Suitability of open-field autorefractors as pupillometers and instrument design effects

Carles Otero, Mikel Aldaba, Oriol Ferrer, Andrea Gascón, Juan C Ondategui-Parra, Jaume Pujol

Davalor Research Center, Technical University of Catalonia, Terrassa 08222, Spain

Correspondence to: Carles Otero. Davalor Research Center, Technical University of Catalonia, Terrassa 08222, Spain. carles.otero.molins@upc.edu

Received: 2016-06-20        Accepted: 2017-02-14

Abstract

AIM: To determine the agreement and repeatability of the pupil measurement obtained with VIP-200 (Neuroptics), PowerRef II (Plusoptix), WAM-5500 (Grand Seiko) and study the effects of instrument design on pupillometry.

METHODS: Forty patients were measured twice in low, mid and high mesopic. Repeatability was analyzed with the within-subject standard deviation ($S_w$) and paired $t$-tests. Agreement was studied with Bland-Altman plots and repeated measures ANOVA. Instrument design analysis consisted on measuring pupil size with PowerRef II simulating monocular and binocular conditions as well as with proximity cues and without proximity cues.

RESULTS: The mean difference ($±$standard deviation) between test-retest for low, mid and high mesopic conditions were, respectively: -0.09 ($±$0.16), -0.05 ($±$0.18) and -0.08 ($±$0.23) mm for Neuroptics, -0.05 ($±$0.17), -0.12 ($±$0.23) and -0.17 ($±$0.34) mm for WAM-5500, -0.04 ($±$0.27), -0.13 ($±$0.37) and -0.11 ($±$0.28) mm for PowerRef II. Regarding agreement with Neuroptics, the mean difference for low, mid and high mesopic conditions were, respectively: -0.48 ($±$0.35), -0.83 ($±$0.52) and -0.38 ($±$0.56) mm for WAM-5500, -0.28 ($±$0.56), -0.70 ($±$0.55) and -0.61 ($±$0.54) mm for PowerRef II. The mean difference of binocular minus monocular pupil measurements was: -0.83 ($±$0.87) mm; and with proximity cues minus without proximity cues was: -0.30 ($±$0.77) mm.

CONCLUSION: All the instruments show similar repeatability. In all illumination conditions, agreement of Neuroptics with WAM-5500 and PowerRef II is not good enough, which can be partially induced due to their open field design.

KEYWORDS: pupil; repeatability; agreement; open-field

DOI:10.18240/ijo.2017.04.11

INTRODUCTION

Nowadays, there is a vast variety of pupillometers specifically designed to determine pupil diameter\(^1\)\(^-\)\(^9\). However, there are several ophthalmic devices that can be found in a clinical setting and report also pupil size but as a secondary function. Among them, the autorefractometers\(^10\) (e.g. WAM-5500 and PowerRef II) and wavefront sensors\(^11\) that allow accommodation measurement are especially interesting due to the linkage between both systems (i.e. accommodative and pupillary system). While the refraction or aberration measurement function of these instruments has been widely assessed there is little information about their performance in terms of precision and accuracy when measuring pupil size. On the one hand, comprehensive studies, involving human pupil measurements, under different light levels is still not found for these ophthalmic devices that report pupil data. Thus, it is still unknown whether their pupillometric function can also be useful not only in the study of accommodation synchronically with pupil responses but also in important applications of pupillometry such as refractive surgery\(^1\) or the study of neurological disorders\(^12\)\(^-\)\(^13\).

On the other hand, there exists a wide variety of designs ranging from binocular\(^14\) to monocular\(^6\)\(^,\)\(^9\) pupillometers and including those that allow open field of view\(^15\), or contrary, those that display a virtual stimuli in a closed field environment\(^14\). Related to this, some evidence exists in the literature regarding accommodation in which concepts as proximal accommodation\(^16\) and instrument accommodation\(^17\) arise and can affect significantly the accommodative response, which in turn can affect the pupil measurement\(^18\). Following the same line of thought, little is known regarding how the instrument design contributes to the variability of the pupil measurement, or in other words, up to what extent the differences in design between pupillometers is a source of disagreement between pupil measurements in a clinical setting. The purpose of this study is dual. Firstly, to compare the pupil measurement at different light levels of a reference pupillometer (VIP-200, Neuroptics) with the two autorefractometers WAM-
Open-field autorefractors as pupillometers

5500 (Grand Seiko) and PowerRef II (Plusoptix). Secondly, to study the potential effects of instrument design on pupil measurements.

SUBJECTS AND METHODS

This cross-sectional study was conducted on healthy subjects recruited from the staff and students of the Technical University of Catalonia (Terrassa, Spain). Only participants that had pupils equal, round, and reactive to light and accommodation, and had no history of ocular disease, surgery, pharmacological treatment or any other condition that could affect the pupil shape or diameter were invited to participate. The study followed the tenets of the Declaration of Helsinki and was approved by the Ethical Committee Board of Mutua de Terrassa Hospital (Terrassa, Spain). All subjects were asked to give their written informed consent after receiving a written and verbal explanation of the nature of the study.

For the repeatability and agreement analysis 40 eyes from 40 subjects were finally included in the study, with a mean±standard deviation (SD) in age of 27.0±7.9y (20 to 59y). A cohort of 18 subjects was used for the second experiment related to the instrument design.

Measurement Devices

Three pupillometers were used. The reference standard considered in this study was the VIP-200 pupillometer (Neuroptics)[2,6,9]. It is a monocular infrared unit that has a gasket that goes against the patient’s face. It autocalibrates and autofocuses although the user is responsible for proper placement and alignment of the device. It takes multiple measurements, discards outliers, and reports the mean pupil diameter and SD values. The WAM-5500 is an openfield autorefractor and pupilometer widely used in accommodation studies that has been described in multiple occasions elsewhere[21-23]. It is worth to recall that it can work in static and dynamic (at 5 Hz) modes. The minimum and maximum pupil diameter is 2.00 and 8.00 mm (in steps of 0.10 mm), respectively. The WAM-5500 allows binocular viewing, but measurements are monocular. The PowerRef II is also an openfield autorefractor and pupilometer widely used in accommodation studies and deeper described elsewhere[21-23]. Its pupil measurable range goes from 3.00 to 8.00 mm in steps of 0.10 mm[20]. It measures in dynamic mode at a frequency of 25 Hz. Unlike the WAM-5500, it performs binocular measurements.

Examination Protocol

Subjects who accepted to participate underwent a standardized examination without cycloplegia to determine whether both pupils from the patient were equal, regular, and reactive to light and accommodation stimuli. For the repeatability and agreement analysis, each subject was measured twice in each device and in the following three lighting conditions (±SD): 0.05 lx (±0.01 lx) (low mesopic), 0.80 lx (±0.30 lx) (mid mesopic) and 20.00 lx (±0.30 lx) (high mesopic). The different illumination levels were achieved in a single ordinary eye examination room in a similar way as in Bradley et al.’s[21] study.

Before testing, subjects were told to wear wraparound completely-dark glasses for at least 5min, then, subjects were left at least two minutes at each illumination condition before start measuring. Pupil measurements were randomized among devices but not among lighting conditions, i.e. subjects were always measured first at the lowest light level (low mesopic) and then atmid and high mesopic, respectively. Both measurements (i.e. test-retest) taken with each device and at each light level were performed in an inter-session mode (centering and focusing of each instrument was performed twice). Notice that all measurements were monocular (randomization of the chosen eye was conducted) with the non-tested eye occluded with an eye patch except when measuring with Neuroptics. Patients were told to fixate on a red dot at 6 m distance and pupil size was averaged in a period of 7s when using the WAM-5500 and PowerRef II. Blinks were removed from data sets and no smoothing filters were applied to raw data.

For the instrument design analysis, on the one hand, it was studied the pupil measurement differences between open field of view with proximity cues versus without proximity cues, on the other hand, it was also studied monocular versus binocular measurements. For that purpose, it was simulated three different configurations at mid mesopic light level (Figure 1) using only PowerRef II, thus minimizing the impact of instrument variability.

The first configuration consisted on a monocular openfield measurement in which the tested eye fixated on the PowerRef II placed at 1 m distance. There were no objects interfering within the field of view of the patient except the pupillometer device. The second configuration consisted on a monocular measurement simulating the handheld Neuroptics measurement, that is, the clinician was placed in front of the patient and occluded one patient eye just like he would do when using the Neuroptics pupillometer. The non-occluded eye...
was the measured eye by the PowerRef II. This configuration is also referred to as a monocular measurement with proximity cues. Notice that proximity cues can be defined as non-optical visual information related to perceptual distance of objects that guide the accommodative system with respect to direction of the blur\[25\]. The last configuration consisted on a binocular openfield measurement without any object interfering within the field of view. All patients were uncorrected in terms of refraction during pupil measurements and all measurements were taken in a single session of 40min long.

### Statistical Analysis

The significance was set at 0.05 and the statistical analysis was performed using SPSS v22 (IBM Corp., USA). Normality of each variable was checked by applying the Shapiro-Wilk test and comparing the skewness and kurtosis statistics to the standard error. Repeatability of each device was analyzed mainly with the within-subject standard deviation (Sw). Further analysis were the mean and SD of the differences between both measurements, paired sample t-tests with the corresponding 95% CI, and also, the number of cases in which the absolute difference was above 0.50 mm\[15-16\]. Agreement of WAM and PowerRef II with respect the reference pupillometer (VIP-200, Neuroptics) was analyzed at each illumination condition by means of Bland and Altman plots. Additionally, the one-way repeated measures ANOVA was used to test the within-participant effects (i.e. the overall significant difference between each device). Where the assumption of sphericity was violated, the Greenhouse-Geisser correction was used. When significance was obtained, pairwise comparisons were examined by t-tests with Bonferroni correction. Analogously to repeatability analysis, the number of cases in which the absolute difference was above 0.50 mm was also considered. Regarding the instrument design analysis, the comparisons among the three configurations were studied in the same way as the agreement between devices.

### RESULTS

#### Repeatability

Test-retest results (repeatability) for each device and at each illumination condition are shown in Table 1. Sw, the mean±SD between two measurements of each device, the t-statistic and the 95% CI can be seen in this table.

<table>
<thead>
<tr>
<th>Light level</th>
<th>Device</th>
<th>Sw  (mm)</th>
<th>Mean±SD (mm)</th>
<th>Ndiff.&gt;0.5 mm</th>
<th>95% CI (mm)</th>
<th>t (DF=39)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low mesopic</td>
<td>Neuroptics</td>
<td>0.12</td>
<td>-0.09±0.16</td>
<td>0</td>
<td>(-0.14, -0.04)</td>
<td>-3.68</td>
</tr>
<tr>
<td></td>
<td>WAM-5500</td>
<td>0.10</td>
<td>-0.05±0.17</td>
<td>0</td>
<td>(-0.11, 0.01)</td>
<td>-1.76</td>
</tr>
<tr>
<td></td>
<td>PowerRef II</td>
<td>0.17</td>
<td>-0.04±0.27</td>
<td>3</td>
<td>(-0.12, 0.05)</td>
<td>-0.81</td>
</tr>
<tr>
<td>Mid mesopic</td>
<td>Neuroptics</td>
<td>0.11</td>
<td>-0.05±0.18</td>
<td>0</td>
<td>(-0.10, 0.01)</td>
<td>-1.61</td>
</tr>
<tr>
<td></td>
<td>WAM-5500</td>
<td>0.16</td>
<td>-0.12±0.23</td>
<td>2</td>
<td>(-0.20, -0.05)</td>
<td>-3.40</td>
</tr>
<tr>
<td></td>
<td>PowerRef II</td>
<td>0.25</td>
<td>-0.13±0.37</td>
<td>8</td>
<td>(-0.25, -0.01)</td>
<td>-2.25</td>
</tr>
<tr>
<td>High mesopic</td>
<td>Neuroptics</td>
<td>0.15</td>
<td>-0.08±0.23</td>
<td>0</td>
<td>(-0.15, 0.00)</td>
<td>-2.11</td>
</tr>
<tr>
<td></td>
<td>WAM-5500</td>
<td>0.24</td>
<td>-0.17±0.34</td>
<td>4</td>
<td>(-0.28, -0.06)</td>
<td>-3.24</td>
</tr>
<tr>
<td></td>
<td>PowerRef II</td>
<td>0.19</td>
<td>-0.11±0.28</td>
<td>2</td>
<td>(-0.20, -0.02)</td>
<td>-2.57</td>
</tr>
</tbody>
</table>

Sw: Within-subject standard deviation; Df: Degrees of freedom; Ndiff.: Number of cases in which the absolute difference was above 0.50 mm.

#### Agreement

On the one hand, the Bland and Altman plots comparing Neuroptics against WAM-5500 and PowerRef II at each illumination condition is shown in Figure 2. On the other hand, the one-way repeated measures ANOV A at low, mid and high mesopic conditions showed statistically significant differences in all cases (P<0.05). The F-statistic at low, mid and high mesopic was, respectively, F (1.56, 51.56)=21.89, F (2.00, 78.00)=59.86 and F (2.00, 78.00)=26.55. The Bonferroni post-hoc test is shown in Table 2. Given the fact that Neuroptics is the reference device, only the comparisons with respect Neuroptics are represented in this table, that is, the mean±SD and the 95% CI.

<table>
<thead>
<tr>
<th>Light level</th>
<th>Comparison</th>
<th>Mean±SD (mm)</th>
<th>Ndiff.&gt;0.5</th>
<th>95% CI (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low mesopic</td>
<td>Neuroptics-WAM-5500</td>
<td>-0.48±0.35</td>
<td>14</td>
<td>(-0.62, -0.34)</td>
</tr>
<tr>
<td></td>
<td>Neuroptics-PowerRef II</td>
<td>-0.28±0.56</td>
<td>12</td>
<td>(-0.50, -0.05)</td>
</tr>
<tr>
<td>Mid mesopic</td>
<td>Neuroptics-WAM-5500</td>
<td>-0.83±0.52</td>
<td>27</td>
<td>(-1.03, -0.62)</td>
</tr>
<tr>
<td></td>
<td>Neuroptics-PowerRef II</td>
<td>-0.70±0.55</td>
<td>29</td>
<td>(-0.91, -0.48)</td>
</tr>
<tr>
<td>High mesopic</td>
<td>Neuroptics-WAM-5500</td>
<td>-0.38±0.56</td>
<td>15</td>
<td>(-0.60, -0.16)</td>
</tr>
<tr>
<td></td>
<td>Neuroptics-PowerRef II</td>
<td>-0.61±0.54</td>
<td>23</td>
<td>(-0.82, -0.39)</td>
</tr>
</tbody>
</table>

Ndiff.: Number of cases in which the absolute difference was above 0.50 mm.

### Effect of Instrument Design

Effect of Instrument Design Firstly, the one-way repeated measures ANOVA showed statistically significant differences [F (1.32, 22.48)=12.48, P<0.05]. The Bonferroni post-hoc test is shown in Table 3. The mean±SD between configurations and the 95% CI can be seen in this table. Notice that on the one hand, the comparisons with the binocular openfield...
measurement are relevant since this is the condition closest to a natural-viewing situation and thus it can be considered the reference condition. On the other hand, the comparisons between the configuration 1 (monocular open field) and 2 (Neuroptics simulation) is also of interest since the only variable that changes is the simple presence of the clinician within the field of view of the patient. Lastly, these same pairwise comparisons are also shown in the form of Bland and Altman plots in Figure 3.

**DISCUSSION**

Pupil diameter is important not only in the field of refractive surgery and in the study of neurological disorders but also in the study of oculomotor functions such as accommodation. This study investigated the repeatability and agreement of the pupil measurement (at three different mesopic light levels) obtained with a reference pupillometer (VIP-200, Neuroptics) and two autorefractometers WAM-5500 (Grand Seiko) and PowerRef II (Plusoptix). Additionally, this study investigated the potential effects of instrument design on pupil measurements.

The three pupillometers studied herein showed similar levels of repeatability across the three mesopic levels. The Sw of the differences between test and retest ranged from 0.10 mm for

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Mean±SD (mm)</th>
<th>Ndiff&gt;0.5</th>
<th>95% CI (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Config.3-Config.1</td>
<td>-0.83±0.87</td>
<td>16</td>
<td>(-1.37, -0.28)</td>
</tr>
<tr>
<td>Config.3-Config.2</td>
<td>-0.53±0.40</td>
<td>8</td>
<td>(-0.77, -0.28)</td>
</tr>
<tr>
<td>Config.2-Config.1</td>
<td>-0.30±0.77</td>
<td>7</td>
<td>(-0.78, 0.18)</td>
</tr>
</tbody>
</table>

Config.: Configuration; Ndiff.: Number of cases in which the absolute difference was above 0.50 mm.
WAM-5500 in low mesopic to 0.25 mm for PowerRef II in mid mesopic. The source of these differences is likely related to the intrinsic variability of each device plus fluctuations in pupil size (which are typically smaller than 0.50 mm in amplitude). Overall, none of the differences showed a clinically significant bias (&lt;0.50 mm), although it must be taken into account that few individuals surpassed this limit value in both autorefractometers (Table 1).

Not surprisingly, Neuroptics showed the best repeatability values in all light levels (0% of the subjects surpassed the clinical limit of 0.50 mm for pupil differences). The results found for Neuroptics are in agreement with the study of Schallenberg et al[6] in which they found SD of the differences of 0.20 mm (at 0.04 lx) and 0.26 mm (at 0.40 lx) whereas in our study we obtained 0.12 mm (at 0.05 lx) and 0.11 mm (at 0.80 lx). Regarding the repeatability of the pupillometric functions of WAM-5500 and PowerRef II, only Jainta et al[5] measured human pupils with PowerRef II, they found mean intra-individual SD around 0.30 mm but considering only 8 subjects, excluding 10% of the raw data (5% at each end of the distribution) and measuring over a period of 2min (room illumination level was not specified). Despite the methodological differences with respect our study, their results are fairly similar to the ones obtained in this study and suggest that the PowerRef II (as well as the WAM-5500) has an acceptable repeatability. It is also worth mentioning that SDs did not systematically change across illumination levels (Table 1), e.g. the PowerRef’s largest variability is obtained in mid mesopic conditions whereas in the case of WAM-5500 and Neuroptics is in high mesopic.

In terms of agreement, neither WAM-5500 nor PowerRef II showed good agreement with Neuroptics in all illumination conditions of this study. Both disagreed similarly with the reference pupilometer (Neuroptics). The frequency (i.e. number of cases) at which the absolute differences were greater than the clinical limit of 0.50 mm is considerably large in all light levels, ranging from 12 individuals (out of 40) for the Poweref II in low mesopic to 29 individuals for the PowerRef II in mid mesopic (Table 2). Our results are fairly similar to those of the study of Bradley et al[13]. They compared the WAM-5500 with Neuroptics and they found at 1.00 lx of illumination an outlier frequency of 23 subjects (out of 49), in our case this was of 27 out of 40 at 0.80 lx. Regarding the PowerRef II, Kasthuriangan and Glasser[22-23] reported an underestimation of 0.50 mm in pupil size (roughly a 10%) when measured with the PowerRef II and using artificial pupils. Interestingly, an overestimation with respect the Neuroptics pupilometer was found in our study (Table 2). In this sense, the magnification factor of the pupil by the cornea lens, which has been estimated to be approximately 10%[26], as well as the instrument design differences (it will be discussed below) could explain these differences.

The fact that the two autorefractometers/pupilometers disagreed similarly with Neuroptics, and at the same time the fact that both devices showed a negative mean bias with respect the Neuroptics raised one question. It seems fair to think that instrument design might actually contribute to these differences. As previously mentioned, concepts such as proximal accommodation and instrument accommodation can explain differences in the accommodative response measured with different instrument designs. To this sense, considering the linkage between the accommodative and the pupillary system, the second experiment aimed to address the potential effects of instrument design (with vs without proximity cues and monocular vs binocular measurement) in pupillometry. As shown in Table 3 and Figure 3, binocular pupil measurements are significantly smaller (differences &gt;0.50 mm) than monocular pupil measurements, which is somehow expected due to halving of the retinal light flux. It is also shown that the presence of proximity cues can significantly affect pupil measurements in some patients, i.e. the inter-subject variability (or in other words, the mean SD of the differences shown in Table 3) is quite large, and it can actually be seen in Figure 3 that the difference between the “Neuroptics simulation” configuration (i.e. configuration 2, monocular measurement with proximity cues) and configuration 1 (i.e. monocular without proximity cues) is greater than 0.50 mm in some patients. Notice that these quite large variability found in pupil measurements is also in agreement with other studies that relate the pupil diameter with different accommodation stimulations[18,22,27]. Additionally, our findings showed a tendency towards a decrease in pupil diameter when the clinician was within the field of view when conducting pupil measurements (the mean bias when comparing configuration 2 minus configuration 1 was -0.30 mm), this tendency leads to consider that proximal accommodation might have induced significant pupil miosis in some patients (in closed-loop conditions)[10]. In fact, there is a counterbalance between the monocular midrasis and the proximal accommodation miosis when comparing the configuration 3 (i.e. a condition closer to natural-viewing) with configuration 2. Nonetheless, these large variabilities found in Table 3 suggest that even though accommodation can play a role in pupil measurements’ variability, there might be other covariables besides accommodation. Further studies that take into account accommodation control and report refraction data in synchronization with pupil data would be beneficial for more robust conclusions. One of the reasons why accommodation was not reported was the fact of not controlling the accommodative response in each patient, that is, pupil measurements were conducted without eye’s correction (as it is usually done in a clinical setting) therefore not all
patients started from the same accommodative level when fixating to the PowerRef II device (placed at 1 m distance from the patient’s eye).

In summary, pupil measurements with WAM-5500, PowerRef II and Neuroptics have similar repeatability. However, agreement between these autorefractometers and neuroptics is not good considering the limit of clinical significance of 0.50 mm in pupil differences. Despite binocular open-view pupillometers have some theoretical advantages over monocular closed-view pupillometers, differences in instrument design can contribute to disagreement in pupil measurements and it should be taken into consideration specially when comparing studies that have used different pupillometers.

ACKNOWLEDGEMENTS

Foundation: Supported by the Spanish Ministry of Economy and Competitiveness under the grant DPI 2014-56850-R.

Conflicts of Interest: Otero C, None; Aldaba M, None; Ferrer O, None; Gascón A, None; Ondategui-Parra JC, None; Pujol J, None.

REFERENCES


