Evaluation of Frequency-Doubling Technology Perimetry as a Means of Screening for Glaucoma and Other Eye Diseases Using the National Health and Nutrition Examination Survey

Michael V. Boland, MD, PhD; Priya Gupta, MD; Fang Ko, MD; Di Zhao, PhD; Eliseo Guallar, MD, DrPH; David S. Friedman, MD, PhD

IMPORTANCE Glaucoma is a significant cause of global blindness and there are, as yet, no effective means of screening.

OBJECTIVE To assess the potential role of frequency-doubling technology (FDT) perimetry in screening for glaucoma using data collected as part of the National Health and Nutrition Examination Survey (NHANES).

DESIGN, SETTING, AND PARTICIPANTS Reanalysis of cross-sectional data of 6797 participants in the 2005-2008 cycles of the NHANES, which evaluated a sample of the noninstitutionalized US population with at least light-perception vision. A subset of optic nerve photographs were regraded by 3 glaucoma specialists in December 2012. Each participant underwent visual field testing, including FDT perimetry screening, and had fundus photographs taken.

MAIN OUTCOMES AND MEASURES Sensitivity and specificity of FDT perimetry to detect glaucoma, macular disease, or decreased visual acuity.

RESULTS A total of 5746 NHANES participants had optic images originally graded. We regraded 1201 images of 1073 eyes of 548 participants with initial cup-disc ratio (CDR) of 0.6 or greater and 423 images of 360 eyes of 180 randomly selected participants with initial CDR less than 0.6. Diagnoses of glaucoma by disc photograph were 1.6% (3 of 180) in the CDR less than 0.6 group and 31.4% (172 of 548) in the CDR of 0.6 or greater group. The sensitivity of FDT was 33% (95% CI, 0%-87%) and specificity was 77% (95% CI, 71%-84%). For the group with at least 1 CDR of 0.6 or greater, sensitivity of FDT was 66% (95% CI, 59%-73%) and specificity was 70% (95% CI, 66%-75%). When analyzed to give FDT credit for identifying glaucoma, macular disease, or decreased visual acuity, the sensitivity of the test was 80% (95% CI, 77%-83%) and the specificity was 83% (95% CI, 82%-84%). Approximately 25% of the NHANES population was not able to successfully complete FDT testing, representing screening failures and decreasing specificity.

CONCLUSIONS AND RELEVANCE Using the 2005-2008 waves of the NHANES as a model of population-based screening for eye disease, FDT perimetry lacks both sensitivity and specificity as a means of screening for glaucoma, the presence of retinal disease, or decreased acuity in a population-based setting. Given that no single test of glaucoma has yet been shown to be appropriate in a screening setting, to our knowledge, investigators should consider novel methods of detecting glaucoma or combinations of tests that might work better in a screening setting.
Frequency-doubling technology (FDT) perimetry has been proposed to screen asymptomatic persons for glaucoma. This recommendation is based on the fact that, compared with standard automated perimetry, the FDT test takes less time and may detect glaucoma damage earlier, both of which make it a candidate as a screening tool.

In this study, we used the following definition of screening: “the systematic application of a test or inquiry, to identify individuals at sufficient risk of a specific disorder to benefit from further investigation or direct preventive action, among persons who have not sought medical attention on account of symptoms of that disorder.”

We were rigorous in our use of the term screening to distinguish it from the concept of case finding, which usually involves an evaluation that includes more than a single test and is intended to be definitive in the determination of disease. Thus far, screening for glaucoma has not been adopted because no system has demonstrated adequate sensitivity and specificity in this relatively low-prevalence disease.

Systematic reviews have not supported screening for glaucoma, and the US Preventive Services Task Force repeatedly found insufficient evidence to support glaucoma screening in the primary care setting. Screening for glaucoma is still an important goal given its impact on vision worldwide and because treatments are available.

Perhaps the most fundamental problems underlying glaucoma screening are the lack of a legitimate gold standard for determining the presence of disease and the lack of a test with appropriate sensitivity and specificity to justify applying it to large populations with low prevalence of disease. While FDT perimetry has shown reasonable performance in clinic-based evaluations, few community-based studies of FDT performance have been performed in the United States. However, it has been evaluated in a screening tool in Japan and China.

The National Health and Nutrition Examination Survey (NHANES) is a cross-sectional survey that uses a stratified design to obtain representative health data of the civilian, noninstitutionalized US population. The NHANES oversamples elderly participants and certain age and minority groups, making it well suited to estimate glaucoma prevalence in the United States. Frequency-doubling technology perimetry was performed as part of the 2005-2008 NHANES evaluations. Understanding the performance of the FDT device in this population will provide guidance for those hoping to use FDT to screen for glaucoma.

Methods

This study was reviewed and approved by the institutional review board of the Johns Hopkins University School of Medicine and adhered to the tenants of the Declaration of Helsinki. Patient consent was not obtained because data were deidentified. The method for performing FDT perimetry on NHANES participants has been described. Briefly, each NHANES participant with at least light-perception vision and without an eye infection underwent a 19-point suprathreshold old screening test in both eyes using the N-30-5 test on the Matrix FDT (Carl Zeiss Meditec). To reduce the false-positive rate, participants were required to complete 2 such tests with reliable results. The NHANES protocol defined a test as unreliable if the false-positive rate was greater than 33%, if there were more than 33% fixation losses by blind spot testing, or if the technician administering the test noted a significant error. The result for a given eye was deemed unreliable if either of the 2 tests was unreliable by these criteria. Visual field loss was defined if at least 2 subfields (test points) in the first and second tests were abnormal at the 1% level and at least 1 abnormal subfield was the same on both tests.

In addition to visual field testing, all participants in the 2005-2008 cycles of the NHANES had fundus photographs taken. These photographs included 45° nonstereo views of the optic nerve and macula and were graded by a reading center to determine the cup-disc ratio (CDR) of the optic nerve and the presence or absence of macular disease including macular edema, panretinal photocoagulation, focal photocoagulation, artery or vein occlusion, diabetic retinopathy, age-related macular degeneration, chorioretinal abnormalities, macular hole, or retinal detachment. The details of the ophthalmic testing performed during these phases of the study (visual acuity and objective refraction) have been described, as have the methods for the overall NHANES protocol.

Grading of Optic Disc Photographs

Because CDR alone is not definitive for diagnosing glaucoma, a subset of the optic disc images was regraded by 3 glaucoma specialists (M.V.B., P.G., and D.S.F.). In December 2012, we reassessed all optic disc photographs from participants who had been graded with CDRs of 0.6 or larger in either eye by the photographic reading center. We also randomly selected 180 participants for whom the reading center determined neither eye had a CDR of 0.6 or greater. Image files were deidentified and transferred into a tablet-based review system (TruthMarker, IDx LLC) where they were graded to determine quality (excellent, good, fair, poor, or ungradable), vertical CDR (0.0 to 1.0 in increments of 0.1), notching in the neuroretinal rim (none, inferior, superior, or both), excavation of the optic cup (no, maybe, yes, or unable), optic disc hemorrhage (no, maybe, yes, or unable), myopic cup, cupping, notching, and retinal detachment.

At a Glance

- The National Health and Nutrition Examination Survey (NHANES) provides an opportunity to evaluate the potential role of frequency-doubling technology (FDT) perimetry in screening for glaucoma.
- Even giving FDT perimetry the benefit of the doubt and allowing any retinal disease or decreased visual acuity to count as a true-positive, the sensitivity and specificity of the test were 80% and 83%, respectively.
- A critical issue with FDT perimetry is the fact that many NHANES participants were unable to complete the test, which would necessitate referral for definitive diagnosis.
- Based on the results of FDT applied to the participants in the NHANES, this test does not appear to be appropriate for glaucoma screening in the general population.
or unable), tilting of the disc (no or yes), disc size (small, average, or large), and the presence of glaucoma (no, possible, probable, definite, or unable).

Three glaucoma specialists graded each image and then the results were adjudicated where necessary using the following algorithms:

1. Image quality and likelihood of glaucoma: if at least 2 of 3 graders agreed and the third grader was within 1 level on the scale, the majority grade was assigned to the image. If there was not agreement between at least 2 graders or if the third grader was off by 2 or more levels, the image was re-reviewed in the presence of all graders.

2. Vertical CDR: if at least 2 of 3 graders agreed within 0.1 on CDR, the median value was assigned to the image. If there was disagreement of at least 0.2 between any 2 graders, the image was re-reviewed in the presence of all graders.

3. Disc hemorrhage and disc size: if at least 2 of 3 graders provided the same grade, that grade was assigned to the image. If there was not agreement between at least 2 graders, the image was reviewed in the presence of all graders.

4. Excavation, notch, or tilted: if at least 2 of 3 graders provided the same grade, that grade was assigned to the image. Images without this degree of consensus were labeled no or none.

Diagnostic Performance of FDT Perimetry

Glaucoma was determined present if the consensus assessment of their optic disc photographs was probable or definite. An abnormal FDT was defined as any result that would have resulted in the patient being referred for further evaluation. This included the test not being done, an abnormal result, insufficient data (only 1 test completed), or an unreliable test. These criteria were used because they represent how the test would function in a true screening mode, with the results of the test used to recommend further evaluation or to reassure participants that they do not have disease.

To account for the possibility of false-positive results caused by nonglaucoma vision loss, we also analyzed the diagnostic performance of FDT perimetry while including as true-positives any NHANES participants who had an abnormal FDT and any finding on their fundus photographs that might also produce vision loss including retinopathy, macular edema, photocoagulation, macular degeneration, vascular occlusion, choroidal scars, macular holes, and retinal detachment. We also counted anyone with visual acuity less than 20/40 and an abnormal or unreliable FDT result as a true-positive.

Values for sensitivity and specificity were calculated for the group with either CDR of 0.6 or greater and for the sample of 180 individuals with CDRs less than 0.6 who had their images regraded. The test performance from the latter group was then used to estimate the performance of FDT perimetry on the entire group with CDRs less than 0.6 (5566 individuals total, 180 of which were regraded).

Statistical Analysis

When comparing the demographics of subgroups of participants, differences in continuous variables were assessed using the Mann-Whitney U test to avoid assumptions regarding normal distribution of values, and differences between proportions were assessed using the Fisher exact test. The performance of FDT perimetry was assessed against the gold standard of optic disc grading by calculating sensitivity and specificity using a contingency table with true-positives, true-negatives, false-positives, and false-negatives.

Although the NHANES oversamples some demographic groups, we did not weight our results because the oversampled groups (Hispanic, African American, and elderly individuals) are ones who would also likely be targets of glaucoma screening.

Quantitative analyses were performed using R (version 3.1.1; The R Foundation; http://r-project.org).

Results

The 2005-2006 and 2007-2008 waves of the NHANES included 2934 and 3863 participants, respectively (6797 total). Of these, 5746 had optic disc images that were originally graded by the reading center. We regraded 1201 images of 1073 eyes (some eyes had multiple photographs taken) from 548 participants who had at least 1 CDR of 0.6 or greater on the grading by the reading center. We also graded 423 images of 360 eyes of 180 randomly selected participants for whom neither CDR was graded 0.6 or larger by the reading center. There was no statistical difference between the graded and ungraded groups for individuals with CDRs less than 0.6 (Table 1). Compared with the ungraded group with CDRs all less than 0.6, the group with at least 1 CDR of 0.6 or greater was statistically older, more male, more likely of black race, and more likely to self-report a diagnosis of glaucoma. The rate of glaucoma based on disc photographs was much higher in the group with at least 1 CDR of 0.6 or greater (31.4% vs 1.6%; Table 1).

To estimate the diagnostic performance of FDT in identifying glaucoma in the NHANES population, we first used the grading of the randomly selected subset of participants with no CDR greater than 0.6 to estimate the performance on that entire subgroup. As described here, an abnormal FDT result was considered to be any outcome that would require the participant to be referred for further evaluation, including test results that were abnormal, not done, unreliable, or insufficient (could not complete 2 tests). Using these criteria, the sensitivity of FDT was 33% (95% CI, 0%-87%) and the specificity was 77% (95% CI, 71%-84%) (Table 2). The wide confidence interval for sensitivity was owing to the small number of glaucoma cases in this group. Diagnostic performance was similarly calculated for the group with at least 1 CDR of 0.6 or greater and the sensitivity was 66% (95% CI, 59%-73%) and the specificity was 70% (95% CI, 66%-75%) (Table 2). These results were combined to estimate the performance for the entire NHANES cohort and the sensitivity was 54.5% (95% CI, 48%-61%) and the specificity was 76.8% (95% CI, 76%-78%) (Table 2).

The analysis here assumed that the only cause of visual field defects was glaucoma and so likely underestimated the
performance of FDT for finding people in need of eye care for other reasons. Therefore, we performed the analysis including presenting visual acuity and information about any macular disease identified on fundus photographs. We also biased the results in favor of FDT by only considering macular disease and visual acuity less than 20/40 as true-positives if the FDT screening test result was also abnormal. This approach maximizes the ratio of true-positives to false-positives. The estimated performance of FDT to identify persons with macular disease, decreased visual acuity, or glaucoma using this approach had sensitivity of 80% (95% CI, 77%-83%) and specificity of 83% (95% CI, 82%-84%) (Table 3).

If these results were extrapolated to a sample of 400 participants, 34 (9%) would be true-positives, 9 (2%) would be false-negatives (ie, those with glaucoma and normal FDT), and 61 (15%) would be false-positives (ie, those without disease and a positive FDT). Because of the relatively low prevalence of glaucoma, the number of false-positives was roughly twice the number of true-positives. A visualization of the limitations of this screening program is displayed in the eFigure in the Supplement.

A significant source of diagnostic error was owing to the proportion of participants who did not successfully complete testing (Table 4).21,22 This was despite systematic and de-

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Group, %</th>
<th>Graded</th>
<th>CDR &lt; 0.6</th>
<th>CDR ≥ 0.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, median (range), y</td>
<td>59 (40-85)</td>
<td>58 (40-85)</td>
<td>63 (40-85)*</td>
<td></td>
</tr>
<tr>
<td>Sex (% female)</td>
<td>50</td>
<td>53</td>
<td>44*</td>
<td></td>
</tr>
<tr>
<td>Race/ethnicity*</td>
<td>15.1</td>
<td>17.2</td>
<td>17.3</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>55.4</td>
<td>58.3</td>
<td>36.3</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>19.2</td>
<td>17.8</td>
<td>34.9</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>3.0</td>
<td>2.8</td>
<td>4.7</td>
<td></td>
</tr>
<tr>
<td>Self-reported glaucoma*</td>
<td>4.4</td>
<td>6.1</td>
<td>21.2</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>95.1</td>
<td>93.9</td>
<td>78.2</td>
<td></td>
</tr>
<tr>
<td>Unsure</td>
<td>0.4</td>
<td>0</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Glaucoma by disc photograph (probable or definite), No. (%)</td>
<td>NA</td>
<td>3 (1.6)</td>
<td>172 (31.4)</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: CDR, cup-disc ratio; NA, not applicable.
* For the CDR ≥ 0.6 group, P ≤ .05 compared with ungraded group.

<table>
<thead>
<tr>
<th>Group</th>
<th>Glaucoma, Macular Disease, or Decreased Visual Acuity</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDR &lt; 0.6*</td>
<td></td>
</tr>
<tr>
<td>FDT result</td>
<td>Abnormal</td>
</tr>
<tr>
<td>Normal</td>
<td>2</td>
</tr>
<tr>
<td>CDR ≥ 0.6*</td>
<td></td>
</tr>
<tr>
<td>FDT result</td>
<td>Abnormal</td>
</tr>
<tr>
<td>Normal</td>
<td>51</td>
</tr>
<tr>
<td>Overall Estimate*</td>
<td></td>
</tr>
<tr>
<td>FDT result</td>
<td>Abnormal</td>
</tr>
<tr>
<td>Normal</td>
<td>123</td>
</tr>
</tbody>
</table>

Abbreviations: CDR, cup-disc ratio; FDT, frequency-doubling technology.
* Macular disease and decreased visual acuity were only included if the FDT result was also abnormal.
* Results for the randomly selected participants with CDRs < 0.6.
* Results for the participants with CDRs ≥ 0.6.
* Estimated diagnostic performance for entire cohort based on the results from the CDR < 0.6 and CDR ≥ 0.6 groups. The sensitivity for the entire cohort was 80% (95% CI, 77%-83%) and the specificity was 83% (95% CI, 82%-84%).

<table>
<thead>
<tr>
<th>Year</th>
<th>Eye, No. (%)</th>
<th>Right</th>
<th>Left</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005-2006</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not done</td>
<td>422 (14.3)</td>
<td>420 (14.3)</td>
<td></td>
</tr>
<tr>
<td>Insufficient</td>
<td>30 (1)</td>
<td>41 (1.4)</td>
<td></td>
</tr>
<tr>
<td>Unreliable</td>
<td>241 (8.2)</td>
<td>324 (11)</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>2096 (71.4)</td>
<td>1971 (67.1)</td>
<td></td>
</tr>
<tr>
<td>Positive (abnormal)</td>
<td>145 (4.9)</td>
<td>178 (6)</td>
<td></td>
</tr>
<tr>
<td>2007-2008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not done</td>
<td>538 (13.9)</td>
<td>535 (13.8)</td>
<td></td>
</tr>
<tr>
<td>Insufficient</td>
<td>70 (1.8)</td>
<td>77 (2)</td>
<td></td>
</tr>
<tr>
<td>Unreliable</td>
<td>365 (9.4)</td>
<td>577 (14.9)</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>2703 (70)</td>
<td>2414 (62.4)</td>
<td></td>
</tr>
<tr>
<td>Positive (abnormal)</td>
<td>187 (4.8)</td>
<td>260 (6.7)</td>
<td></td>
</tr>
</tbody>
</table>
tailed instructions given to each participant. The proportions of incomplete, insufficient, and unreliable tests suggest that approximately 25% of participants undergoing FDT perimetry for glaucoma screening would be referred for further evaluation. This phenomenon accounts for a large percentage of the false-positive results, thereby decreasing the specificity of FDT for detecting eye disease.

**Discussion**

Using the 2005-2008 waves of the NHANES as a model of population-based screening for eye disease, FDT perimetry lacks both sensitivity and specificity as means of screening for glaucoma or the presence of retinal disease or decreased acuity in a population-based setting. Even with an analytic approach that gave FDT credit for finding glaucoma, macular disease, or decreased visual acuity, there were many false-positives and false-negatives, limiting its ability to screen general populations.

These findings are consistent with the results of FDT perimetry applied to populations in Japan and China. The Tajimi Study assessed all residents of Tajimi City, Japan, who were 40 years or older. The criteria for an abnormal FDT result were less rigorous than those applied to the NHANES population; the test was given once for each participant and was considered abnormal if 1 or more points were abnormal. The proportion of participants successfully completing the test was much higher in the Tajimi Study (98%) than in the NHANES (75%) (Table 4). Using a combined assessment of the comprehensive eye examination and automated perimetry as the standard for diagnosing glaucoma, the Tajimi Study found FDT perimetry had a sensitivity of 55.6% and a specificity of 92.7%. The sensitivity was worse than what we found in our best-case analysis (80.4%) while the specificity was better (83% in the NHANES), although their results were reported on a per-eye basis, not a per-participant basis. Higher testability in the Japanese population likely contributes to the better specificity.

The Beijing Eye Study similarly evaluated more than 4000 individuals in Northern China. The authors also relied on optic nerve photographs to diagnose glaucoma. Using the same abnormality criterion used in the Tajimi Study, the Beijing Eye Study reported a sensitivity of 72% and a specificity of 86.1%, similar to our findings. The sensitivity in the Beijing Eye Study decreased further (58.3%) when they required 2 points to be abnormal.

Taken together, our results for FDT performance as a screening test and those from the Tajimi and Beijing studies are generally worse than what has been reported for FDT in nonpopulation settings. This difference was expected given that individuals solicited to participate in a study are likely to be more motivated and potentially more experienced with testing than those selected in a population-based setting. One striking difference between the results of FDT testing in the NHANES compared with the Tajimi and Beijing studies is that the latter studies had a much higher proportion of their participants successfully complete the test. Because we counted the FDT failures as positives, thereby producing false-positives, our specificity values were lower than for the other 2 studies.

The primary limitation of this investigation was that it was not an evaluation of screening per se. On the other hand, it did have the advantage of allowing us to evaluate the performance of FDT perimetry in a population that represents what a screening program would encounter. This is particularly true in the NHANES because it was designed to oversample black and Hispanic populations, both of which might be targeted in screening programs. We were also limited by the fact that we had to rely only on evaluation of optic disc photographs to determine the presence of glaucoma, although it is standard practice to use a structural measure of glaucoma to assess a functional one.

The results of FDT perimetry in the NHANES population do not support its role as a screening test. Because of the relatively low prevalence of glaucoma in the adult population, a test with even a low rate of false-positives will generate many cases of persons misidentified as having disease. This is a problem because those false-positives will later consume resources for definitive glaucoma evaluation and because they have incorrectly been told they may have a blinding disease. Perhaps more concerning, both our results and the results of the other 2 cited population studies found high rates of false-negative results. This means a significant number of those screened who actually have glaucoma would be told they did not have disease and would therefore be unlikely to seek further care.

**Conclusions**

Given that multiple population-based studies have demonstrated that FDT perimetry has substantial limitations as a glaucoma screening test and that systematic reviews of glaucoma screening have shown no evidence to support any single screening test, further work is needed to determine how best to identify this disease that is typically asymptomatic until late in the course. Future approaches to glaucoma screening should consider novel methods of detecting glaucoma, combinations of tests, and looking for people likely to be affected by glaucomatous vision loss in their lifetime, perhaps with more severe disease.
for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Boland, Ko, Friedman. Acquisition, analysis, or interpretation of data: All authors. Drafting of the manuscript: Boland, Friedman. Critical revision of the manuscript for important intellectual content: Boland, Gupta, Ko, Zhao, Guallar. Statistical analysis: Boland, Ko, Zhao, Guallar, Friedman. Administrative, technical, or material support: Boland, Friedman. Study supervision: Boland, Friedman.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Boland reported receiving research support from Heidelberg Engineering and Glaukos Corp. Dr Friedman reported consulting for Alcon, Bausch and Lomb, Merck, Pfizer, Allergan, Nidek, and QLT. No other disclosures were reported.

Funding/Support: The Wilmer Eye Institute receives support from the National Eye Institute (grant EY01766) and Research to Prevent Blindness. This project also received support from the Centers for Disease Control and Prevention.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Disclaimer: Dr Boland is the Web Editor for JAMA Ophthalmology but was not involved in the editorial review or the decision to accept the manuscript for publication.

Additional Contributions: We thank Ana Terry, MC, RD, of the Centers for Disease Control and Prevention for help with the logistics of transferring the fundus images. She did not receive additional compensation for her contribution.

REFERENCES