Journal of Clinical Neuroscience 33 (2016) 79–82

Contents lists available at ScienceDirect

Journal of Clinical Neuroscience

journal homepage: www.elsevier.com/locate/jocn

Clinical Study

Inter-device reliability of the NPi-100 pupillometer

Weidan Zhao [a], Sonja Stutzman [b], DaiWai Olson [b], Ciji Saju [b], Margaret Wilson [b], Venkatesh Aiyagari [a,b,*]

[a] Department of Neurological Surgery, UT Southwestern Medical Center, Dallas, TX, USA
[b] Department of Neurology and Neurotherapeutics, UT Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75093, USA

ARTICLE INFO

Article history:
Received 18 February 2015
Accepted 1 January 2016
Available online xxxx

Keywords:
Interrater reliability
Neurological illness
Pupil assessment
Pupillometery
Traumatic brain injury

ABSTRACT

The purpose of this study was to explore the inter-device reliability of NPi-100 pupillometers (NeuroOptics, Inc.). The pupillary examination is a fundamental element of the neurological exam. Current evidence suggests that the traditional examination of the pupil with a hand held flashlight has limited inter-rater reliability. Automated pupillometers were developed to provide an objective scoring of pupil size and reactivity. However, there are no data examining inter-device reliability of automated pupil assessments. This study included 210 paired pupillometer measurements were obtained by 33 practitioners from 20 patients at risk for cerebral edema. There was no statistically significant difference between the mean maximum pupil size at rest, the minimum pupil size during light stimulation, and the mean pupil reactivity, for both the right and left eye, when assessed by two investigators, each with a different pupillometer. In addition, Cohen's Kappa assessments of pupil size and reactivity revealed an almost perfect agreement between the two pupillometers for the maximum pupil size, the minimum pupil size, and for pupil reactivity for both eyes. There is a high inter-device reliability of automated pupillary assessments by two practitioners examining the same patient using different NPi-100 pupillometers.

1. Introduction

Examination of the pupil is an essential part of the neurological examination [1]. It provides the clinician with data concerning the functional statuses of cranial nerves II and III and the sympathetic nervous system. Under normal conditions, a bright light stimulates the retinal cells and the pupil afferents are carried by the optic nerve (CN-II) and optic tract to synapse in the pretectum of the midbrain. Fibers project from the pretectum to the Edinger-Westphal nucleus in the dorsal midbrain. The stimulus is processed and parasympathetic pupillary efferents are carried by the oculomotor (CN-III) nerve, which travel through the cavernous sinus to the superior orbital fissure of the eye and the ciliary ganglion, where they synapse with the short posterior ciliary nerves that innervate the iris. CN III controls the parasympathetic pupil constriction in response to light stimulation. The pupilloconstrictor parasympathetic input is balanced by pupillodilator sympathetic input from the superior cervical ganglion. When both the left and right afferent and efferent pathways are functionally intact, both pupils should be equal in size and constriction when stimulated with the same light source. Differences in either measurement are often linked to neurological conditions, such as damage to CN III due to brainstem compression or transtentorial herniation [2–4].

Traditionally, the pupillary examination was scored strictly as a subjective assessment of the size and shape of the pupil prior to the manual application of a light source and the presence and speed of pupil reactivity to a bright light. However, studies suggest that the manual examination may have limited inter-rater reliability, as the examiners are of different levels of skill and training and are allowed to use a variety of non-standardized flashlights, penlights, and other handheld light sources [5–8]. To overcome this, automated pupillary assessment technology has been developed that is able to provide an objective measurement of the initial and final size of the pupil and to grade the speed of pupil contraction in response to a light stimulus.

The NPi-100 pupillometer (Neuroptics Inc.) is a portable, battery-operated device with a light source, a liquid crystal display (LCD) screen and a digital video camera that measures, records, and analyzes both the pupil size and reactivity. Once the device has focused on the target pupil, a white light stimulus is flashed. The pupil light response is then calculated immediately and a Neurological Pupil index (NPi) is reported thereafter. A number of pupil variables such as size, latency, constriction velocity and dilatation velocity are parameters of the NPi algorithm. Each of

* Corresponding author. Tel.: +1 214 648 6410; fax: +1 214 648 0341.
E-mail address: venkatesh.aiyagari@utsouthwestern.edu (V. Aiyagari).

http://dx.doi.org/10.1016/j.jocn.2016.01.039
0967-5868/

these variables is compared against a mean of a reference distribution of healthy subjects, taking the difference and standardizing it. The set of all the standardized differences are then combined to fall within a scale set between 0 and 5. A score equal to or above 3 indicates that the pupil measurement is within the boundaries of normal pupil behavior as defined by the NPi model. A score below 3 suggests that the reflex is abnormal, and the absence of a constriction is reported as zero. Additionally, a difference between the left and right NPi scores (even if both are above 3) is a sign of pupil abnormalities.

As the pupillometer is slowly becoming more common in hospitals across the US, it is important to validate its effectiveness [9,10]. Although past studies have demonstrated that the pupillometer is more accurate and reliable when compared to the traditional manual pupillary reflex assessment, no studies have focused on the inter-device and inter-rater reliability of the device [11,12]. The purpose of this prospective study is to explore the inter-rater reliability of different pupillometers when operated by different users. We hypothesize that the pupillometers have a statistically significantly high inter-device and inter-rater reliability.

2. Methods

Twenty patients admitted to the Neurocritical Care Unit or Acute Stroke Unit at the University of Texas Southwestern Medical Center, and the Surgical Intensive Care Unit of Parkland Memorial Hospital located in Dallas, TX were consented to participate in the study. Patients were considered eligible to participate if they were admitted with a neurological or neurosurgical diagnosis that placed them at risk for cerebral edema. All patients in the study had pre-existing medical orders for serial neurologic and pupillary examinations prior to consent. Examiners eligible for participation included medical students, neuroscience registered nurses (RN), nurse practitioners (NP), resident physicians, fellow physicians, researchers and attending (faculty) physicians.

Staff that normally provided care to consented subjects was asked to perform a pupillometer assessment using the NPi-100 pupillometer. Pupillometers (Neuroptics, Inc) were programmed to a research mode which blinded the examiner to the results. The pupillometer screen would display 'Ok' only when the device obtained an adequate signal to provide size and reaction values. The examiners all received a short education (less than 5-minutes of verbal training) on how to use the device and were instructed to make a maximum of three attempts to obtain a pupillometer reading (scored by device as 'Ok'). Paired pupillometer assessments were completed within a five-minute period and consisted of two separate assessments by two different clinicians in the same session and under identical ambient conditions. To assess inter-device reliability, each staff member used a different pupillometer device on the same subject. Four different pupillometers were used during the study and data from each were downloaded and converted to msExcel for analysis. Statistical analyses were performed using SAS for windows v9.3. Pupil size was rounded to the nearest whole number and treated as a continuous variable. NPi scores were initially examined as continuous variables and then dichotomized to represent normal (NPI $\geqslant 3$) versus abnormal (<3.0) based on manufacturer recommendations. Dichotomized variables were treated as nominal data.

3. Results

The study included 210 prospective paired pupillometer measurements obtained between July and August of 2014 by 33 practitioners (28 RNs, two MDs, one NP, one medical student and one PhD research coordinator). Two separate examiners each used a different pupillometer device to score 105 OS and 105 OD (210 paired assessments) pupillary responses from 20 patients. There were 56/420 (13.3%) independent assessments with incomplete (missing) data. In 9/210 (4.3%) paired assessments (OS = 5, OD = 4), neither examiner was able to obtain a pupillometer reading, and there was one subject for whom neither examiner was able to obtain a reading for either eye.

Mean values for size and reactivity measures from the two devices were similar (Table 1, Fig. 1). There was no statistically significant difference in the mean maximum pupil size (at rest and before stimulation) for the OS assessments, comparing readings from the first pupillometer (PM1) and second pupillometer (PM2) device (3.8 mm vs 4.0 mm; p = 0.27), nor for OD assessments (3.6 vs 3.8; p = 0.74). There was no statistically significant difference in the mean minimum pupil size (size after light stimulation) for the OS assessments, comparing PM1 versus PM2 (2.8 mm vs 3.0 mm; p = 0.64), nor for OD assessments (2.6 mm vs 2.6 mm; p = 0.44). There was no statistically significant difference in the mean NPi for the OS assessments, comparing PM1 versus PM2 (3.9 mm vs 3.9 mm; p = 0.36), nor for OD assessments (4.2 mm vs 4.3 mm; p = 0.82).

Cohen's Kappa (k) assessments of size and reactivity were examined for agreement using interpretation described by Viera and Garrett [13] based on work of Landis and Koch [14]. There was almost perfect agreement between PM1 and PM2 for the maximum pupil size of OS observations (k = 0.97), and of OD observations (k = 0.91). Excluding observations of non-reactive (fixed) pupils, there was high correlation between PM1 and PM2 for the minimum pupil size of OS observations (k = 0.96), and of OD observations (k = 0.98). Correlation values were slightly higher when readings from non-reactive pupils were imputed for extremely small NPi values (e.g., 0.1 imputed to 0.0). Excluding zero values (0.0) as non-reactive pupils, and retaining these small values (0.1), there was moderate agreement for NPi measures from PM1 and PM2 for the OS readings (k = 0.50), and OD readings (k = 0.90). Including non-reactive pupils (0.1 imputed to 0.0), there was almost perfect agreement for the OS readings (k = 0.99) and OD readings (0.90). In five observations where both devices scored the pupil as non-reactive, two independent observers also scored the pupil as non-reactive. In two observations where both devices scored the pupil as non-reactive, one observer scored the pupil as non-reactive and the other observer scored the pupil as sluggishly reactive.

Table 1
Comparing mean size and reactivity for two devices by left and right eye

	Left Eye		P-Value	Right Eye		P-Value
	Device 1	Device 2		Device 1	Device 2	
Maximum Size*	3.8 (1.1)	4.0 (1.6)	0.27	3.6 (1.1)	3.8 (1.1)	0.74
Minimum Size*	2.8 (1.2)	3.0 (1.5)	0.64	2.6 (1.0)	2.6 (0.6)	0.44
NPi*	3.9 (1.2)	3.9 (1.4)	0.36	4.2 (1.0)	4.3 (0.8)	0.82

* Values displayed as mean (s.d.).
NPi = Neurological Pupil index.