

The Test–Retest Reliability of the Photopic Negative Response (PhNR)

Jessica Tang¹, Thomas Edwards², Jonathan G. Crowston^{1,2}, and Marc Sarossy^{1–3}

¹ Centre for Eye Research Australia, Melbourne, Australia

² The Royal Victorian Eye and Ear Hospital, Melbourne, Australia

³ RMIT University, Melbourne, Australia

Correspondence: Marc Sarossy, The Royal Victorian Eye and Ear Hospital, 32 Gisborne Street, East Melbourne, Victoria 3002, Australia; e-mail: marc.sarossy@eyeandear.org.au

Received: 3 June 2014

Accepted: 24 August 2014

Published: 3 November 2014

Keywords: electroretinogram; glaucoma; statistics; photopic negative response; reliability

Citation: Tang J, Edwards T, Crowston JG, Sarossy M. The test–retest reliability of the photopic negative response (PhNR). *Trans Vis Sci Tech.* 2014;3(6):1, <http://tvstjournal.org/doi/full/10.1167/tvst.3.6.1>, doi:10.1167/tvst.3.6.1

Purpose: The photopic negative response (PhNR) may be useful as a tool to monitor longitudinal change in retinal ganglion cell (RGC) function. The goal was to assess PhNR test–retest reliability, and to estimate the amount of change between tests that is likely to be statistically significant for an individual test subject.

Methods: Photopic electroretinograms (ERGs) were recorded from 49 visually normal subjects (mean age, 38.9 years; range, 21–72 years). Signals were acquired using Dawson-Trick-Litzkow (DTL) electrodes in response to red stimulus at four flash energies (0.5, 1, 2.25, 3 cd·s/m²) on a blue background (10 cd/m²). The PhNR amplitude was recorded from prestimulus baseline to trough (BT), prestimulus baseline to fixed time point (BF), and b-wave peak to trough (PT). The ratio of baseline PhNR to b-wave amplitude (BT/b-wave) was calculated. Reliability was assessed using the intraclass correlation coefficient (ICC_{2,1}) and coefficient of repeatability (CoR).

Results: Flash energy of 1.00 cd·s/m² produced reliable, well-defined traces. At this stimulus, the a- and b-wave amplitudes were reproduced with moderate reliability (ICC, 0.62; CoR%, 90.0%; and ICC, 0.74; CoR%, 54.3%; respectively). For PhNR, the order from most to least reliable measurement was: PT (ICC, 0.64; CoR%, 59.1%), BT (ICC, 0.40; CoR%, 148.3%), and BF (ICC, 0.22; CoR%, 166.1%). The BT/b-wave did not improve reliability (ICC, 0.37; CoR%, 181.5%).

Conclusion: The b-wave peak-to-PhNR trough amplitude produced the most reliable measurement.

Translational Relevance: A relatively large magnitude of change in PhNR amplitude is required to make clinical inferences about changes in RGC function. Refinement to the technique of acquisition and/or processing of the PhNR is recommended to improve reliability.

Introduction

The photopic negative response (PhNR) is a slow, negative wave that follows the b-wave of the full-field, photopic electroretinogram (ERG, Fig. 1). In animal models, the PhNR has been shown to originate from the spiking activity of retinal ganglion cells (RGCs) and their axons with additional contribution from amacrine cells and glia.^{1–5} Studies in humans have reported the PhNR amplitude to be significantly reduced in glaucoma,^{6–8} optic nerve atrophy,^{9,10} and ischemic retinal diseases.^{11–13} Further, the decrease in PhNR in these conditions may precede detectable changes in morphology and visual function, enabling

the possibility of early object assessment of retinal dysfunction.^{7,14}

The PhNR has advantages over the conventional pattern electroretinogram (PERG) to assess RGC function. Unlike the PERG, the PhNR does not require clear optics, adequate refractive correction, or exact foveal placement of the stimulus, and, therefore, may be easier to record.¹ A recent cross-sectional study comparing PhNR to PERG found both tests to have similar sensitivity and specificity in detecting early glaucoma.¹⁵

Given its clinical potential in the diagnosis and monitoring of diseases involving RGC injury, the motivation for this study was to evaluate further the test–retest reliability and determine the most reliable

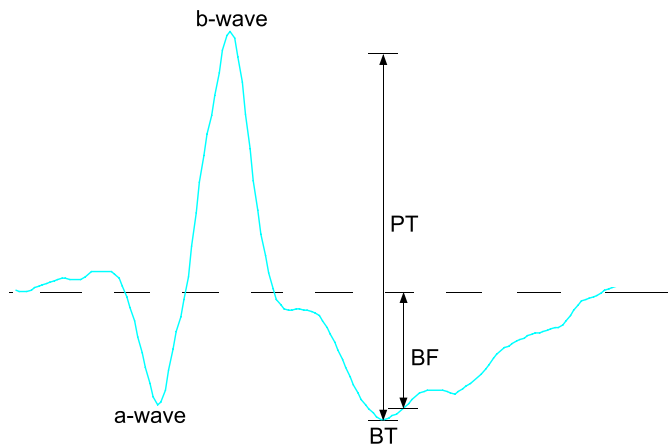


Figure 1. Illustration of measurement methods from the ERG waveform (2.25 cd.s/m²): a-wave measured from prestimulus baseline to first negative trough. The b-wave measured from a-wave to positive peak. The PhNR amplitude was measured from prestimulus baseline to negative trough following B-wave (BT), prestimulus baseline to fixed time point of 75 ms (BF), and B-wave to negative trough following B-wave (PT).

definition of the PhNR. This would enable the determination of the magnitude of change in an individual test subject to be considered significant.

In this study, we used a protocol that might be suitable in the setting of a glaucoma clinic. Therefore, a relatively brief adaptation period has been used as well as a relatively small number of sweeps per average.

Method

Subjects

A total of 49 visually-normal subjects participated in this study (mean age, 38.9 years; range, 21–72 years). Electrophysiological tests as detailed below were performed on two occasions by the same operator (mean days between test–retest, 7.9 days; range, 6–20 days).

The research was conducted after ethics approval by the Royal Victorian Eye and Ear Hospital Ethics Committee and under the tenets of the Declaration of Helsinki. Informed consent was obtained from all subjects after explanation of the nature and possible consequences of the study.

Electrophysiology

Pupils were dilated to at least 7 mm using 1% tropicamide (Mydriacyl). Eyes were preadapted for at least 1 minute with background room light (of 0.92

cd/m²). An Espion system (E2/ColorDome; Diagnosys LLC, Lowell, MA, USA) was used for stimulus generation and data acquisition. Brief preadaptation to the blue background of approximately 1 minute was performed before the first stimulus.

Brief, red (peak wavelength, 635 nm) stimulus at four flash energies (0.5, 1, 2.25, and 3 cd.s/m²) was delivered via a Ganzfeld sphere on a blue background (of 10 cd/m², peak wavelength 465 nm). Flashes were 4 ms in duration and presented at 1 Hz. The response was recorded using a Dawson-Trick-Litzkow (DTL) fiber electrode placed inside the lower lid conjunctival fornix of each eye.¹⁶ The ground electrode was attached to the forehead and the reference electrode was attached to the lateral canthus.

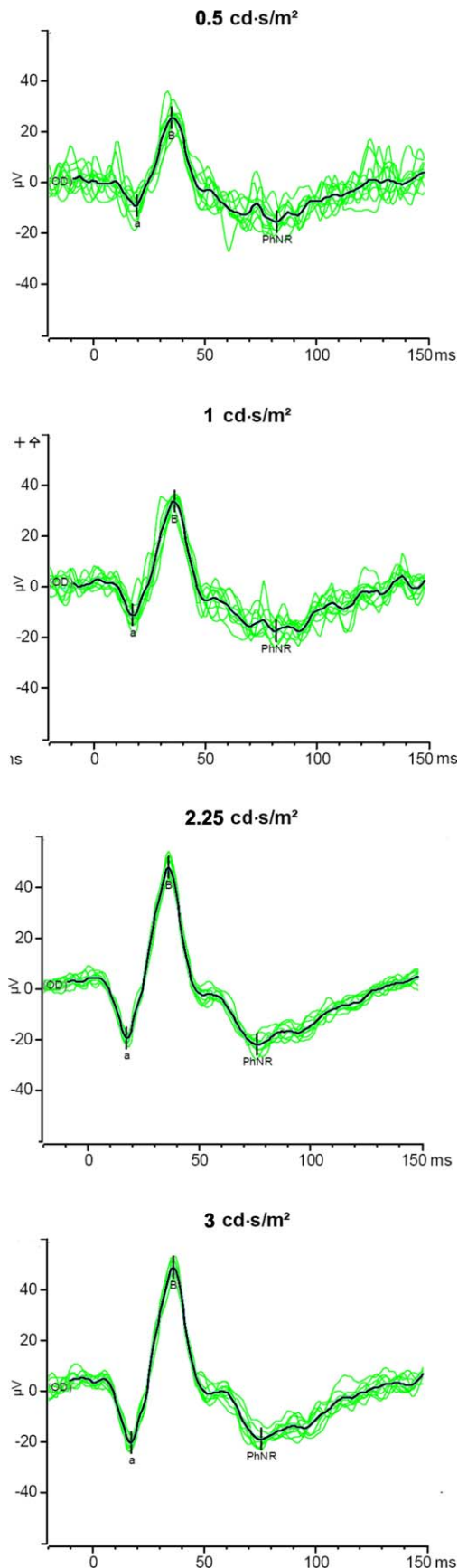
The waveforms were averaged over 10 sweeps at each stimulus level and signals were filtered from 0.15 to 100 Hz. An automatic rejection system removed large artefacts secondary to blink and eye movements.

Signal Analysis

The methods used to measure the a-wave, b-wave, and PhNR are shown in Figure 1. The a-wave amplitude was measured from the prestimulus baseline at time zero to the first negative trough. The b-wave amplitude was measured from the a-wave trough to the peak of the following positive wave. The PhNR amplitude was recorded from prestimulus baseline to trough following the b-wave (BT),^{1,3,6,9,11,12,17,18} the b-wave peak to PhNR trough (PT),^{18,19} and amplitude at a fixed time point with respect to prestimulus baseline (BF). The fixed time point was determined by averaging the BT implicit times of all right eye signals at the first visit. This was found to be 75 ms (95% confidence interval [CI], 73–76 ms). The PhNR was measured from b-wave peak to the negative trough at this fixed time point, even if this did not coincide with the trough on visual inspection.^{2,7,10,18} In addition, the ratio of baseline PhNR/b-wave (BT/b-wave) was calculated.^{14,15,19}

Statistical Analysis

Although both eyes were tested, only the results from the right eyes were included in the statistical analysis to exclude any effects of statistical dependency between the eyes. As amplitudes of ERG waveforms are distributed commonly with a positive skew, nonparametric Wilcoxon matched-pair signed-rank for related samples was used to evaluate intersession changes.



Relative reliability was analyzed using the intra-class correlation coefficient (ICC), which is a measure of the proportion of the total variance that is due to the variability between individuals. The 2-way, random-effect model (ICC_{2,1}) was chosen to account for systematic and random error, and to enable generalization of the reliability data beyond the confines of this study.²⁰ According to Fleiss,²¹ ICC values >0.75 represent “excellent reliability,” values between 0.4 and 0.75 represents “fair to good reliability,” and values <0.4 represents “poor reliability.”

Absolute reliability was assessed using coefficient of repeatability (CoR), which provides an interval within which 95% of test–retest measurement differences lie.^{22,23} This was calculated by ± 1.96 multiplied by the standard deviation of the mean difference, and the 95% CI was constructed as described previously.^{18,19} The CoR also was expressed as a percentage of the mean test–retest value (CoR%).

All statistical analyses were performed with SPSS (Released 2009. PASW Statistics for Windows, Version 18.0; SPSS, Inc., Chicago, IL, USA)

Results

Figure 2 show responses at the four flash energies from one typical subject. Overall, the amplitudes of all three waveforms grew with increasing flash energies. In the remaining Figures and analyses, we selected ERG responses to flash energy of 1.00 cd-s/m² (see also [Supplementary Table S1](#)). This stimulus was chosen because the PhNR responses were well-defined, with fewer artefacts caused by blink and eye movements compared to the higher flash energies, a finding consistent with other studies.^{15,24} The results for other flash energies are included in the supplementary material (see [Supplementary Tables S2, S3](#)).

The test–retest mean amplitudes and the mean of the differences in amplitudes are presented in [Table 1](#). There were no significant differences between test and retest measures (Wilcoxon matched-pair signed rank at $P = 0.05$). This is suggested further in the plots of test against retest amplitudes for each waveform ([Fig. 3](#)).

The PhNR amplitudes also are plotted as a

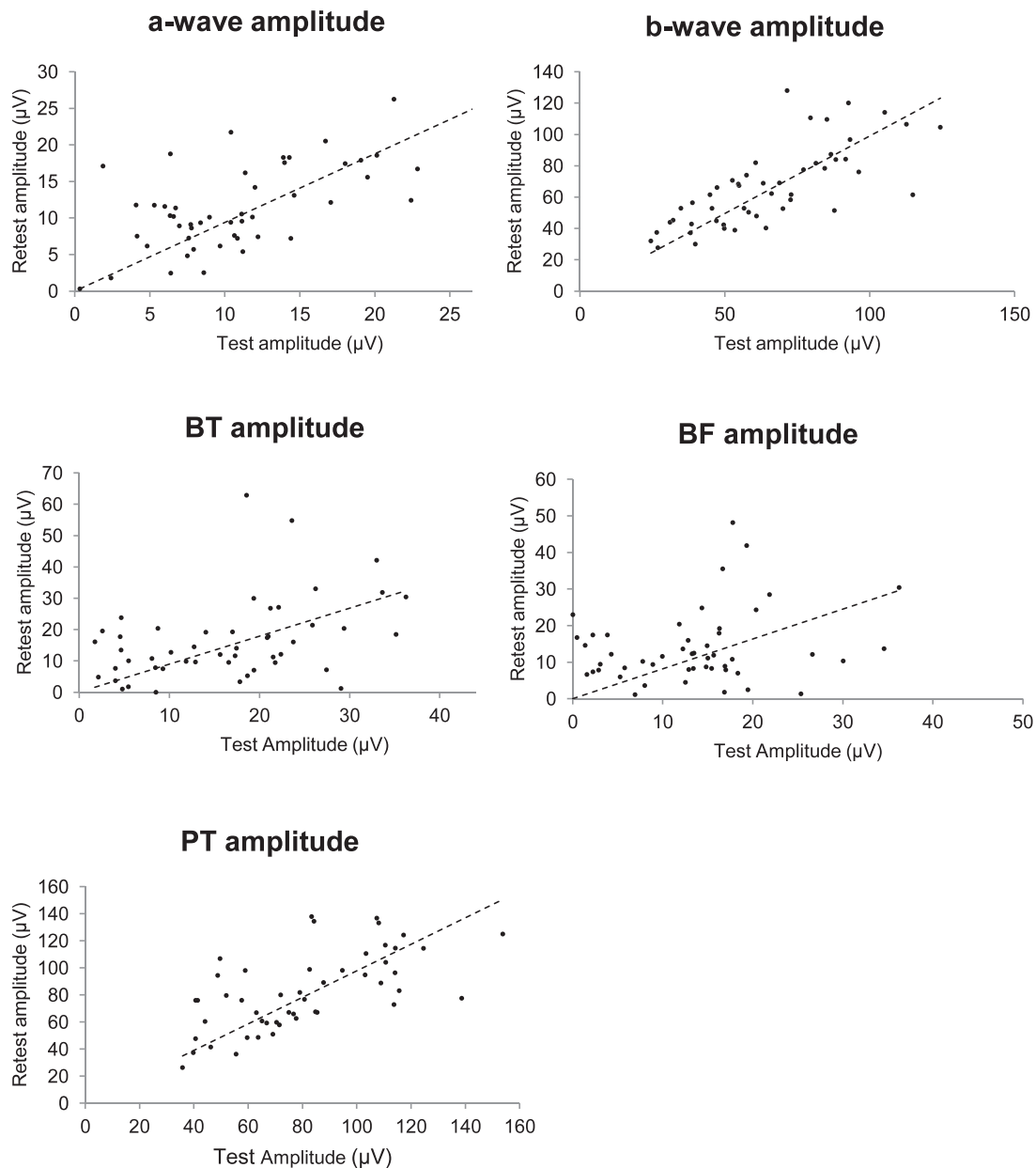
←

Figure 2. Representative responses to the four flash energies from one subject. *Thin green traces:* single responses. *Thick black trace:* average response.

Table 1. Mean Amplitude and Mean Test–Retest Difference in Amplitude for Flash Energy of 1.00 cd·s/m²

	Mean Amplitude (SD)		Mean Test–Retest Difference (% of Mean Amplitude)	95% CI of Test–Retest Difference	Sig
	Test	Retest			
a-wave	11.09 (6.07)	11.55 (5.79)	0.46 (4.1)	–1.02–1.95	0.75
b-wave	64.82 (24.74)	66.34 (25.40)	1.52 (2.3)	–3.69–6.74	0.49
BT	16.34 (9.56)	16.44 (12.81)	0.11 (0.7)	–3.45–3.67	0.63
BF	13.37 (8.46)	13.87 (9.90)	0.50 (3.6)	–2.81–3.82	0.90
PT	80.57 (28.59)	82.21 (28.59)	1.64 (2.0)	–5.41–8.69	0.87

Test–retest difference was calculated by subtracting test from retest amplitude. Significance was assessed using Wilcoxon match-pair signed rank at $P=0.05$. All mean amplitudes and mean difference between test–retest amplitudes are expressed in μV .

**Figure 3.** Test against retest amplitudes for each ERG waveform. *Dotted line:* represents perfect correlation.

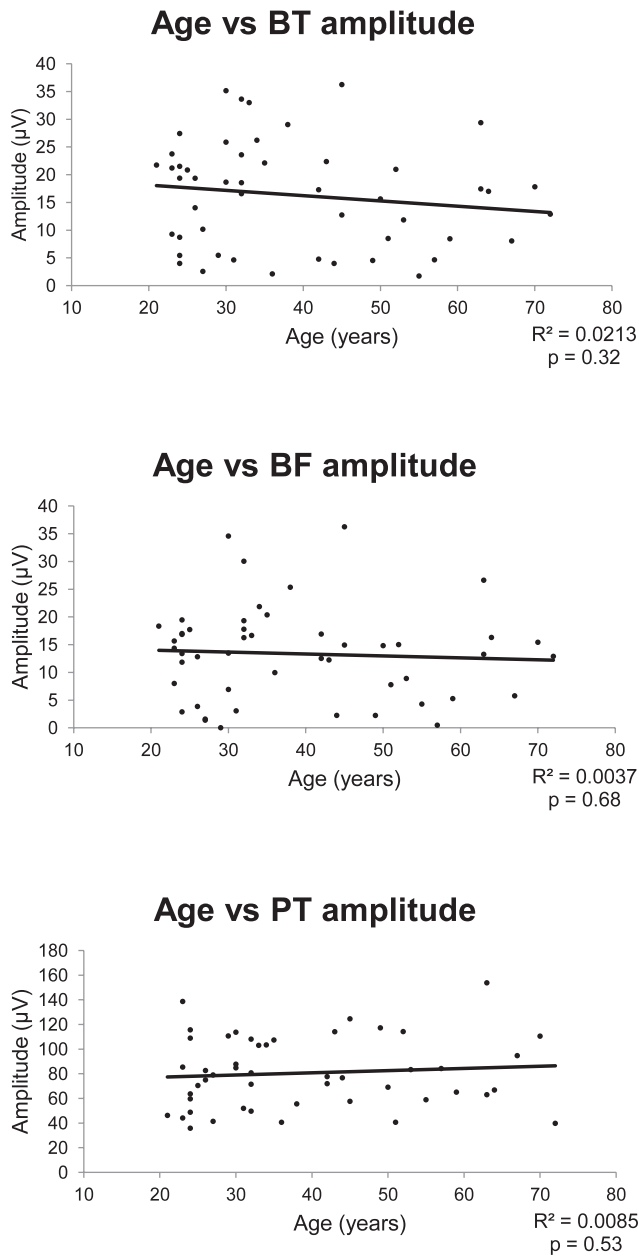


Figure 4. The PhNR as a function of age. *Straight line:* represents the best-fitting linear regression of these data.

Table 2. Reliability Indices for Flash Energy of 1.00 cd·s/m²

	ICC	ICC 95% CI	CoR, μV	CoR 95% CI	CoR%
a-wave	0.62	0.41–0.77	10.2	7.6–12.8	90.0
b-wave	0.74	0.58–0.85	24.3	18.1–30.4	54.3
BT	0.40	0.14–0.61	20.6	15.4–25.9	148.3
BF	0.22	0.00–0.47	22.6	16.9–28.4	166.1
PT	0.64	0.43–0.78	48.1	35.9–60.3	59.1
BT/b-wave	0.37	0.01–0.59	0.52	0.39–0.65	181.5

function of age in Figure 4. There was no statistically significant correlation between PhNR and age.

The ICCs for the a-wave, b-wave, and PhNR amplitudes, as well as the PhNR ratio calculations are presented in Table 2. According to categorization of ICC reliability reported by Fleiss,²¹ a- and b-waves showed “fair to good reliability.” Of the three PhNR amplitude measures, PT was the most reliable (ICC, 0.64), while BF was the least reliable (ICC, 0.22). The ratio of BT/b-wave was similarly poorly reliable (ICC, 0.37).

Table 2 also presents the CoR in absolute values and expressed relative to the mean value of test and retest combined (CoR%). For the a- and b-waves, 95% of test–retest difference is expected to fall within ± 10.2 (90.0%) and ± 24.3 (54.3%) μV , respectively. The PT had the lowest CoR (± 48.1 μV , 59.1%), while both baseline measurements had similarly high CoR (BT, ± 20.6 μV and 148.3%; BF, ± 22.6 μV and 166.1%). The BT/b-wave had the highest CoR% (± 0.52 μV , 181.5%).

Discussion

The present study demonstrated that the reliability of the PhNR varies depending on the method of amplitude measure. To date, although some studies have reported coefficient of variability as a measure of within-subject variation,^{6,14,25} few have reported PhNR test–retest reliability.

Viswanathan et al.⁷ found that on repeated recording in visually-normal subjects, baseline to PhNR trough measures was within $\pm 13\%$ of the mean amplitude; however, the study population was small ($n = 6$). Mortlock et al.¹⁸ reported much larger variation of $\pm 88.4\%$ of mean amplitude. We found the test–retest variation of baseline to PhNR trough amplitude to be even higher (within $\pm 148.3\%$). Measuring the trough at a fixed time point, which may be useful where the PhNR trough is not well-defined,^{2,7} had similarly high CoR% of $\pm 166.1\%$. The

relatively large CoR% can be accounted for by the smaller mean values of these measurements compared to their absolute CoR value, resulting in a larger percentage.¹⁸ However, taken together with the ICC values, amplitudes measured with reference to the baseline (BT and BF) are the least reliable (ICC, <0.4) and a relatively large magnitude of change may be required in repeated testing to be confident that the difference is significant.

We found the most reliable PhNR measure to be peak-to-trough (ICC, 0.64) where 95% of test–retest difference is expected to lie within $\pm 59.1\%$. This finding is slightly higher than that of Mortlock et al.¹⁸ ($\pm 42\%$, $n = 16$). The underlying process responsible for the PhNR response is of small amplitude and most likely commences before the b-wave is complete. Measurement of the PhNR in this way is analogous to the method of measuring the b-wave from the trough of the a-wave. It would be expected that the reliability of measuring the PhNR in this way would be affected by the reliability of the peak of the b-wave itself. There is, however, no simple way to evaluate the significance of this. It should be noted that this study addressed the reliability of measurement of the PhNR and did not attempt to investigate which method of measurement is most sensitive or specific for detection of longitudinal change in an individual test subject. Further studies are required to assess the sensitivity and specificity of PT compared to other methods in RGC dysfunction.

It has been suggested that the PhNR/b-wave amplitude ratio would show less variability and might prove to be a more useful measure than absolute PhNR amplitude.^{7,15,19} While Mortlock et al.¹⁸ reported the ratio to improve reliability, they calculated ratio of b-wave to peak-to-trough, which again comprises mainly of the b-wave amplitude. In our study, the ratio of baseline PhNR and b-wave amplitude was poorly reliable (ICC, 0.37; CoR%, 181.5) and likely reflects the variability of baseline PhNR measurements. This finding is consistent with an earlier study that found the reproducibility of the ratio was no better than absolute amplitude.¹⁴

A limitation of our study is that only 10 sweeps were averaged for each recording and we acknowledge that reliability may be improved by increasing the number of sweeps. Using more sweeps, however, would assume stationarity of the process and the absence, for example, of adaptation of the response. That question has not been investigated in this study. Our results do, however, highlight the importance of establishing laboratory norms of test–retest measures

in visually-normal subjects and those with RGC dysfunction before evaluating changes in repeated measures.

In summary, while the PhNR has clinical potential in the early detection and monitoring of RGC disease, refinements to the technique of acquisition and processing of the amplitude are required to improve test–retest reliability and increase the confidence in making inferences about changes in RGC function.

Acknowledgments

The authors thank Kristen Geddes for her guidance in conducting electroretinograms.

Disclosure: **J. Tang**, None; **T. Edwards**, None; **J.G. Crowston**, None; **M. Sarossy**, None

References

1. Viswanathan S, Frishman LJ, Robson JG, Harwerth RS, Smith EL III. The photopic negative response of the macaque electroretinogram: reduction by experimental glaucoma. *Invest Ophthalmol Vis Sci.* 1999;40:1124–1136.
2. Rangaswamy NV, Frishman LJ, Dorotheo EU, Schiffman JS, Bahrani HM, Tang RA. Photopic ERGs in patients with optic neuropathies: comparison with primate ERGs after pharmacologic blockade of inner retina. *Invest Ophthalmol Vis Sci.* 2004;45:3827–3837.
3. Li B, Barnes GE, Holt WF. The decline of the photopic negative response (PhNR) in the rat after optic nerve transection. *Doc Ophthalmol.* 2005;111:23–31.
4. Chrysostomou V, Trounce IA, Crowston JG. Mechanisms of retinal ganglion cell injury in aging and glaucoma. *Ophthalmic Res.* 2010;44:173–178.
5. Machida S, Raz-Prag D, Fariss RN, Sieving PA, Bush RA. Photopic ERG negative response from amacrine cell signaling in RCS rat retinal degeneration. *Invest Ophthalmol Vis Sci.* 2008;49:442–452.
6. Colotto A, Falsini B, Salgarello T, Iarossi G, Galan ME, Scullica L. Photopic negative response of the human ERG: losses associated with glaucomatous damage. *Invest Ophthalmol Vis Sci.* 2000;41:2205–2211.
7. Viswanathan S, Frishman LJ, Robson JG, Walters JW. The photopic negative response of the flash electroretinogram in primary open angle

- glaucoma. *Invest Ophthalmol Vis Sci.* 2001;42: 514–522.
8. North RV, Jones AL, Drasdo N, Wild JM, Morgan JE. Electrophysiological evidence of early functional damage in glaucoma and ocular hypertension. *Invest Ophthalmol Vis Sci.* 2010;51: 1216–1222.
 9. Gotoh Y, Machida S, Tazawa Y. Selective loss of the photopic negative response in patients with optic nerve atrophy. *Arch Ophthalmol.* 2004;122: 341–346.
 10. Wang J, Cheng H, Hu YS, Tang RA, Frishman LJ. The photopic negative response of the flash electroretinogram in multiple sclerosis. *Invest Ophthalmol Vis Sci.* 2012;53:1315–1323.
 11. Chen H, Wu D, Huang S, Yan H. The photopic negative response of the flash electroretinogram in retinal vein occlusion. *Doc Ophthalmol.* 2006; 113:53–59.
 12. Chen H, Zhang M, Huang S, Wu D. The photopic negative response of flash ERG in nonproliferative diabetic retinopathy. *Doc Ophthalmol.* 2008;117:129–135.
 13. Machida S, Gotoh Y, Tanaka M, Tazawa Y. Predominant loss of the photopic negative response in central retinal artery occlusion. *Am J Ophthalmol.* 2004;137:938–940.
 14. Machida S, Gotoh Y, Toba Y, Ohtaki A, Kaneko M, Kurosaka D. Correlation between photopic negative response and retinal nerve fiber layer thickness and optic disc topography in glaucomatous eyes. *Invest Ophthalmol Vis Sci.* 2008;49: 2201–2207.
 15. Preiser D, Lagréze WA, Bach M, Poloschek CM. Photopic negative response versus pattern electroretinogram in early glaucoma. *Invest Ophthalmol Vis Sci.* 2013;54:1182–1191.
 16. Hébert M, Vaegan, Lachapelle P. Reproducibility of ERG responses obtained with the DTL electrode. *Vision Res.* 1999;39:1069–1070.
 17. Chrysostomou V, Crowston JG. The photopic negative response of the mouse electroretinogram: reduction by acute elevation of intraocular pressure. *Invest Ophthalmol Vis Sci.* 2013;54: 4691–4697.
 18. Mortlock KE, Binns AM, Aldebasi YH, North RV. Inter-subject, inter-ocular and inter-session repeatability of the photopic negative response of the electroretinogram recorded using DTL and skin electrodes. *Doc Ophthalmol.* 2010;121:123–134.
 19. Fortune B, Bui BV, Cull G, Wang L, Cioffi GA. Inter-ocular and inter-session reliability of the electroretinogram photopic negative response (PhNR) in non-human primates. *Exp Eye Res.* 2004;78:83–93.
 20. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979; 86:420–428.
 21. Fleiss JL. *The Design and Analysis of Clinical Experiments. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics.* New York, NY: Wiley; 1986:432.
 22. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1:307–310.
 23. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999;8:135–160.
 24. Kremers J, Jertilla M, Link B, Pangeni G, Horn FK. Spectral characteristics of the PhNR in the full-field flash electroretinogram of normals and glaucoma patients. *Doc Ophthalmol.* 2012;124:79–90.
 25. Machida S, Toba Y, Ohtaki A, Gotoh Y, Kaneko M, Kuosaka D. Photopic negative response of focal electroretinograms in glaucomatous eyes. *Invest Ophthalmol Vis Sci.* 2008;49:5636–5644.