

Author Response: Test–Retest Reliability of Hemifield, Central-Field, and Full-Field Chromatic Pupillometry for Assessing the Function of Melanopsin-Containing Retinal Ganglion Cells

We thank Zhou, Lou, Pan, and Huang¹ for their interest in our paper.² We appreciate their assertion that the postillumination pupil response (PIPR) method we described has clinical utility. However, we would like to assert from the outset that our findings in the present paper by no means imply that this approach is ready for immediate clinical implementation without modification. In fact, we actively encourage tuning of the chromatic pupillometry methodology to target specific disorders before conducting larger-scale clinical trials, where the test–retest repeatability (TRT) and precision calculations the Zhou group suggest would become relevant.

The Zhou group has argued that the variation we report in our intraclass correlation coefficient (ICC) assessment of the hemifield PIPR measurements may undermine its clinical utility. The introduction of our paper clearly states that this study was designed as a proof of concept to “develop a PIPR testing method in which stimulation area/location can be adjusted to present hemifield, central-field, and full-field stimulation” (the second paragraph of the Introduction).² This goal was novel in the chromatic pupillometry literature. Despite being a promising new tool, the clinical utility of hemifield PIPR testing remains an open question. Accordingly, we feel it is premature to judge our methods using the International Organization for Standardization (ISO) standards typically applied to a well-established clinical diagnostic modality (TRT and precision studies).³ Our work to date represents the first steps toward a potential standardized and validated clinical test, which can be a lengthy process.

We interpret the relatively weaker hemifield PIPR ICC values as the likely result of reduced signal-to-noise ratios, which is supported by the higher median coefficient of variation (CV) values we observed under hemifield stimulation. We do note that defining the PIPR as the mean pupil size during a 20-second interval from 10 to 30 seconds post illumination is a very conservative index. Others have employed less stringent methods, sampling earlier postillumination and over shorter intervals.⁴ This conservative index is likely to be more susceptible to natural fluctuations in pupil contractility when the melanopsin-driven photo activity is weaker. There is additional evidence that the early rapid pupil constriction to a blue light stimulus also receives significant influence from the melanopsin-driven activity of intrinsically photosensitive retinal ganglion cells (ipRGC).⁵ Considering that our hemifield protocol was able to induce highly repeatable maximum pupil constrictions (MPC), we remain positive that our protocol has induced melanopsin-driven ipRGC activity in a repeatable manner. This does not preclude the evolution of a better index of melanopsin activity in the future as investigators develop a

better understanding of the influence of melanopsin on the pupil response.

We thank Zhou and coauthors¹ for their suggestions regarding statistics to assess whether a particular method is precise enough for clinical diagnostics; but as we have noted, we feel this is premature for an empirical pilot study based on 10 visually normal participants. We do agree that statistical power is an important consideration, and it was computed and utilized in our experimental design.

We calculated an a priori sample size estimation for our current study from the results of our previous study.⁶ The primary objective of our current study was to “develop a PIPR testing method in which stimulation area/location can be adjusted to present hemifield, central-field, and full-field stimulation.” Accordingly, the chief statistical test of interest to us was a one-way repeated measures ANOVA model on the mean PIPR measures with Testing Condition (five levels: lower hemifield, upper hemifield, central field, full-field blue, and full-field red) as the within-subjects factor. The effect size estimates (required for the a priori sample size calculations) were computed from the partial eta-squared (η^2) values we obtained for the 400 cd/m², 400 ms central-field blue, full-field blue, and full-field red conditions in our previous study (see Fig. 4 in Ref. 6). Our current empirical investigation was the first to examine the modulation of chromatic pupillometry PIPR when light stimulation was restricted to one hemifield at a time. Since there were no available data in the literature on chromatic pupillometry-derived hemifield PIPR measures, we used the η^2 from the 200 cd/m² full-field blue condition from our previous study (see Fig. 2 in Ref. 6) to estimate the expected effect size for the two hemifield conditions. Our reasoning was that the overall photon exposure on the retina would be similar between the 200 cd/m² full-field and the 400 cd/m² hemifield conditions we planned to use. This a priori analysis indicated that we needed a sample size of at least seven participants to attain a power value of 0.80. Indeed, when we calculated a post hoc power analysis on the primary statistical test used in our current study, we had achieved an omnibus power of 0.99 with the current sample size of 10 people.

The ICCs for the individual testing conditions were calculated subsequent to fitting a separate one-way random effects ANOVA model⁷ to the PIPR data. The requested post hoc power by testing condition for the PIPR ICC data are given in the Table. Note that all the statistical tests expected to elicit melanopsin-driven responses (blue light conditions) exceeded the desired power of 0.80. The statistical test on the full-field red condition exhibited low power of 0.40, possibly because this melanopsin-silent control condition had a low signal-to-noise ratio (CV = 0.87). The Table clearly shows that all the statistical tests we used to calculate the ICCs were more than adequately powered with our current sample size of 10 individuals. Care should be taken while interpreting post hoc power values, as it is generally not advisable to calculate the power of a study retrospectively.⁸ This is because a statistically significant result will always yield high power, as the effect size

TABLE. Post Hoc Power Values for the PIPR Intraclass Correlation Coefficient (ICC) Measures by Stimulation Condition

	Lower Hemifield	Upper Hemifield	Central-Field	Full-Field Blue	Full-Field Red
3 repetition model	0.99	0.98	0.99	0.99	0.40
6 repetition model	0.99	0.99	N/A	N/A	N/A

N/A, not applicable.

from which that power value was calculated is substantially high (since it is significant).

Zhou and coauthors¹ have suggested that the physiological diurnal variation in corneal thickness and shape should be taken into consideration in our study. It should be noted that the mean amplitude of diurnal changes of refractive power that was reported by Read and Collins⁹ was small (0.36 ± 0.11 D), with the largest change observed immediately after subjects woke from sleep or upon walking.⁹ There is no evidence that such a small variation in refractive error would have a discernible impact on chromatic pupillometry measurements of PIPR, which was induced by diffuse, nonpatterned illumination. The PIPR response reflects the activity of an irradiance detection pathway responsible for subconscious, nonvisual photoperception. We agree that ipRGCs also exhibit circadian variation¹⁰; however, it appears that Zhou and coauthors¹ have cited the wrong paper to support their argument. The 2002 paper by Hattar and coworkers (Ref. 7 in the Zhou group's letter) does not address this issue. In any case, the circadian variation of ipRGC responses was controlled in our study, as all our pupillometry tests were conducted within a narrow time window during the day (see the last sentence of the Experimental Conditions and Procedure: "All experiments were conducted during the day between 8 AM and 2 PM."). Our recordings were made by the same examiner using the same protocol in all instances.

In closing, our effect size estimates for the main statistical test and hence sample size calculations were derived empirically using the best information available at the time we designed our study. The TRT reliability values of our PIPR measures based on our current sample size of 10 visually normal participants are not meant to be applied directly to clinical practice before further development. That said, we fully support the notion that the chromatic pupillometry stimulation and response quantification methods may well need to be adjusted according to the specific disease process under scrutiny. In fact, this work is currently underway in our laboratory for a number of disorders.

Shaobo Lei¹
Herbert C. Goltz^{1,2}
Manokaraanantban Chandrakumar¹
Agnes M. F. Wong¹⁻³

¹Program in Neurosciences and Mental Health, The Hospital for Sick Children Research Institute, Toronto, Canada; ²Department of Ophthalmology and Vision Sciences, University of Toronto, Toronto, Canada; and ³Department of Ophthalmology

and Vision Sciences, The Hospital for Sick Children, Toronto, Canada.

E-mail: agnes.wong@sickkids.ca

References

1. Zhou W, Lou Y, Pan B, Huang J. Reliability of field chromatic pupillometry for assessing the function of melanopsin-containing retinal ganglion cells. *Invest Ophthalmol Vis Sci.* 2015;56:2519.
2. Lei S, Goltz HC, Chandrakumar M, Wong AMF. Test-retest reliability of hemifield, central-field, and full-field chromatic pupillometry for assessing the function of melanopsin-containing retinal ganglion cells. *Invest Ophthalmol Vis Sci.* 2015;56:1267-1273.
3. ISO 5725-2. *Accuracy (Trueness and Precision) of Measurement Methods and Results - Part 2: Basic Method for the Determination of Repeatability and Reproducibility of a Standard Measurement Method.* 1994. Available at: http://www.iso.org/iso/catalogue_detail.htm?csnumber=11834.
4. Park JC, Moura AL, Raza AS, Rhee DW, Kardouk RH, Hood DC. Toward a clinical protocol for assessing rod, cone, and melanopsin contributions to the human pupil response. *Invest Ophthalmol Vis Sci.* 2011;52:6624-6635.
5. McDougal DH, Gamlin PD. The influence of intrinsically-photosensitive retinal ganglion cells on the spectral sensitivity and response dynamics of the human pupillary light reflex. *Vision Res.* 2010;50:72-87.
6. Lei S, Goltz HC, Chandrakumar M, Wong AM. Full-field chromatic pupillometry for the assessment of the postillumination pupil response driven by melanopsin-containing retinal ganglion cells. *Invest Ophthalmol Vis Sci.* 2014;55:4496-4503.
7. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86:420-428.
8. Hoenig JM, Heisey DM. The abuse of power: the pervasive fallacy of power calculations for data analysis. *Am Stat.* 2001;55:19-24.
9. Read SA, Collins MJ. Diurnal variation of corneal shape and thickness. *Optom Vis Sci.* 2009;86:170-180.
10. Zele AJ, Feigl B, Smith SS, Markwell EL. The circadian response of intrinsically photosensitive retinal ganglion cells. *PLoS One.* 2011;6:e17860.

Citation: *Invest Ophthalmol Vis Sci.* 2015;56:2520-2521.
doi:10.1167/iovs.15-16847