Critical Literature Review

Alhadeff PA, De Moraes CG, Chen M, Raza AS, Ritch R, Hood DC (2017). The association between clinical features seen on fundus photographs and glaucomatous damage detected on visual fields and optical coherence tomography scans. *J. Glaucoma* 26 498–504. 10.1097/IJG.0000000000000640

Preface

This is a critical literature review submitted to meet a CORE requirement for the Glaucoma Diplomate program. Also attached is a completed checklist of the 2015 STARD Guidelines. This criteria was selected as this article meets the EQUATOR criteria for STARD analysis. STARD stands for "Standards for Reporting Diagnostic accuracy studies". The list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. It can be used by authors, editors and peer-reviewers to evaluate manuscripts for completeness. A diagnostic accuracy study evaluates the ability of one or more medical tests to correctly classify study participants as having a target condition. This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called index test.[1] A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the reference

standard. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.[2] Never having completed one of these in my professional career I have relied on several sources for guidance. First of these was the EQUATOR network. I have been in contact with several individuals who have been helpful to me in navigating the system. I have also benefitted from a pre-emptive web search for guidelines in critical review of the scientific literature and these are cited in the references.[3,4,5] Finally, I applied where possible the criteria used by the Americaan Academy of Optometry in refereeing articles for the OVS in which I have participated.

This article was chosen to review out of two potential candidates because of its clinical relevance and familiarity with the past and current work of one of the authors (Donald Hood, PhD) on this topic.

Introduction

The ability of ancillary testing to accurately identify patients with glaucomatous optic atrophy (GON) is typically tested against the gold standard of either optic nerve stereo photography and/or standard automated perimetry (SAP).[6,7,8] This has been reported widely in the literature, some of which have been cited in previous case reports.[9] A plethora of papers have attempted to determine which SD-OCT parameter has the best predictive value for Glaucoma.[10,11,12,13] The purpose of the study in this paper is to take the opposite approach. Specific clinical features of the optic nerve evaluated by stereo photography are compared with nerves previously determined

to be glaucomatous by OCT and SAP. This purports to be the first paper to investigate this relationship.[14]

Historically, optic nerve head stereo photographs and standard automated perimetry have been considered the gold standard for glaucoma diagnosis.[8,15,16] The advent of new imaging technologies such as Spectral Domain Ocular Coherence Tomography (SD-OCT) have had significant impact on the diagnostic paradigm. Many studies have been performed to validate the performance of these new technologies.[17,18,19,20] Their accuracy has been determined by comparison to the existing Gold standard of stereo photographs and visual fields.[21,22] In this study the opposite approach was taken. In a prospective sample of patients, stereo photographs of optic nerves were graded by two masked experts as Glaucoma, Glaucoma suspect or normal based on the presence or absence of specified features of GON. The results were then compared to those determined by OCT and SAP of these same patients to assess the frequency of which features of GON on stereo photos correlate with the ancillary tests. The study's findings revealed a significant correlation, especially for vertical cup to disc ratio (VCDR) greater than 0.6. A surprising finding was the poorer predictive power of localized thinning. It was hypothesized that localized thinning perhaps represents an earlier stage of the disease and they recommended longitudinal follow up to clarify this.

Methods

One eye of 100 glaucoma patients or suspects and 62 healthy subjects was prospectively selected. Although the study was approved by the review boards of Columbia University and the

New York Eye and Ear infirmary of Mount Sinai there is no indication of from what clinic the patients were selected or when. There is no explanation for the reason for the selection of one eye or the other or not both. There is also no indication as to whether the selection was random or not. This makes it difficult to evaluate the sample for bias.

The inclusion criteria stipulated that to be identified as Glaucoma or suspected glaucoma patients were required to have optic disc abnormalities based on clinical evaluation including intraocular pressure and family history, a mean deviation on Humphrey Field Analyzer (HVFA) 24-2 algorithm better than −6.00 dB, ≤75 years old, refractive error with spherical equivalent less −6.00 diopters or better, reliable performance on both visual fields, open angles and a best-corrected visual acuity equal or better than 20/40. Patients also had to meet criteria for an experienced field taker. Normal (Healthy) subjects were identified as having normal optic disc appearance, IOP < 22 mmHg, and normal 24-2 and 10-2 VF results. Patients with other ocular pathology were excluded.

Optic nerve stereo photographs were evaluated by 2 Glaucoma specialists simultaneously. Both graders were co-authors of the study. 12 features of GON were identified and evaluated: vertical cup-to-disc ratio (VCDR) of greater than 0.6; inter-eye VCDR asymmetry of greater than 0.2, small disc with significant cupping, optic disc pit, focal and diffuse neuroretinal rim thinning, disc hemorrhage, b-zone peripapillary atrophy (bPPA), nasal cupping, violation of the “ISNT” (inferior rim greater than superior greater than nasal greater than temporal)[23], and focal and diffuse retinal nerve fiber layer loss.

Spectral Domain OCT was performed using a Heidelberg Spectralis Tomographer (Heidelberg Engineering GmbH Dossenheim, Germany). Both the combined retinal ganglion cell and inner plexiform layers (RGC+) of the SD-OCT macular scans and the retinal nerve fiber layer (RNFL) of discs scans were segmented and evaluated. Each hemifield was classified separately and a significant defect on the probability cluster plot was defined as three abnormal points horizontally or vertically adjacent to each other in the same hemifield, representing a 2%, 2%, 1% or worse defect for macular cube or 2%, 2%, 5% for disc cube. These criteria were selected to produce an approximately 5% false positives rate in both cases. Scans determined to be of poor quality were rejected.

Threshold visual fields were performed using the Humphrey VF Analyzer (Carl Zeiss Meditec, Inc., Dublin, CA) Both 24-2 and 10-2 strategies were performed. The classification of 24-2 VF and 10-2 VF was also based upon cluster criteria. Each hemifield was classified separately and considered abnormal if at least 3 contiguous test points respecting the horizontal midline were abnormal (at 5%, 5%, 1% or 5%, 2%, 2%) on either total deviation (TD) or pattern deviation (PD) probability plots. This is consistent with the widely utilized Anderson-Patella's criteria.[24]

The eyes were classified into three groups: “suspects” (both OCT and VFs normal in both hemifields); “glaucoma” (both the 24-2 SAP and disc cube SD-OCT and/or both 10-2 SAP and macula cube SD-OCT were abnormal in the same matching hemifield); “controls” (healthy subjects with normal discs, both OCT and VFs normal in both hemifields); and other (the remaining eyes). This classification was done to “To minimize false-positives and false-negatives” although exactly how this is achieved by this choice is never clarified.

Two aspects of this classification scheme are problematic. First, “suspects” as noted above were defined as having a normal OCT and VFs normal in both hemifields. This may represent an overly conservative assessment and may be a contributing factor to the unexpected weakness in predictive value of GON features such as focal thinning. The authors make no further elaboration on the decision for this classification, other than “To minimize false-positives and false-negatives” as noted above. Other studies have used more expansive criteria.[25,26,27,28] This would clearly eliminate pre-perimetric disease, for example.

It appears that some of these eyes then wound up in the category of “other”. One problem with this is that this group is poorly defined. In addition, this presents a fair number of eyes (n=39) which were lost to analysis. It is stated in the paper that the results of these patients were subjected to analysis but not included in the results.

Results

The association between clinical signs of GON and glaucoma vs. suspect vs. control groups was tested with multivariate regression, which is a technique that estimates a single regression model with more than one outcome variable. It differs from multiple (or multivariable) regression in that several dependent variables are jointly regressed on the same independent variables.[29] The analysis showed a statistically significant difference between the clinical signs of GON found in glaucoma patients and glaucoma suspects as compared with normal eyes. VCDR of greater than 0.6 was the most common feature, present in 92% of glaucoma eyes, 70% of glaucoma suspects

and only in 3% of false positive healthy eyes. Violation of the ISNT rule was the second most significant sign, present in more than 80% of glaucoma or glaucoma suspect eyes and 5% of false positive controls. Close correspondence was found with the classification based upon visual field and OCT results.

The Breusch-Pagan Test[30] was employed to determine statistical significance of the variable independence. The Breusch-Pagan Test is a test for heteroscedasticity of errors in regression. Heteroscedasticity means "differently scattered". This is the opposite of homoscedastic which means "same scatter." Homoscedasticity in regression is an important assumption; if the assumption is violated, you won't be able to use regression analysis.[31] Using the Breusch-Pagan test most GON features were able to discriminate the glaucoma and suspect groups from controls. The pattern of clinical signs of GON seen on stereo photographs was statistically different between glaucoma ($P<0.001$) and suspects ($P<0.001$) vs. controls and explained up to 68% of the total variance of the diagnosis based upon SD-OCT and VFs. Vertical cup-to-disc (VCDR) >0.6, focal neuroretinal rim thinning, focal RNFL loss, and violation of the ISNT rule had the best performance to differentiate glaucoma and suspects from controls. Compared to the suspect group, glaucoma eyes (abnormal SD-OCT and VF tests) were more likely to have VCDR >0.6 (92 vs. 69%, $P=0.003$), diffuse rim (53 vs. 9%, $P<0.001$) and RNFL (61 vs. 26%, $P<0.001$) thinning, and beta-zone parapapillary atrophy (68 vs. 17%, $P<0.001$). The exception was the bPPA ($P=0.517$), which was statistically as frequent in suspects as in controls.

When comparing suspects and Glaucoma only, the eyes classified as glaucoma were more likely to have VCDR > 0.6, diffuse rim thinning, diffuse RNFL loss, bPPA, and nasal cupping. Three patients were identified with optic nerve hemorrhages. All three were classified as Glaucoma.

As noted previously, eyes in the suspect group were more likely to have focal rim thinning and focal RNFL loss than those classified as glaucoma. This was most likely to occur in the inferotemporal region, an area identified with early Glaucoma.[32,33,34]Along with eyes classified as "other" (6 of which were identified as "pre-perimetric Glaucoma") this imparts some degree of concern about selection bias. It could be argued that the results would have been different with more inclusive selection criteria. In any event it certainly merits a more careful interpretation of the results.

Discussion

The study's findings suggest classifying features of glaucomatous optic atrophy on optic nerve stereo photography correlate with the results of spectral domain optical coherence tomography and standard automated perimetry for detecting early signs of glaucomatous optic neuropathy and differentiate healthy eyes from those with glaucoma or suspected glaucoma. However, further follow-up might be necessary in some of the eyes to differentiate false positives from early signs of the disease. The findings and statistical analysis justify the authors' premise and conclusions.

Reviewing the paper for compliance with STARD guidelines proved to be an interesting exercise. The STARD criteria is part of the EQUATOR network collection of guidelines for evaluating the scientific literature. The EQUATOR network (https://www.equator-network.org/) provides toolkits with a flowchart to determine what guidelines are appropriate to evaluate a particular study. This study fell under the STARD guidelines as it met the criteria for diagnostic/prognostic studies.[35] There were 30 different items in the criteria, some with multiple sub-categories. A completed evaluation for this study is attached. For a paper accepted to a known, peer-reviewed journal it was surprising to find a number of missing features. As noted previously, there was a dearth of information presented to justify the selection criteria and, more important, the designation of the groups. This was particularly apparent in the distinction between Glaucoma and suspect groups. This could have had a significant impact on the results reported.

The discussion was remarkable for a lack of evaluation of the study's strengths and weaknesses. There was no real consideration of sources of error or bias, other than the recognition that poor prognostic power of focal rim thinning could represent an earlier stage of the disease. A simple assessment would be that the strengths of the study were the clinical relevance of the subject matter, but that were several weaknesses including sample selection, choice of definition of presence or absence of the condition and failure to meet a number of basic STARD criteria.

An explanation for choice of GON parameters - how they were arrived at and why would have been helpful. There was a comparison to Jost Jonas' seminal work on optic nerve head anatomy.[36] which was reviewed in a previous case report.[9] It could be inferred, although not

specifically stated, that the choice was related. The authors go on to state that their results were "in consonance" with Jonas et al. Finally, in addition to the suggestion that there be longitudinal follow up, it would be helpful to either re-evaluate the selection definitions or more thoroughly explain the reasons.

Comment

The authors present a clinically relevant study with respect to optic nerve head evaluation. Although there have been many papers to assess the accuracy of SD-OCT in identifying patients with for at risk for Glaucoma, this is purportedly the first paper to study the opposite scenario- how accurate are GON features identified on Stereo photographs in discriminating normals from those with Glaucoma or suspect by SD-OCT or SAP. Several features are demonstrated to be more clinically significant than others, but the general premise is statistically validated by the results which suggests an evaluation of a common disease process.

STARD CRITERIA
(http://www.equator-network.org/)

TITLE OR ABSTRACT

1 Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)

No. Not specifically stated. As the purpose is to "To facilitate retrieval of their article, authors can explicitly identify it as a report of a diagnostic accuracy study. This can be performed by using terms in the title and/or abstract that refer to measures of diagnostic accuracy, such as 'sensitivity', 'specificity', 'positive predictive value', 'negative predictive value', 'area under the ROC curve (AUC)' or 'likelihood ratio'." Does not meet this criteria.

ABSTRACT

2 Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)

There are multiple criteria for this item as defined by "STARD for Abstracts: essential items for reporting diagnostic accuracy studies in journal or conference abstracts". The paper meets some but not all of the criteria.

INTRODUCTION

3 Scientific and clinical background, including the intended use and clinical role of the index test

Yes.

4 Study objectives and hypotheses

Yes "In particular, assuming that the sdOCT and SAP results are abnormal and suggestive of glaucoma, what are the features of GON and their frequency when described by glaucoma-trained specialists using stereo-photography? We investigated how clinicians classified the optic discs of patients in which ancillary diagnostic tests were suggestive of glaucoma."

METHODS

Study design

5 Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)

Prospective study.

Participants

6 Eligibility criteria

Yes.

7 On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)

Yes.

8 Where and when potentially eligible participants were identified (setting, location and dates)

No.

9 Whether participants formed a consecutive, random or convenience series

Not specifically stated.

10a Index test, in sufficient detail to allow replication

Yes.

10b Reference standard, in sufficient detail to allow replication

Yes.

11 Rationale for choosing the reference standard (if alternatives exist)

Yes.

12a Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory

Yes.

12b Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory

Yes.

13a Whether clinical information and reference standard results were available to the performers/readers of the index test

No. They were masked to the reference test results.

13b Whether clinical information and index test results were available to the assessors of the reference standard

No.

Analysis

14 Methods for estimating or comparing measures of diagnostic accuracy

Yes.

15 How indeterminate index test or reference standard results were handled

No. They were identified- “Thirty-nine eyes (39%) had sdOCT and/or SAP abnormalities in nonmatching hemifields (other). Only 6 eyes (15.38%) showed preperimetric glaucoma (ie, sdOCT cluster criteria abnormal and SAP VFs within normal limits). Interestingly, 28 eyes (71.8%) of this group had at least 1 abnormal SAP hemifield, but the sdOCT was normal.” But no specifics provided of what was done with the data.

16 How missing data on the index test and reference standard were handled

Yes.

17 Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory

No.

18 Intended sample size and how it was determined

No.

RESULTS

Participants

19 Flow of participants, using a diagram

No.

20 Baseline demographic and clinical characteristics of participants

Yes.

21a Distribution of severity of disease in those with the target condition

No.

21b Distribution of alternative diagnoses in those without the target condition

No.

22 Time interval and any clinical interventions between index test and reference standard

No.

Test results

23 Cross tabulation of the index test results (or their distribution) by the results of the reference standard

No.

24 Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)

Yes.

25 Any adverse events from performing the index test or the reference standard

No.

DISCUSSION

26 Study limitations, including sources of potential bias, statistical uncertainty, and generalisability

No.

27 Implications for practice, including the intended use and clinical role of the index test

Yes.

OTHER INFORMATION

28 Registration number and name of registry

DOI Registered only.

29 Where the full study protocol can be accessed

No.

30 Sources of funding and other support; role of funders

Yes.

---

[1] *Šimundić AM. Measures of Diagnostic Accuracy: Basic Definitions. EJIFCC. 2009;19(4):203-11. Published 2009 Jan 20.*

[2] ***STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies*** *http://www.equator-network.org/reporting-guidelines/stard/*

[3] *Wood, J.M. (2003).Research Lab Guide. MICR*3260 Microbial Adaptation and Development Web Site. Retrieved July 31, 2006, from http://www.uoguelph.ca/mcb/teaching/micr3260/research_lab/guide.shtml*

[4] *Step by step guide to reviewing a manuscript. https://onlinelibrary.wiley.com*

[5] *https://twp.duke.edu/sites/twp.duke.edu/files/file-attachments/scientific-article-review.original.pdf*

[6] *Greenfield DS, Weinreb RN. Role of optic nerve imaging in glaucoma clinical practice and clinical trials. Am J Ophthalmol. 2008;145:598-603.*

[7] *Jampel HD, Friedman D, Quigley H, et al. Agreement among glau- coma specialists in assessing progressive disc changes from photographs in open-angle glaucoma patients. Am J Ophthalmol. 2009;147:39-44.*

[8] *Sharma P, Sample PA, Zangwill LM, Schuman JS. Diagnostic tools for glaucoma detection and management. Surv Ophthalmol. 2008;53 Suppl1(SUPPL1):S17-32.*

[9] *Case Report 4: Mild/Moderate Primary Open-angle Glaucoma*

[10] *Saikumar SJ, Jose S, Bhat S, Mahesh G, and Giridhar A. Diagnosis of Preperimetric Glaucoma using Optical Coherence Tomography. Kerala J Ophthamol 2008; XX: 39-42.*

[11] *Lisboa R, Paranhos A, Weinreb RN, Zangwill LM, Leite MT, Medeiros FA. Comparison of Different Spectral Domain OCT Scanning Protocols for Diagnosing Preperimetric Glaucoma. Investigative Ophthalmology & Visual Science. 2013;54(5):3417-3425.*

[12] *Rao HL, Zangwill LM, Weinreb RN, Sample PA, Alencar LM, Medeiros FA. Comparison of different spectral domain optical coherence tomography scanning areas for glaucoma diagnosis. Ophthalmology . 2010; 117: 1692–1699.*

[13] *Wang X Li S Fu J Comparative study of retinal nerve fibre layer measurement by RTVue OCT, GDx VCC. Br J Ophthalmol . 2011; 95: 509–513.*

[14] *Alhadeff P. A., De Moraes C. G., Chen M., Raza A. S., Ritch R., Hood D. C. (2017). The association between clinical features seen on fundus photographs and glaucomatous damage detected on visual fields and optical coherence tomography scans. J. Glaucoma 26 498–504.*

[15] *Banegas SA, Antón A, Morilla-Grasa A, et al. Agreement among spectral-domain optical coherence tomography, standard automated perimetry, and stereophotography in the detection of glaucoma progression. Invest Ophthalmol Vis Sci. 2015;56(2):1253-1260.*

[16] *Lemij, H. Glaucoma Diagnosis – The role of optic nerve examination. European Ophthalmic Review, 2007:15-7*

[17] *Wollstein G, Ishikawa H, Wang J, Beaton SA, Schuman JS. Comparison of three optical coherence tomography scanning areas for detection of glaucomatous damage. Am J Ophthalmol. 2005;139:39–43.*

[18] *Mwanza JC, Oakley JD, Budenz DL, Anderson DR, The Cirrus OCT Normative Database Study Group Ability of CirrusTM HD-OCT optic nerve head parameters to discriminate normal from glaucomatous eyes. Ophthalmology. 2011;118:241–248.e1*

[19] *Akashi A, Kanamori A, Nakamura M, Fujihara M, Yamada Y, Negi A. Comparative assessment for the ability of Cirrus, RTVue, and 3D-OCT to diagnose glaucoma. Invest Ophthalmol Vis Sci. 2013;54:4478–4484.*

[20] *Ojima T, Tanabe T, Hangai M, et al. Measurement of retinal nerve fiber layer thickness and macular volume for glaucoma detection using optical coherence tomography. Jpn J Ophthalmol. 2007;51:197–203.*

[21] *Bussel II, Wollstein G, Schuman JS. OCT for glaucoma diagnosis, screening and detection of glaucoma progression*
*British Journal of Ophthalmology 2014;**98:**ii15-ii19.*

[22] *Kotowski J, Folio LS, Wollstein G, et al. Glaucoma discrimination of segmented cirrus spectral domain optical coherence tomography (SD-OCT) macular scans.British Journal of Ophthalmology 2012;**96:**1420-1425.*

[23] *Harizman N, Oliveira C, Chiang A, et al. The ISNT Rule and Differentiation of Normal From Glaucomatous Eyes. Arch Ophthalmol. 2006;124(11):1579–1583. doi:10.1001/archopht.124.11.1579*

[24] *Anderson DR Patella VM. Automated Static Perimetry. 2nd ed. St. Louis, MO: Mosby; 1999.*

[25] *World Glaucoma Association Consensus series 8. Progression of Glaucoma. eds. R.N. Weinreb & D.F. Garway-Heath & C. Leung & J.G. Crowston & F.A. Medeiros.. Amsterdam: Kugler; 2011.*

[26] *Kim KE , Jeoung JW , Kim DM , Ahn SJ , Park KH , Kim SH. Long-Term Follow-up in Preperimetric Open-Angle Glaucoma: Progression Rates and Associated Factors. American Journal of Ophthalmology , Volume 159 , Issue 1 , 160 - 168.e2*

[27] *Ophir A. First-Visit Diagnosis of Preperimetric Glaucoma. The Open Ophthalmology Journal. 2010;4:22-27.*

[28] *Lu AT, Wang M, Varma R, Schuman JS, Greenfield DS, Smith SD. Combining nerve fiber layer parameters to optimize glaucoma diagnosis with optical coherence tomography. Ophthalmology. 2008;115:1352–7.*

[29] *Alexopoulos EC. Introduction to multivariate regression analysis. Hippokratia. 2010;14(Suppl 1):23-8.*

[30] *Breusch, T.S.; Pagan, A.R. (1979). "Simple test for heteroscedasticity and random coefficient variation". Econometrica (The Econometric Society) **47** (5): 1287–1294.*

[31] *Statistics Solutions. (2013). Homoscedasticity [WWW Document].*

[32] *Lloyd MJ, Mansberger SL, Fortune BA, et al. Features of optic disc progression in patients with ocular hypertension and early glaucoma. J Glaucoma. 2013;22(5):343-8.*

[33] *Chen JJ, Kardon RH. Avoiding Clinical Misinterpretation and Artifacts of Optical Coherence Tomography Analysis of the Optic Nerve, Retinal Nerve Fiber Layer, and Ganglion Cell Layer. J Neuroophthalmol. 2016;36(4):417-438.*

[34] *Choi JA, Shin H-Y, Park H-YL, and Park CK. "The Pattern of Retinal Nerve Fiber Layer and Macular Ganglion Cell-Inner Plexiform Layer Thickness Changes in Glaucoma," Journal of Ophthalmology, vol. 2017.*

[35] *Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open 2016;**6:**e012799. doi: 10.1136/bmjopen-2016-012799*

[36] *Jonas JB, Bergua A, Schmitz–Valckenberg P, Papastathopoulos KI, Budde WM. Ranking of Optic Disc Variables for Detection of Glaucomatous Optic Nerve Damage. Invest. Ophthalmol. Vis. Sci. 2000;41(7):1764-1773.*