are more specific for glaucoma damage, whereas the temporal and nasal disc cup/rim/NFL are not only nonspecific but normally substantially thinner [14], thus having inherently smaller useful signals before reaching baseline values. Consideration may be given to use of a smaller data subset such as clock hour or macular regional data.

Statistical abnormality

The level of statistical abnormality denoted as definite in OCT and field parameters provide useful criteria, although it was recognized that the generalizability of normative databases is highly dependent on the size and composition of the 'normal' volunteers. However, raw values of thickness or threshold cannot be considered useful.

Levels of severity

It was realized that at least one lower level of certainty for the GON definition should be considered between definite GON and not GON. Examples are eyes with nonmatching positions of OCT and field defects, those with defects only on either OCT or field, eyes with unusual disc configurations.

Asymmetry

Regional asymmetric differences have not been properly accounted in normative OCT or field databases and whereas asymmetry is more the rule than the exception in OAG, the decision that the GON definition will consider each eye separately precludes its use.

Reliability or quality

Criteria for both field and OCT data in these measures needs to be included and will need to be set for each instrument individually. This is a feature that needs to be included in any validating dataset and the decision to use a 'cutoff' for quality needs justification in terms of predictive power and generalizability.

Confirmation

The issue of having a second confirming OCT and field test can best be solved by having a validating dataset in which the usefulness of the second test is measured.

Visual field criteria

The criteria for field loss in the Foster *et al.* classification were considered: 'The glaucoma hemifield test (GHT) graded 'outside normal limits' and a cluster of three contiguous points abnormal at the 5% level on the pattern deviation plot, using the threshold test strategy with the 24-2 test pattern of the Zeiss-Humphrey Field Analyser 2'. The Zeiss-Humphrey perimeter is not the only acceptable tool for field analysis, but represents a standard against which others should be validated. Haag Streit (Octopus) or other instruments that have equivalent features and have been shown to have a high level of correlation with the above criteria could be substituted. A major problem in data presentation for all automated field instruments is that the upper or lower position of the GHT abnormality must be manually determined.

Intraocular pressure

OAG occurs at all levels of untreated IOP; hence, there is no defining IOP for OAG.

VALIDATION OF A DEFINITION

A new GON definition requires validation by large datasets of representative eyes. Such validations could use one of several approaches. The approach most often suggested in the discussions was to compare objective criteria to the current gold standard–'clinician diagnosis'. Many experts feel that they know glaucoma when they see it, much as a famous U.S.A. Supreme Court justice said: 'I can't write a definition of obscenity, but I know it when I see it'.

Two approaches are ongoing as validation studies. In the first method, clinicians provide a clinical diagnosis on patients seen in their own clinic in whom the OCT and visual field are available. They categorize the eye as 'definite', 'probable', or 'not' GON. A second approach asks clinicians to scale the likelihood of GON from 0 to 100, based on clinical photographs, OCT, and visual field data that they view online. These approaches have the advantage that they provide objective data from OCT and automated perimetry to compare against the existing standard, but they assume that the existing standard is reasonable. Because there are many clinicians providing data, the consistency among them can also be measured.

A third analytic method would be to apply artificial intelligence (AI), with its various forms. If these methods are applied using a clinician definition, they are called supervised learning. AI can also be unsupervised, excluding clinician diagnosis and simply determining how the method categorizes patients. After the analysis, the result can be compared to what clinicians thought.

A final type of definitive test would be to acquire a dataset with longitudinal data on known progressive glaucoma eyes from various centers and test the definition on them. This should include many eyes with early degrees of injury.

THE FIRST VALIDATION APPROACH: CLINICIAN DIAGNOSIS AGAINST OBJECTIVE PARAMETERS

The first validation method has undergone initial analysis to compare objective features of structure and function to clinician diagnosis. The participants in the online discussion were solicited to submit data from eyes seen at their clinical practices. In each case, the clinician categorized the eye as definite GON, probable GON or not GON. Clinicians were aware of past history, the clinical exam, and the OCT and field data in making this determination. The issue was not whether the eye was likely to develop glaucoma, but rather whether it exhibited GON at that time, independent of the other risk factors present. Because these patients attended glaucoma clinics worldwide, this is not a population-based sample. However, the purpose of the definition is not to identify unselected persons for OAG prevalence, but to set a standard for what constitutes glaucoma as it is included in clinical glaucoma research. The very patients who would participate in those research projects are in this group.

Each eye entered into this database had substantial information entered from typical OCT and field outcomes. Both retinal NFL thickness and segmented macular inner ganglion cell/inner plexiform layer thicknesses were solicited. Both 24° and 10° Zeiss-Humphrey program data were included, but not Octopus data, as it was not possible to determine against an eccentricity-weighted normative database whether criterion field loss was localized in the upper or lower hemifield. The actual values of thickness and threshold were not entered, but rather the degree of abnormality against the instrument's normative database, both globally and by regions. For each eye, two tests of each type (OCT and field) were provided to test confirmation. Eyes were excluded if there were other potential disease-related causes for the defects in either modality.

The results of this validation investigation will be presented fully in another report. Eyes were entered from 16 centers: USA, Europe, South America, Korea, Japan, Hong Kong, and Australia/New Zealand. Of just under 2000 eyes entered from six of the centers thus far, half were clinically classified as definite GON, and ~25% each were probable or not GON. Among a large number of statistical parameters entered, the retinal NFL quadrant being statistically outside normal limits in either superior or inferior quadrant was 82% sensitive for definite GON, but insufficiently specific, with 11% of not GON eyes meeting the criterion. Likewise, the better of visual field criteria – abnormal GHT with a cluster of three nonedge points abnormal at $P < 5\%$ – was 83% sensitive for definite GON, but was likewise too nonspecific, identifying 15% of not GON eyes. In present evaluation, the best criteria were: a statistically abnormal OCT NFL superior or inferior quadrant with positionally matching abnormal field GHT, with a sensitivity of 72% for definite GON and specificity = 98% (i.e. 2% of not GON misclassified as definite). As expected, the eyes that met the objective criteria had somewhat greater mean deviation in field compared to those who were categorized as 'possible' and 'not' GON by the new definition (−7.5 dB vs. −2.2 dB vs. −1.2 dB). The requirement for a second test of each type that confirmed the categorization caused a lower sensitivity (66%), but did not improve on the already nearly perfect specificity. There were insufficient 10° fields or OCT macular thickness data submitted at this time to include them in the analysis.

SECOND VALIDATION APPROACH: SCALING THE CLINICIAN DECISION

Although the first approach used the clinical diagnosis of each eye from one (treating) expert, the second approach differs by having multiple experts grade each eye as GON on a 0–100 likelihood scale. Experts around the world will assess a clinical photograph of disc, OCT NFL thickness, optic nerve head neuroretinal rim measurements, macular ganglion cell thickness, and 24° visual field tests (Humphrey or Octopus), but will not have any other clinical history. The full range of damage will be represented. A web application will present eyes to the experts in a random order and ensure that every eye is evaluated by the same number of specialists. A pilot analysis of grading variability among experts using this methodology was presented at the ARVO 2019 Annual Meeting (Invest Ophthalmol Vis Sci. 2019;60:6152).

A series of possible glaucoma definitions will be created from objective field and OCT criteria. Field and OCT criteria will be tested in isolation, then combined with AND/OR logic, including consideration of anatomical correspondence. This approach will estimate the average and range of GON likelihood for any chosen OCT/field criteria. Unlike the first approach, the raw OCT and field data can be considered as well as their statistical likelihood. Both intra- and inter-expert reproducibility will be evaluated. A database of approximately 1000 eyes from 15 countries has been acquired and 500 glaucoma specialists will be the goal for participation.

These two approaches offer evidence-based validations of one or more GON definitions to be used in clinical research. Additionally, the data may provide insight into parameters that are more strongly valued by specialists when assessing the likelihood of GON. The data and web application also can be adapted into educational tools. The comparison across groups derived from different continents will assess whether some regionally resident persons differ in GON phenotype, or, whether experts evaluating GON in eyes from their own region differ from assessments by others less familiar with that group.

REMAINING ISSUES – ASSESSMENT OF PROGRESSION

A number of glaucoma expert groups have collected longitudinal data on carefully selected glaucoma patients, in some cases to study different racial groups. For datasets with consistent follow-up over several years of OCT and field data, it will be interesting to apply the point in time definitions from these initial approaches to eyes at the start and the end of follow-up. There are many ways to define progressive change and these may affect the view of the usefulness of each definition.

CAVEATS

Other than approaches with unsupervised AI, the approaches here may have reduced predictive power or greater variation in assessment by clinicians if there are phenotypic differences among eyes with GON, or, inherent baseline structural or functional categories that differentially affect the testing. An example often pointed out during the initial online discussion was significant myopia, more prevalent in OAG in general, and particularly so in Asian populations.

There are off-target effects of this effort that should be avoided. Those seeking to limit reimbursement for glaucoma care should not be allowed to take a method of research standardization and co-opt it as a restrictive process. Nor would it be appropriate for governmental regulatory agencies to base approval of treatments for glaucoma on such criteria. The process suggested and its outcome do not deal with whether an eye should be treated or how aggressively. As useful as the Foster *et al*. definitional structure has been in literature, there is no evidence that such concerns have materialized.

CONCLUSION

With support from many centers globally, it is possible to achieve an objective definition for GON among patients with OAG at good sensitivity and high specificity. In comparisons among clinical research investigations, it would be useful to assure that a standardized GON definition is available on those included. A set of criteria that have high specificity, but identify only three-fourth of those felt by clinicians to have definite GON is not so stringent as to restrict inclusion of data. A more complete analysis using the first approach is forthcoming that includes 10-2 fields and OCT macular thickness. The second approach, based on the clinician-scaling, will likely provide a more complete determination of features of structure and function that clinicians view as defining GON. Analysis using a standard definition would not preclude use of other definitions by individual investigators.

Acknowledgements

None.

Financial support and sponsorship

None.

Conflicts of interest

There are no conflicts of interest.

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

- ■ of special interest
- ■■ of outstanding interest

1. Hollows FC, Graham PA. Intra-ocular pressure, glaucoma, and glaucoma suspects in a defined population. Br J Ophthalmol 1966; 50:570–586.
2. Tielsch JM, Sommer A, Katz J, *et al.* Racial variations in the prevalence of primary open angle glaucoma: The Baltimore Eye Survey. JAMA 1991; 266:369–374.
3. Quigley HA. The number of people with glaucoma worldwide. Br J Ophthalmol 1996; 80:389–393.
4. Quigley HA, Vitale S. Models of open-angle glaucoma prevalence and incidence in the United States. Invest Ophthalmol Vis Sci 1997; 38:83–91.
5. ■■ Foster PJ, Buhrmann R, Quigley HA, Johnson GJ. The definition and classification of glaucoma in prevalence surveys. Br J Ophthalmol 2002; 86:238–242.

This paper represented the first definition system for glaucoma for use in clinical research in glaucoma and is the basis for the updated process of definition using more modern methods.

6. Tham YC, Li X, Wong TY, *et al.* Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology 2014; 121:2081–2090.
7. Google Scholar data, August 8, 2017.
8. Schuman JS, Hee MR, Arya AV, *et al.* Optical coherence tomography: a new tool for glaucoma diagnosis. Curr Opin Ophthalmol 1995; 6:89–95.
9. Coleman AL, Sommer A, Enger C, *et al.* Interobserver and intraobserver variability in the detection of glaucomatous progression of the optic disc. J Glaucoma 1996; 5:384–389.
10. Werner EB, Bishop KI, Koelle J, *et al.* A comparison of experienced clinical observers and statistical tests in detection of progressive visual field loss in glaucoma using automated perimetry. Arch Ophthalmol 1988; 106:619–623.
11. https://www.rand.org/topics/delphi-method.html [Accessed 1 July 2019]
12. Quigley HA, Hohman RM, Addicks EM, *et al.* Morphologic changes in the lamina cribrosa correlated with neural loss in open-angle glaucoma. Am J Ophthalmol 1983; 95:673–691.
13. Danesh-Meyer HV, Boland MV, Savino PJ, *et al.* Optic disc morphology in open angle glaucoma compared with anterior ischemic optic neuropathies. Invest Ophthalmol Vis Sci 2010; 51:2003–2010.
14. Quigley HA. Examination of the retinal nerve fiber layer in the recognition of early glaucoma damage. Trans Am Ophthalmol Soc 1986; 84:920–966.