

Interrater Reliability of Pupillary Assessments

DaiWai M. Olson¹ · Sonja Stutzman¹ · Ciji Saju¹ · Margaret Wilson¹ · Weidan Zhao¹ · Venkatesh Aiyagari^{1,2}

Published online: 17 September 2015

© Springer Science+Business Media New York 2015

Abstract

Background Subjective scoring of pupil reactivity is a fundamental element of the neurological examination for which the pupillometer provides an objective measure.

Methods This single-blinded observational study examined interrater reliability of pupil exam findings between two practitioners and between practitioners and a pupillometer.

Results From 2329 paired assessments, the interrater reliability between practitioners was only moderate for pupil size ($k = 0.54$), shape ($k = 0.62$), and reactivity ($k = 0.40$). Only 33.3 % of pupils scored as non-reactive by practitioners were scored as non-reactive by pupillometry.

Conclusions Despite the strong emphasis placed on the traditional pupil examination, especially for patients with a neurological illness, there is limited interrater reliability for subjective scoring of pupillary assessments. Thus, the use of automated pupillometers should be examined as a potential method to increase the reliability of measuring of pupil reactivity.

Keywords Pupillometry · Interrater reliability · Neurological illness · Traumatic brain injury · Pupil assessment

On behalf of The University of Texas Southwestern Medical Center, Neurology & Neurotherapeutics.

✉ DaiWai M. Olson
daiwai.olson@utsouthwestern.edu

¹ Department of Neurology & Neurotherapeutics, The University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75390-8855, USA

² Department of Neurological Surgery, The University of Texas Southwestern Medical Center, Dallas, TX, USA

Introduction

Careful examination of the pupil by a trained practitioner is an established element of the neurologic examination [1–4]. In normal conditions, the pupils are equal in both size and reactivity to light [5]. The detection of a non-reactive pupil in a patient with an acute neurological disease is considered to be an event of vital importance and often triggers a variety of diagnostic (e.g., stat brain CT) and therapeutic (e.g., mannitol infusion) maneuvers. Numerous clinicians have considered pupil size and reactivity as notable in multi-parameter predictive models of outcomes and used such models to detect and define the location of the intracranial mass lesions [6–9]. In the management and prognosis of traumatic brain injury (TBI), for example, abnormalities of pupillary response or anisocoria (pupil size asymmetry) have been associated with neurologic deterioration and secondary brain injury and are correlated with poor neurologic outcomes [10, 11]. The literature is storied with triage, prognostication, and treatment algorithms based in whole or part on pupillary abnormality [12–17].

The traditional and most common method of pupillary assessment is performed by a practitioner using a handheld light source [2]. Automated pupillary assessment is now commercially available and is being utilized for patients hospitalized with neurologic injury [18, 19]. For example, the NeuroOptics® NPⁱTM-100 Pupillometer (NeuroOptics Inc., Irvine, CA, USA) is a portable handheld device that provides automated readings of one pupil at a time. Using this device, the observer targets the pupil which is then scored for size using proprietary software. Next, a burst (0.8 s) of light is emitted at a fixed distance from the patient and a high-speed video recording of the pupillary response is

used to measure the maximum pupil size, minimum pupil size, constriction velocity and neuro pupillary index (NPiTM) [20]. Automated pupillometry may reduce observer bias and provide insight to intracranial pathology [21, 22]. However, three citations, a case study [23], a study of healthy volunteers receiving phenylephrine versus sterile water [24], and an interrater reliability including only nurse examiners find limitations in pupillometry [25]. This study fills an important gap by examining the interrater reliability of subjective pupillary assessments in a diverse group of practitioners, and the comparison between subjective and objective pupillary assessments.

Methods

This prospective, blinded, observational study was approved by the Institutional Review Board at the University of Texas Southwestern Medical Center and is registered with ClinicalTrials.gov (NCT02296606). Only patients with a neurological or neurosurgical diagnosis with pre-existing orders for serial pupil/neurologic exams were approached for consent. Hospital practitioners, both physicians (MDs) and nurses (RNs), were informed via email and at meetings that they would be asked to participate as assessors. Only practitioners who would routinely perform pupil assessments were invited to participate (e.g., neurologist, neurosurgeon, or neuroscience RN). Each study subject was examined by a convenient pairing of practitioners so that study observations could be obtained without interrupting daily routines.

Following consent, pupil exams were performed and independently scored by two practitioners within a 5-min time window. Each observation required each practitioner to assess two (OU) pupils. Thus, each observation provided three sets of scores (two practitioner scores and one device score) for the size, shape (traditional assessments only), and reactivity of the pupils. Care was taken to ensure that the ambient lighting and the physiological state of the patient was identical for both measurements. Each practitioner used a flashlight or penlight to observe and score the left (OS) and right (OD) pupil for the initial size (in mm), shape (round or irregular), and reactivity (brisk (normal), sluggish (abnormal), or non-reactive (fixed)). To ensure that the environment of study was representative of practice, practitioners were permitted to use the light source they would routinely use when performing a pupil exam.

After practitioners had completed the subjective pupil assessment, a member of the study team obtained OS and OD pupillometer readings using the automated pupillometer (NPiTM-100) within 5 min the practitioner exams, again ensuring similar lighting and physiological

conditions. The pupillometer was set to ‘research mode’ so that practitioners and study team members were blinded to the actual size and reactivity measures obtained by the pupillometer.

Participants

Study subjects and practitioners were from two hospitals (a county teaching hospital and a university hospital) and from four units, including two neurocritical care units, a stroke unit, and a general neurology ward. One hundred and twenty-seven patients participated as study subjects (Table 1). Study subjects were predominantly male (56.7 %), Caucasian (88.1 %), and did not require surgical intervention (55.1 %). Nine (7.1 %) subjects had a prior ocular history that may have impacted pupillary exams (physical eye injury, cataract, glaucoma, or prosthetic implant). Study subjects were observed an average of 8.8 times (IQR = 3–9) during the study. Two hundred and twenty-two practitioners (RN = 194, MD = 28) participated in the study and performed traditional pupil assessments. The automated pupillometer (NPiTM-100) assessments were performed by three trained research investigators.

Statistical Analysis

Statistical analyses were performed using SAS v 9.3 for Windows. Cohen’s Kappa coefficient (*k*) is a measure of interrater agreement and was calculated for each pupil size, shape, and reactivity, first as a composite score and then examined as separate scores [26]. Kappa requires agreement for the number of columns and rows. Therefore, when present, null cells were inserted and assigned a weight of 0.000001 for interrater reliability [27]. Bland–Altman plots were constructed to examine measurement differences between human observers and pupillometer readings; regression lines were then fitted to examine for bias. Subjective pupil size estimates were scored as equal if the difference between the two raters was ≤ 1.0 mm and dichotomized as ≤ 3 versus > 3 mm. Kappa values were interpreted as slight (0–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), or almost perfect (0.81–1.0) [28].

Sample Size

Power calculation was performed using SAS v9.3, assuming $K_0 = 0.5$, alpha 0.05, and a power of 0.8 ($\beta = 0.20$). A targeted sample size of 1163 paired observations was based on the desire to remain conservative given the limited literature available from which to estimate the required sample size [20, 24, 29].

Table 1 Demographics for patient-subjects

	All Patients (<i>N</i> = 127)	Surgical (<i>N</i> = 57)	Non-surgical (<i>N</i> = 70)	<i>p</i> value
Age mean (S.D.)	55.1 (15.6)	54.7 (15.7)	55.4 (15.6)	0.81
Sex				
Male	72 (56.7 %)	34 (59.6 %)	38 (54.3 %)	0.55
Female	55 (43.2 %)	23 (40.4 %)	32 (45.7 %)	
Race				
Caucasian	103 (81.1 %)	47 (82.5 %)	56 (80 %)	0.78
African Am.	12 (9.4 %)	4 (7.0 %)	8 (11.4 %)	
Other	12 (9.4 %)	6 (10.5 %)	6 (8.6 %)	
Ethnicity				
Non-Hispanic	106 (83.5 %)	47 (82.5 %)	59 (84.3 %)	0.81
Hispanic	21 (16.5 %)	10 (17.5 %)	11 (15.7 %)	
Primary diagnosis				
Ischemic stroke	43 (33.9 %)	7 (12.3 %)	36 (51.4 %)	<0.001
Hemorrhagic stroke	40 (31.5 %)	20 (35.1 %)	20 (28.6 %)	
Neoplasm	23 (18.1 %)	18 (31.6 %)	5 (7.1 %)	
Other	21 (16.5 %)	12 (21.0 %)	9 (12.9 %)	
Vision or ocular History	9 (7.1 %)	5 (8.8 %)	4 (5.7 %)	0.51

Results

There were a total of 1166 observations. Practitioners participated in an average of 6.5 observations (IQR 1–6). The results in Table 2 summarize findings from 2329 (1166 OS, 1163 OD) paired subjective pupillary assessments by practitioners, and 2192 (1099 OS, 1093 OD) automated pupillometer device assessments (3 OD observations were null for a patient with a prosthetic eye). For the sake of brevity, only the results of the OU examination for size, shape, and reactivity are presented. The results of the OS and OD examination were similar and are presented in Table 2. Practitioner agreement for the omnibus pupil assessment (size, shape, and reactivity) was low ($k = 0.26$, 95 % CI 0.23–0.29).

Agreement on Pupil Size and Shape

Pupil size was dichotomized as being ≤ 3 versus > 3 mm in size. Practitioner agreement on pupil size was moderate ($k = 0.54$; 95 % CI 0.50–0.57). Agreement with device on pupil size was fair ($k = 0.29$; 95 % CI 0.27–0.32 and $k = 0.31$; 95 % CI 0.28–0.34 for the first and second practitioners, respectively). Figure 1 provides the full range of pupil size estimates for human and pupillometer observations. Anisocoria was scored as present if the difference between OS and OD was > 1 mm. Practitioner agreement on anisocoria was moderate ($k = 0.60$; 95 % CI 0.54–0.64). Shape was scored as round or irregular.

Agreement on pupil shape was moderate ($k = 0.62$; 95 % CI 0.55–0.69).

Agreement on Pupil Reactivity

Practitioner agreement for reactivity, (reactive versus fixed), was moderate ($k = 0.64$; 95 % CI 0.58–0.71). Agreement with automated pupillometry device on pupil reactivity (dichotomized as reactive or fixed) was moderate ($k = 0.52$; 95 % CI 0.44–0.60) for the first practitioner and fair ($k = 0.40$; 95 % CI 0.32–0.49) for the second practitioner. There were 189 practitioner observations of a fixed pupil. Of these, 94/189 (49.7 %) were scored as fixed by both practitioners and 58/189 (33.3 %) were scored as fixed by pupillometry. Practitioner agreement on pupil reactivity scored as fixed (non-reactive), sluggish, or brisk was fair ($k = 0.40$; 95 % CI 0.36–0.44). There were 83 observations of non-reactive pupil as scored by pupillometer. Of these, the first practitioner also scored the pupil as non-reactive in 58/83 (69.9 %) observations and the second practitioner scored the pupil as non-reactive in 46/83 (55.4 %) observations. When the sample was confined to only observations of fixed (non-reactive) pupils, practitioner agreement on reactivity was only fair [$k = 0.28$ (OS); $k = 0.47$ (OD)]. Notably, there were 21 observations where the pupillometer could not be used to determine if the pupil was reactive or not (e.g., periorbital edema prevented the examiner from fully visualizing the pupil when using the pupillometer). Of 28 OS, and 17 OD,

Table 2 Practitioner agreement for pupillary observations

	Left eye		Right eye		Both eyes	
	Kappa	% agree	Kappa	% agree	Kappa	% agree
Entire cohort (<i>N</i> = 1166)						
Size	0.54 (0.49–0.59)	78.0	0.54 (0.49 to .59)	79.8	0.54 (.50–.58)	78.9
Anisocoria	–	–	–	–	0.60 (.55–.64)	80.8
Shape	0.68 (.61–.76)	94.3	0.06 (–.07 to .19)	97.6	0.62 (.55–.69)	96.0
Reactivity	0.48 (.43–.53)	80.0	0.30 (.23 to .36)	78.0	0.40 (.36–.44)	79.0
Reactive versus fixed	0.69 (.62–.77)	95.0	0.51 (.37 to .64)	96.7	0.64 (.58–.71)	95.9
Two RN examiners (<i>N</i> = 952)						
Size	0.54 (.49–.60)	78.6	0.52 (.46 to .58)	79.7	0.53 (.49–.57)	79.1
Anisocoria	–	–	–	–	0.60 (.54–.65)	80.8
Shape	0.70 (.62–.78)	94.7	0.01 (.005 to .01)	97.8	0.64 (.56–.72)	96.3
Reactivity	0.49 (.43–.55)	80.3	0.33 (.26 to .40)	78.7	0.42 (.38–.47)	79.5
Reactive versus fixed	0.69 (.61–.78)	95.2	0.59 (.45 to .73)	97.1	0.67 (.59–.74)	96.2
Two MD examiners (<i>N</i> = 33)						
Size	0.49 (.20–.78)	74.3	0.77 (.55 to .98)	88.6	0.63 (.45–.81)	81.4
Anisocoria	–	–	–	–	0.70 (.42–.97)	87.8
Shape	0.64 (.19–1.0)	94.2	a	96.9	0.55 (.11–.99)	95.6
Reactivity	0.73 (.45–1.0)	91.4	b	78.8	0.54 (.30–.79)	85.3
Reactive versus fixed	0.79 (.38–1.0)	97.1	b	93.9	0.55 (.09–1.0)	95.6
One RN and one MD examiner (<i>N</i> = 179)						
Size	0.51 (.39–.64)	76.0	0.56 (.44 to .68)	78.8	0.54 (.45–.63)	77.4
Anisocoria	–	–	–	–	0.57 (.44–.69)	79.7
Shape	0.61 (.42–.79)	92.1	0.24 (.14 to .63)	96.6	0.55 (.38–.73)	94.4
Reactivity	0.38 (.24–.51)	75.8	0.07 (–.07 to .20)	74.0	0.25 (.15–.36)	74.9
Reactive versus fixed	0.67 (.49–.85)	93.8	–0.02 (–0.04 to 0.002)	94.9	0.54 (.36–.72)	94.4

*95 % CI reported within parentheses for Kappa values. There was no agreement between two MDs for the one pupil scored as irregular (a), nor for the two pupils scored as fixed (b)

observations of a dilated pupil (>6.0 mm), practitioner agreement on reactivity was moderate or high [$k = 0.78$ (OS); $k = 0.88$ (OD)]. Of 407 OS and 450 OD observations of a small pupil (<3.0 mm), practitioner agreement on reactivity was fair [$k = 0.59$ (OS); $k = 0.23$ (OD)].

Agreement for RN and MD Practitioners

Variability within and between practitioners with different training was explored for four subsets: (1) the entire cohort, (2) both practitioners were RNs, (3) both practitioners were MDs, and (4) one RN and one MD performed observations (Table 2). Agreement for pupil size (≤ 3 vs. > 3 mm) was similar for all four subsets ($k = 0.54$, 0.53 , 0.63 , and 0.54 respectively). Agreement for pupil reactivity (fixed versus reactive) was similar for all four subsets ($k = 0.64$, 0.67 , 0.55 , and 0.54 , respectively). Agreement for pupil size (≤ 3 vs. > 3 mm) was fair ($k = 0.30$; 95 % CI 0.27–0.32) between RNs and the device, and fair ($k = 0.38$; 95 % CI

0.31–0.45) between MDs and the device. Agreement for pupil reactivity (fixed versus reactive) was moderate ($k = 0.47$; 95 % CI 0.40–0.53) between RNs and the automated device, and moderate ($k = 0.42$; 95 % CI 0.22–0.61) between MDs and the automated device.

Discussion

This study explored interrater reliability of the traditional pupil exams performed by two independent practitioners, and the relationship between manual examinations and automated pupillometer results. The finding of limited interrater reliability for the size, shape, and reactivity scores between two practitioners confirms the findings from a smaller study examining traditional pupil exams from six practitioners and 20 patients. The high percent of agreement seen in Table 2 likely reflects the high number of normal pupillary findings. The majority of pupil exams

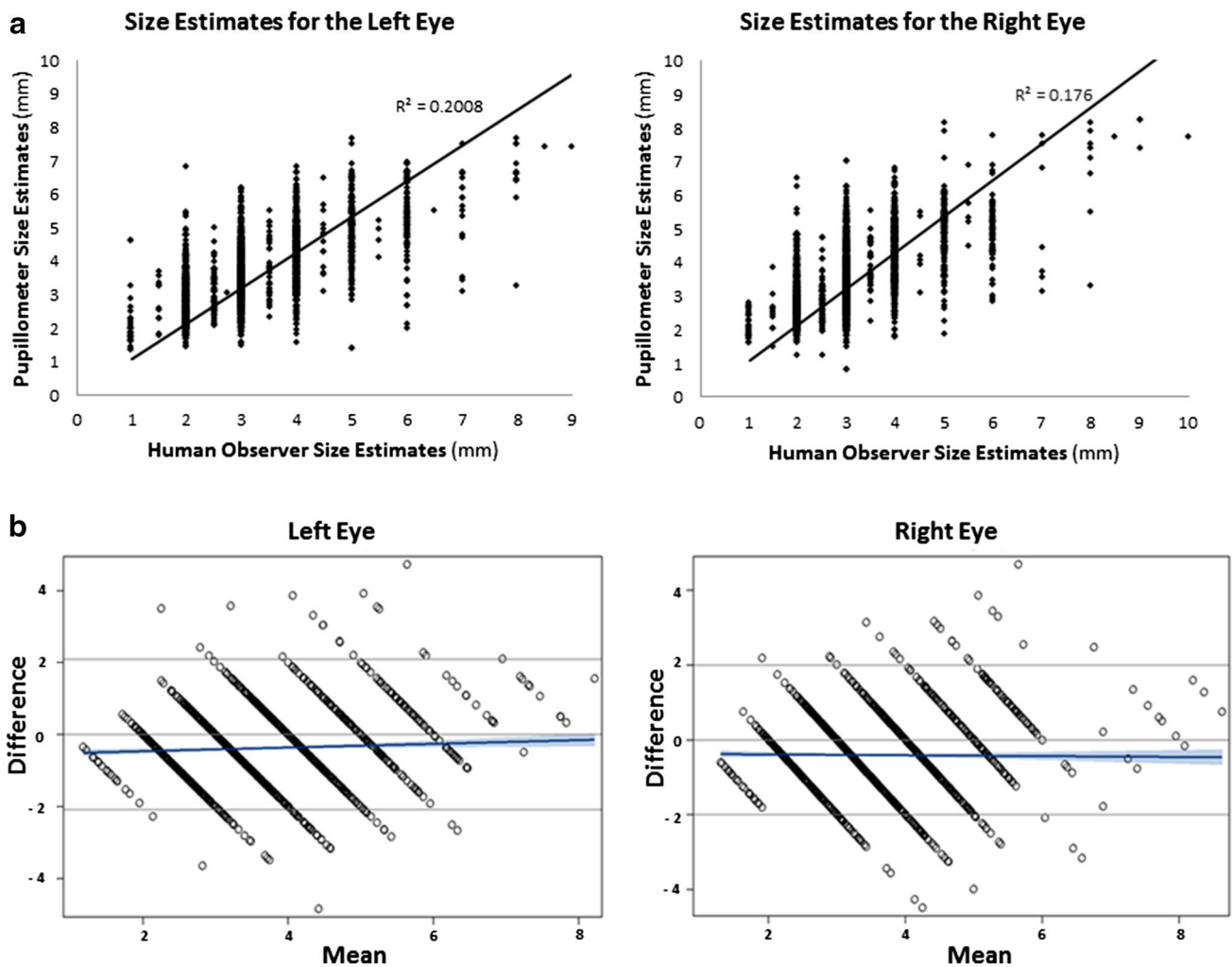


Fig. 1 **a** Comparison of pupil size estimates by human observers and automated pupillometer. **b** Bland–Altman plots for difference in size (includes regression line with 95 %CI)

were documented as round (96.9 %), briskly reactive (79.0 %), and 3–4 mm (50.9 %) in size.

Cohen’s Kappa is more robust when one option (e.g., reactive) is more common by chance or guessing. Lower kappa values likely reflect the lower rates of agreement, seen when the pupil finding is abnormal [30]. This is demonstrated by the data presented in Table 3 where reactivity is dichotomized. The raw percent for the expected outcome (reactive) is 95.7 % (2135/2230). The raw percent for the unexpected outcome (fixed) is 49.7 % (94/189). Thus, kappa provides an omnibus index of the level of agreement.

Although there was high agreement that a pupil was reactive (95.7 % or 2135/2230), there was only 49.7 % (94/189) agreement that the pupil was fixed. These data suggest that practitioners are generally in agreement that a pupil is normal, but often disagree when one practitioner reports an abnormal finding. Agreement on reactivity was

highest when the pupil was >6.0 mm in size (91.1 % [41/45]). Automated pupillometry provides a fixed light source, for a fixed period of time, and then provides an objective measure of the change in pupil size [31]. Subjective pupil examination by a human observer is inadequate, especially when the findings are clinically relevant. The low agreement on a fixed pupil has specific clinical implications given that a physician may decide to obtain diagnostic

Table 3 Comparison of manual observations between two observers with pupil reactivity scored as reactive or fixed

		Observer 2		
		Reactive	Fixed	
Observer 1	Reactive	2135	48	2183
	Fixed	47	94	141
		2182	142	2324

imaging (brain CT), or treat intracranial pressure (e.g., mannitol), based on the finding of a fixed or non-reactive pupil.

Two abnormal findings, anisocoria and non-reactive pupils were explored in depth. There is disagreement on the definition of anisocoria ranging from pupil size difference of 0.25 mm to >1.0 mm [1, 3, 32]. To remain conservative, we scored anisocoria only if the OS to OD difference was >1.0 mm. The low agreement of anisocoria between practitioner and device ($k = 0.14$) assessments likely reflects the fewer number of anisocoria events (153/1166) noted in device readings compared to the number noted by the first (450/1161) or second practitioner (445/1159). It is noteworthy that anisocoria is a dynamic phenomenon and the low agreement may be accounted to temporal, rather than actual state changes.

The absence of a pupillary light reflex (fixed or non-reactive pupil) as a new finding in a hospitalized patient may signal compression of the oculomotor nerve or distortion of the midbrain and warrants emergency assessment and treatment. When pupil exam findings were dichotomized as reactive versus non-reactive, there was poor agreement among the entire cohort ($k = 0.64$). It is unknown whether the finding that practitioners scored 189 pupils as non-reactive, compared to 83 scored as non-reactive by pupillometry reflects an inability of the human eye to perceive slight or slow movement. On the other hand, Kramer et al. [23], recently reported a case study wherein a neurologist that observed the pupil for 7–9 s was able to detect a 1 mm size change that was undetectable by pupillometry. This suggests a potential limitation in examining interrater reliability across a diverse population (e.g., a neurologist may have greater expertise). However, the selection of a diverse group of practitioners was purposeful and strengthens the external validity of the study. Internal validity of this study could have been enhanced by limiting enrollment to only attending neurointensivists performing examinations on patients in primary position with standardized light sources and environmental conditions (background illumination). The diverse group of practitioners in this study includes registered nurses, nurse practitioners, neurologists, neurosurgeons, and resident physicians with varied experience performing the exam that they would normally perform. Thus, our sample more accurately reflects the population of practitioners who would perform patient pupil exams throughout the day, document the results of their exam, and compare those results against prior results. Therefore, the findings from this study would be expected to have high generalizability.

Practitioners were not completely blinded to aims of the study, though they were blinded to the findings of the

second practitioner and the automated pupillometer. They were aware that their results were being compared and that a pupillometer device was being used. Each practitioner examined and scored pupils as they normally would. The limitation of this approach is lower internal reliability but the more accurate depiction of true practice enhances external validity, and thus the results are more easily generalizable across practice settings. Figure 1 demonstrates heteroscedasticity of size estimates; underestimating average size in smaller pupils, and overestimating average size in larger pupils. It is arguable that practitioners may have tried harder than normal to obtain an accurate assessment (reflected as higher interrater reliability); or, knowing that there was a second examiner may have led them to be less precise (reflected as lower interrater reliability). [20] However, given that 222 practitioners performed 4658 pupil observations, it seems reasonable to reject excessive observer effect.

A final limitation is the number of times an automated pupillometer reading could not be obtained (5.9%). Baseline data on pupil function, shape, or reactivity were not collected, nor were any data collected on whether patients had a history of glaucoma or iridectomy, and it is possible that this could have impacted agreement. Automated pupillometer readings were generally completed in less than 1-min and research staff were instructed to make no more than three attempts to obtain any one reading. The most common reasons for inability to obtain readings included periorbital edema, patient movement (especially in the patient with impaired cognition), cataract or prosthetic eye (scored as reactive by three practitioners). Although not statistically different, the number of missing pupillometer readings was higher during the first half of the study compared to the second half, suggesting that there may be an operator learning curve.

Conclusion

There is low interrater reliability among diverse practitioners performing a manual pupillary exam. These findings confirm and extend prior work suggesting that there is inadequate agreement between practitioners who solely rely upon traditional pupillary assessments for patients with neurologic injury [33]. There is a need to standardize the assessment of pupillary function in order to provide higher reliability. The use of automated pupillometry could be considered as a mechanism to standardize practice when there is a need for accurate assessment of the pupil size and reactivity.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Funding The study received no direct funding. Pupillometers were loaned to the university by Neuroptics Inc., at no cost for the duration of the study.

References

- Greenberg MS, Arredondo N. Handbook of neurosurgery. 6th ed. Lakeland: Greenberg Graphics, Thieme Medical Publishers; 2006.
- AANN. AANN Core curriculum for neuroscience nursing. 5th ed. Glenview: American Association of Neuroscience Nurses; 2010.
- Campbell WW. DeJong's The Neurologic examination. 7th ed. Philadelphia: Lippincott Williams & Wilkins; 2005.
- Plum F, Posner JB. The diagnosis of stupor and coma. Ed. 3. ed. Philadelphia: Davis; 1980.
- Blumenfeld H. Neuroanatomy through clinical cases. Sunderland: Sinauer; 2002.
- Marmarou A, Lu J, Butcher I, et al. Prognostic value of the Glasgow Coma Scale and pupil reactivity in traumatic brain injury assessed pre-hospital and on enrollment: an IMPACT analysis. *J Neurotrauma*. 2007;24(2):270–80.
- Narayan RK, Greenberg RP, Miller JD, et al. Improved confidence of outcome prediction in severe head injury. A comparative analysis of the clinical examination, multimodality evoked potentials, CT scanning, and intracranial pressure. *J Neurosurg*. 1981;54(6):751–62.
- Chesnut RM, Gautille T, Blunt BA, Klauber MR, Marshall LE. The localizing value of asymmetry in pupillary size in severe head injury: relation to lesion type and location. *Neurosurgery*. 1994;34(5):840–5 (**discussion 845–846**).
- Majdan M, Steyerberg EW, Nieboer D, Mauritz W, Rusnak M, Lingsma HF. Glasgow coma scale motor score and pupillary reaction to predict six-month mortality in patients with traumatic brain injury: comparison of field and admission assessment. *J Neurotrauma*. 2015;32(2):101–8.
- Braakman R, Gelpke GJ, Habbema JD, Maas AI, Minderhoud JM. Systematic selection of prognostic features in patients with severe head injury. *Neurosurgery*. 1980;6(4):362–70.
- Ritter AM, Muizelaar JP, Barnes T, et al. Brain stem blood flow, pupillary response, and outcome in patients with severe head injuries. *Neurosurgery*. 1999;44(5):941–8.
- Tien HC, Cunha JR, Wu SN, et al. Do trauma patients with a Glasgow Coma Scale score of 3 and bilateral fixed and dilated pupils have any chance of survival? *J Trauma*. 2006;60(2):274–8.
- Manley GT, Larson MD. Infrared pupillometry during uncal herniation. *J Neurosurg Anesthesiol*. 2002;14(3):223–8.
- Goebert HW Jr. Head injury associated with a dilated pupil. *Surg Clin North Am*. 1970;50(2):427–32.
- Clusmann H, Schaller C, Schramm J. Fixed and dilated pupils after trauma, stroke, and previous intracranial surgery: management and outcome. *J Neurol Neurosurg Psychiatry*. 2001;71(2):175–81.
- Andrews BT, Pitts LH. Functional recovery after traumatic transtentorial herniation. *Neurosurgery*. 1991;29(2):227–31.
- Sakas DE, Bullock MR, Teasdale GM. One-year outcome following craniotomy for traumatic hematoma in patients with fixed dilated pupils. *J Neurosurg*. 1995;82(6):961–5.
- Schallenberg M, Bangre V, Steuhl KP, Kremmer S, Selbach JM. Comparison of the Colvard, Procyon, and Neuroptics pupillometers for measuring pupil diameter under low ambient illumination. *J Refract Surg*. 2010;26(2):134–43.
- Zafar SF, Suarez JJ. Automated pupillometer for monitoring the critically ill patient: a critical appraisal. *J Crit Care*. 2014;29(4):599–603.
- Meeker M, Du R, Bacchetti P, et al. Pupil examination: validity and clinical utility of an automated pupillometer. *J Neurosci Nurs*. 2005;37(1):34–40.
- Fountas KN, Kapsalaki EZ, Machinis TG, Boev AN, Robinson JS, Troup EC. Clinical implications of quantitative infrared pupillometry in neurosurgical patients. *Neurocrit Care*. 2006;5(1):55–60.
- Munoz Negrete FJ, Rebolledo G. Automated evaluation of the pupil. *Arch Soc Esp Ophthalmol*. 2013;88(4):125–6.
- Kramer CL, Rabinstein AA, Wijdicks EF, Hocker SE. Neurologist versus machine: is the pupillometer better than the naked eye in detecting pupillary reactivity. *Neurocrit Care*. 2014;21(2):309–11.
- Witting MD, Goyal D. Interrater reliability in pupillary measurement. *Ann Emerg Med*. 2003;41(6):832–7.
- Wilson SF, Amling JK, Floyd SD, McNair ND. Determining interrater reliability of nurses' assessments of pupillary size and reaction. *J Neurosci Nurs*. 1988;20(3):189–92.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–74.
- Gwet KL. Inter-rater reliability using SAS : a practical guide for nominal, ordinal, and interval data. Gaithersburg: Advanced Analytics, LLC; 2010.
- Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med*. 2005;37(5):360–3.
- Chen Y, Kardon RH. Studying the effect of iris mechanics on the pupillary light reflex using brimonidine-induced anisocoria. *Invest Ophthalmol Vis Sci*. 2013;54(4):2951–8.
- McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276–82.
- Chen JW, Gombart ZJ, Rogers S, Gardiner SK, Cecil S, Bullock RM. Pupillary reactivity as an early indicator of increased intracranial pressure: the introduction of the Neurological Pupil index. *Surg Neurol Int*. 2011;2:82.
- Song Z, Zheng W, Zhu H, et al. Prediction of coma and anisocoria based on computerized tomography findings in patients with supratentorial intracerebral hemorrhage. *Clin Neurol Neurosurg*. 2012;114(6):634–8.
- Du R, Meeker M, Bacchetti P, Larson MD, Holland MC, Manley GT. Evaluation of the portable infrared pupillometer. *Neurosurgery*. 2005;57(1):198–203 (**discussion 198–203**).