Title

Normative Databases for Imaging Instrumentation

Journal

Journal of Glaucoma, 24(6)

ISSN

1057-0829

Authors

Realini, Tony
Zangwill, Linda M
Flanagan, John G
et al.

Publication Date

2015-08-01

DOI

10.1097/ijg.0000000000000152

Peer reviewed

HHS Public Access

Author manuscript

J Glaucoma. Author manuscript; available in PMC 2016 August 01.

Published in final edited form as:

J Glaucoma. 2015 August ; 24(6): 480–483. doi:10.1097/IJG.0000000000000152.

Normative Databases for Imaging Instrumentation

Tony Realini[1], **Linda Zangwill**[2], **John Flanagan**[3], **David Garway-Heath**[4], **Vincent Michael Patella**[5], **Chris Johnson**[6], **Paul Artes**[7], **I. Ben Gaddie**[8], and **Murray Fingeret**[9]

[1]Department of Ophthalmology, West Virginia University Eye Institute, Morgantown, West Virginia

[2]Hamilton Glaucoma Center, Department of Ophthalmology, University of California, San Diego, La Jolla, California

[3]Department of Ophthalmology and Vision Sciences, University of Toronto, Toronto Western Hospital, Toronto, Canada; School of Optometry, University of Waterloo, Waterloo, Ontario, Canada

[4]National Institute for Health Research Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, United Kingdom

[5]Carl Zeiss Meditec, Inc. Dublin, CA

[6]Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, Iowa

[7]Department of Ophthalmology and Vision Sciences, Dalhousie University, Halifax, Nova Scotia, Canada

[8]Louisville, KY

[9]VA New York Harbor Health Care System, Brooklyn, NY; SUNY College of Optometry, New York, NY

Abstract

Purpose—To describe the process by which imaging devices undergo reference database development and regulatory clearance. The limitations and potential improvements of reference (normative) data sets for ophthalmic imaging devices will be discussed.

Methods—A symposium was held in July 2013 in which a series of speakers discussed issues related to the development of reference databases for imaging devices.

Results—Automated imaging has become widely accepted and used in glaucoma management. The ability of such instruments to discriminate healthy from glaucomatous optic nerves, and to detect glaucomatous progression over time is limited by the quality of reference databases associated with the available commercial devices. In the absence of standardized rules governing the development of reference databases, each manufacturer's database differs in size, eligibility criteria, and ethnic make-up, among other key features.

Corresponding author: Murray Fingeret, OD, 183 Lakeview Drive, Hewlett, NY 11557, murrayf@optonline.net, Phone 516 532 3598, Fax 516 569 3566.

Conflict of Interest

No other conflicts exist among the authors.

Conclusions—The process for development of imaging reference databases may be improved by standardizing eligibility requirements and data collection protocols. Such standardization may also improve the degree to which results may be compared between commercial instruments.

Keywords

Glaucoma; imaging; perimetry; reproducibility; database

In 2012, a joint symposium organized by the American Glaucoma Society and the United States Food and Drug Administration (FDA) was devoted to explore the role of posterior segment imaging in clinical glaucoma management, as well as the design of studies evaluating glaucoma therapies. A webcast of that meeting is available.[1] A key finding from that event was that the development of reference databases (also called normative databases) for imaging platforms is limited by 1) a lack of standardization of methodologies for developing such databases and 2) the existence of minimal guidelines informing the design of such databases.

In response to these findings, a subsequent symposium was organized by the Optometric Glaucoma Society. This symposium, entitled “Normative Databases for Imaging Instrumentation,” was held at the World Glaucoma Congress in Vancouver, BC, on July 17, 2013. While this symposium emphasized development of normative databases for glaucoma management, the principles exposed may also be applicable to databases designed as aids in the management of other diseases. A webcast of this symposium is available online.[2] This paper represents the proceedings of that symposium.

Considerations in Glaucoma Imaging Device Normative Database Construction and Utilization

The past two decades have seen a proliferation of posterior segment imaging devices that augment stereoscopic optic disc and retinal nerve fiber layer (RNFL) photography. These new platforms—from scanning laser polarimetry to confocal scanning laser ophthalmoscopy to optical coherence tomography—offer objective and quantitative anatomical measurements not available with standard ophthalmic photography. There is ample data in the literature suggesting that these devices can enhance clinical decision-making in glaucoma management, providing diagnostic performance that compares favorably to that of glaucoma specialists.[3]

To optimize the utility of these instruments, their output is usually compared to normative ranges developed from databases of subjects who do not have disease. Such databases have been called normal databases but are probably better referred to as a reference databases, as the word, normal does not necessarily imply healthy, and the word, abnormal does not necessarily imply disease. Another use for these devices is to identify statistically significant change over time. In this case, changes seen in a subject’s data would be compared to the test-retest variability seen in reference subjects who have not had true structural change. Such information most often is developed by imaging reference subjects many times in a short time interval. The goal is to differentiate between changes that likely are attributable to

testing variability, versus changes that are larger than the limits of typical testing variability known to occur in patients – and therefore are more likely to have been due to some disease process. There can also be patient-specific reference databases, as in the Topographic Change Analysis utilized by the Heidelberg Retina Tomograph (HRT), in which the patient's own test-retest variability serves as a reference for identifying statistically significant change. In any case, the empirically determined reproducibility of instrument findings can be used to help ensure that any change observed from test to test is true structural change and not background test-retest variability.

The development of a reference database involves many complex considerations including eligibility criteria, sample size, data stratification needs, and methods of presenting the results and the number of parameters to be assessed, as well as measurement variability:

Eligibility criteria

Who should be included in a reference database and who should not be? The reference cohort should be representative of patients who will be tested using the instrument. Subjects should be drawn from the same general population as clinical patients. Exclusion criteria should be minimized to ensure that the database reflects profiles of co-morbidities that are similar to patients. For instance, if subjects with cataract are excluded from the reference database, then its use in patients with cataract may be invalid. However, the reference database should be composed of people who do not have the disease for which the test subjects will be evaluated.

For an instrument to assess the optic nerve of a glaucoma suspect, the reference database should be free of subjects with glaucoma. However, the definition of glaucoma used to exclude glaucoma patients from the reference database *cannot* be based on the anatomical features for which normative limits are sought. Excluding subjects from a reference database intended to evaluate the optic nerve cannot be based on the optic nerve appearance or else only subjects with very healthy optic nerves will be included. Therefore, an optic nerve reference database might exclude people with elevated intraocular pressure (IOP) and/or those with abnormal visual field tests. This approach won't exclude every subject with glaucoma—for instance, the normal-tension glaucoma patient with pre-perimetric glaucoma—but it will keep the proportion of these patients low in the reference database, so that their effect will be minimized.

Size

The number of subjects included should be large enough to sufficiently characterize the reference population, including important covariates, with the limitation that developing reference databases is costly. With reference databases, we are most interested in the tails of the distribution—the extreme values—and are asking whether a tested subject's value is sufficiently different from the average to be statistically unlikely (say, 5% [p=0.05] or 1% [p=0.01]) to be considered within normal limits.

Stratification

Some covariates, such as age, refractive error/axial length, race/ethnicity, and disc tilt, are known to affect optic disc imaging parameters. If the effect of such a covariate on the measurement is known and is large enough to make a clinically-relevant difference, then stratification may be justified. For instance, retinal ganglion cells and their axons are lost throughout life, which is why the statistical analysis of standard automated perimetry data is age-stratified.

Presentation of Results and Analysis

Reports generated by imaging devices should communicate the test subject's results as well as the comparison to the normative range. This should be in an easy to interpret format that identifies statistically significant deviations from the reference database. Presentation might also report the value of important covariates compared to the distribution in the reference database, so that clinicians can be alerted to any potential outliers. For instance, a 110-year-old patient might have a statistically significant finding, but there may be so few 110-year-old subjects in the reference database that any comparison might not be valid. The limitations of the analysis should be made clear. For instance, a result printed in red does not necessarily mean that the finding is abnormal, but only that it differs statistically from the reference values. The clinician remains obliged to interpret the test output in light of the entire clinical picture.

Measurement variability

Measurement variability refers to the similarity of values obtained from the same test subject in more than one test when there has been no real change in the test subject's status. It is in essence the reproducibility of the measurement. There are many sources of imprecision in clinical practice, including factors related to the device, the operator, the patient and the session. Device factors are largely controlled by the manufacturer but also require proper maintenance of the device. Other contributors to measurement variability can be minimized by adhering to a consistent imaging protocol. It is helpful to quantify typical measurement variability so that clinicians can understand the range of test results that may be found on re-imaging. For instance, a result that is only slightly outside normal limits may be just barely within the same limits upon repeat imaging.

Single versus multiple parameters

In glaucoma, we are often interested in the topography of the optic nerve head, including the rim and the cup, the peripapillary RNFL, and the retinal ganglion cell thickness and macular thickness—often in multiple sectors per eye (superior, inferior, nasal, temporal). Basing classification upon more than one parameter in a single examination inevitably results in increased numbers of false-positive classifications. For instance, the probability that either one of two parameters will fall below the 5% value in a disease-free subject is greater than 5%. Optimally, false positive rates associated with use of multiple parameters should be quantified and available to users.

Comparison between Perimetric and OCT normative database construction

Our approach to development of reference databases for optical coherence tomography has been informed by extensive experience producing similar databases for perimetry. However, perimetry and imaging differ in many ways, and these differences may require adjustments to that approach when establishing an OCT reference database.

Subject effects

Perimetry is a task, while OCT is an image. In many ways, this works in imaging's favor. Perimetry—and the reference databases used to analyze perimetry results—are subject to all the limitations related to human behavior. Learning effects can introduce change over time in visual field testing but are not a factor in imaging. Likewise, fatigue effects can limit both the quantity and quality of perimetric data, but generally are not an issue for imaging. Operator technique and skill level are in play for both devices, although imaging is less dependent upon patient instruction, coaching and supervision.

Sample and data size

Regulatory agencies typically require approximately about 300 subjects in a reference database. Data size is also important and differs between perimetry and imaging. The average file for a perimetry test is 50 to 100 bytes, while OCT imaging files often are more in the range of 50 megabytes, and with further improvements in resolution may soon approach one gigabyte. The mass of data associated with imaging poses several issues. Comparison of patient results to reference data cannot be practically undertaken at the level of raw data. Summary parameters have been derived (for example, mean RNFL thickness by sector and overall) but whether these are the optimal parameters for classification and progression detection remains to be seen.

Eligibility criteria

For both perimetry and imaging reference databases, a normal clinical examination is a necessary subject inclusion requirement. However, one cannot require a normal visual field for inclusion in a visual field reference database, nor a normal optic nerve for an optic nerve imaging reference database. Doing so would restrict variability of the reference values to be less than the real variability in the target population, which would adversely affect classification accuracy.

Covariates

Refractive error and axial length might be especially important covariates in imaging reference databases, given that myopic eyes have characteristic optic nerve appearances and are thought to have thinner retinas. This might be relevant when the target population has a high prevalence of myopia, such as in Asian countries. Visual field reference databases include racially diverse samples but the statistical analyses usually have not been stratified by race. Comparative studies show similar distributions of standard automated visual field parameters in people of European versus African descent.[4] While there may not be enough differences in Standard Automated Perimetry values among races to warrant racial stratification, in imaging, there is evidence of structural heterogeneity among races.[5]

Measurement variability

Perimetric measurement variability depends more upon retinal location and the extent of damage than in OCT. Likewise, cataract affects perimetric results more than OCT. Learning effects are exclusively associated with perimetry.

Inclusion criteria for an OCT normative database

A standardized methodology for developing reference databases is desirable. Still, no set of eligibility criteria can completely exclude non-normal subjects—which is one reason why the term “reference database” is preferable to “normative database.”

The Matrix frequency-doubling technology perimeter’s reference database was developed in the United States using a racially diverse reference sample. When the device was utilized in Japan, approximately 20–25% of normal subjects had “abnormal” tests based on comparison to the US-derived reference database. Similarly, while it has been shown that mean values of SAP and SWAP parameters are consistent between those of European versus African descent, FDT mean deviation scores have been found to be statistically different (worse) in the latter group.[6] This raises an important question: should reference databases be ethnic-specific?

One thought is to impose strict exclusion criteria so that reference databases have a tight homogenous group of healthy normal subjects. Others argue the opposite: that we should enroll everyone, regardless of comorbidities because this reflects the target population we will be testing. Proponents of strict eligibility criteria would argue that if too many people with abnormal fields were included in the reference database, the test would have minimal discriminative power to detect an abnormal visual field because the device will think that all field defects are “normal.”

What is the normal visual field for the elderly?

Almost all OCT devices use a “normal” visual field as a requirement for inclusion of subjects in their reference databases. Yet, it is unclear what criteria need to be met for a field to be called normal. Visual field tests can be difficult when performed for the first time, and a small fraction of test results from healthy people are, by definition outside of normative limits. The requirement for a normal visual field is the most common reason why healthy subjects are excluded from imaging reference databases.

The mean sensitivity of the visual field declines with age, and this decline accelerates later in life.[6] Despite this, most visual field analysis applications (e.g. Statpac for the Humphrey Field Analyzer) assume a linear rate change, perhaps because individuals age at different rates, making it is likely that the between-subject variability of sensitivity also increases with age. If true, such an increase would make the use of non-linear age corrections less rewarding. There are also locational aging differences in the visual field; for instance, the superior field declines more steeply with age compared to the inferior field.[6]

It makes sense not to insist on a “perfect” visual field as an inclusion criterion. Rather, a clearly pathological visual field should be used as an exclusion criterion. This would have

the effect of allowing normal subjects having less than perfect visual fields to be included, as long as their visual field changes are not found to be associated with disease. Overall, the inclusion of less-than-perfect visual fields into a reference database should not adversely affect its overall diagnostic performance. If anything, it will shift the balance between sensitivity and specificity towards a more conservative (more specific, less sensitive) classification. Since the false-positive rate with diagnostic devices has often been higher than expected, a shift towards more conservative classification might be a positive change.

Should we have imaging normative databases that are based upon ethnicity?

In spite of evidence supporting the existence of racial differences in retinal structure, no US FDA cleared reference databases for OCT devices are currently race-specific or make any adjustments for race. The issue of race is complex with most studies using self-reporting as the means to record an individual's ethnicity. The definition of what constitutes a specific ethnic group is not always clear and there may be overlap between racial groups.

Disc size—usually measured by spectral-domain OCT as disc area—has been found to be larger in people of African and Hispanic descent than in people of Caucasian descent, although there is overlap in these distributions.[7,8] Likewise, there is evidence that Hispanics have thicker RNFL and that individuals of African descent have thinner ganglion cell complexes than other ethnic groups; these differences are small but statistically significant.[8, 10]

It may be possible to account for racial variation in optic nerves by adjusting for disc size. Disc size has been shown to be an important consideration for identifying glaucomatous changes.[8] Optic nerve racial differences have been found to be well accounted for when disc size is incorporated into analysis models.[8,9] Further, disc size correlates with both disc and cup parameters on OCT but not with RNFL thickness parameters.[9] Thus, optic disc size explains a large proportion of observed racial variability in optic nerve head parameters and explains much of the variability by race without the need for separate ethnicity-specific reference databases. Further, it is a readily available parameter measured by current OCT instruments.

Evidence that ethnicity-specific reference databases have improved the diagnostic performance of imaging devices is weak. An Indian reference database did not improve the diagnostic performance of HRT for glaucoma detection.[10] Although race-specific reference databases for the HRT 3 did improve sensitivity while maintaining specificity for Caucasians, they increased sensitivity at the expense of specificity in people of African descent.[11] Also, how do we address observed anatomical variability within broad race categories? Should southern and northern Indians be considered separately? And, how do we define specific racial groups? For example, does African descent include Africans and Afro-Caribbeans and African Americans? Are North Africans included? Likewise, what does it mean to be Asian? Are ocular biometrics consistent between the Chinese, Japanese, Koreans, Vietnamese, and Indian populations? How would we handle multiracial patients who are increasing as a proportion of the population?

Thus, while there is clear evidence of statistically significant racial differences in optic nerve head structure, these differences are small, and overlap significantly, and optic disc area explains most of the differences. A reasonable compromise would be to develop reference databases that are ethnically diverse and then to adjust normative limits based upon optic disc size.

Consideration of ethnic origin in construction of RNFL normative limits may be useful. However, it is unclear how best to define ethnic groups in practical and useful ways that support everyday clinical practice.

The process for developing an imaging reference or normative database may be improved by standardizing eligibility requirements and data collection protocols. Such standardization may also improve the degree to which results may be compared between commercial instruments. Visual fields are the most common tool used for eligibility. An important question is what constitutes an acceptable visual field to allow inclusion into the database. Currently it is based upon the criteria of a glaucoma defect. The dataset may improve if the criteria were based upon a healthy field as seen in an elderly individual. Finally, it is important that the reference set be as similar to the population being examined as possible.

Acknowledgments

Sources of Funding

This symposium was arranged with funding from the Optometric Glaucoma Society. Vincent Michael Patella is an employee of Carl Zeiss Meditec, Inc.

References

1. FDA/American Glaucoma Society. [Accessed August 26, 2013] Workshop on the Validity, Reliability, and Usability of Glaucoma Imaging Devices. at http://www.fda.gov/MedicalDevices/NewsEvents/WorkshopsConferences/ucm318305.htm#date
2. Optometric Glaucoma Society. [Accessed August 26, 2013] Normaltive Databases for Imaging Instrumentation. at https://client.blueskybroadcast.com/WGA/2013/202-203/index.html
3. Reus NJ, de Graaf M, Lemij HG. Accuracy of GDx VCC, HRT I, and clinical assessment of stereoscopic optic nerve head photographs for diagnosing glaucoma. Br J Ophthalmol. 2007; 91:313–8. [PubMed: 17035283]
4. Racette L, Liebmann JM, Girkin CA, et al. African Descent and Glaucoma Evaluation Study (ADAGES): III. Ancestry differences in visual function in healthy eyes. Arch Ophthalmol. 2010; 128:551–9. [PubMed: 20457975]
5. Racette L, Boden C, Kleinhandler SL, et al. Differences in visual function and optic nerve structure between healthy eyes of blacks and whites. Arch Ophthalmol. 2005; 123:1547–53. [PubMed: 16286617]
6. Spry PG, Johnson CA. Senescent changes of the normal visual field: an age-old problem. Optom Vis Sci. 2001; 78:436–41. [PubMed: 11444634]
7. Girkin CA, McGwin G Jr, Sinai MJ, et al. Variation in optic nerve and macular structure with age and race with spectral-domain optical coherence tomography. Ophthalmology. 2011; 118:2403–8. [PubMed: 21907415]
8. Zangwill LM, Weinreb RN, Berry CC, et al. Racial differences in optic disc topography: baseline results from the confocal scanning laser ophthalmoscopy ancillary study to the ocular hypertension treatment study. Arch Ophthalmol. 2004; 122:22–8. [PubMed: 14718290]

9. Knight OJ, Girkin CA, Budenz DL, Durbin MK, Feuer WJ. Effect of race, age, and axial length on optic nerve head parameters and retinal nerve fiber layer thickness measured by Cirrus HD-OCT. Arch Ophthalmol. 2012; 130:312–8. [PubMed: 22411660]

10. Rao HL, Babu GJ, Sekhar GC. Comparison of the Diagnostic Capability of the Heidelberg Retina Tomographs 2 and 3 for Glaucoma in the Indian Population. Ophthalmology. 2009

11. Zelefsky JR, Harizman N, Mora R, et al. Assessment of a race-specific normative HRT-III database to differentiate glaucomatous from normal eyes. J Glaucoma. 2006; 15:548–51. [PubMed: 17106370]