Impact of an Automated Pupillometer on the Reliability of the Pupillary Assessment: A Systematic Review

by

Jessica L. Keough, BSN, RN, CCRN

A Major Paper Submitted in Partial Fulfillment

of the Requirements for the Degree of

Master of Science in Nursing

in

The School of Nursing

Rhode Island College

2018

Abstract

Advancements in medical technology offer health care providers the tools needed to deliver high quality health care and the means to generate improvements in the quality of that care. The automated pupillometer device is an advanced assessment device that may significantly improve the accuracy and reliability of pupillary assessments. For patients with life-threatening neurologic complications, more reliable pupillary assessments may lead to lower rates of morbidity and mortality. However, more research is needed to determine the effect of pupillometers on health outcomes. The purpose of this systematic review was to examine the effect the pupillometer has on the accuracy and reliability of the pupillary assessment in comparison to traditional, manual assessments. Polit and Beck's guidelines for developing a research question and conducting a literature review were followed. Additionally, Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines were observed and all required elements were incorporated in the review. The strength and validity of each study was evaluated following the Critical Apraisal Skills Programme (CASP) Cohort Study Checklist, which allowed the reviewer to employ a standardized, consistent, and reliable appraisal method. Using data collected in tables formulated by the author of this review, a cross study analysis was completed whereby the studies were analyzed for emerging trends, patterns, and themes. The pupillometer was found to provide a more accurate and reliable measure of both pupil size and reactivity as compared to the traditional assessment tools such as using flash lights and the naked eye to estimate pupil size and reactivity. By incorporating the use of the pupillometer device in routine monitoring of patients at risk for neurologic deterioration, the danger of undetected life-threatening changes in condition may be reduced.

Table of Contents

Impact of an Automated Pupillometer on the Reliability of the Pupillary Assessments

A Systematic Review

Background/Statement of the Problem

Management of patients with acute neurologic injuries involves with the neurological assessment, and assessment of the pupillary light reflex (PLR) is one of its' key components. The PLR represents the function of both the optic or second cranial nerve (CN II) and oculomotor or third cranial nerve (CN III) and provides vital information about brain function (Wilson-Pauwels & Akesson, 2001). The conventional pupil assessment is performed using a penlight, flashlight, or other light source and the naked eye. Although this technique involves multiple steps, it only takes a few seconds to complete. First, the examiner measures the diameter of a resting pupil in ambient light, then shines a light at the same pupil. Next, the examiner grades the intensity of the pupillary constriction, subjectively rating it brisk, sluggish, or non-reactive, in response to the light stimulus and measures the size of the constricted pupil. This process is then repeated on the alternate side.

Under normal conditions, the pupils are equal in size, are round in shape, and react briskly to light (Hickey, 2009; Singhal & Josephson, 2014). Anisocoria refers to any difference in size between pupils; a difference in size of 1 mm or greater is considered an abnormal finding (Hickey; Singhal & Josephson). Changes in pupil size or reactivity may be due to several causes. However, when the change is unilateral, especially when manifesting as dilated pupils, there is cause for concern that there may be external compression of CN III from impending herniation and increased intracranial pressure (ICP) (Meeker et al., 2005; Singhal & Josephson, 2014). Alternatively, such findings may

also be found in cases of severely diminished cerebral perfusion (Portran, Cour, Hernu, de la Salle, & Argaud, 2017; Ritter et al., 1999). When present in a patient with an acute brain injury or altered level of consciousness, such pupillary changes necessitate further diagnostic workup and potentially life-saving interventions (Meeker et al., 2005). If left undetected and untreated, the neurologic damage that occurs may become irreversible and can rapidly progress to coma and death (Hoffmann et al., 2012). Accurate and reliable pupil assessments are critical in detecting life-threatening conditions for which pupillary changes may represent an early sign.

However, the traditional pupil assessment carries with it a significant degree of assessor subjectivity, which ucan lead to inconsistencies in findings between examiners (Olson & Fishel, 2016). Several factors may contribute to inconsistencies between examiners, including clinical experience and skill level, variations in assessment techniques (e.g. validation with an adjunctive pupil gauge), differences in a given light source's illumination intensity, and differences in ambient light conditions (Olson & Fishel). This carries practical consequences, as multiple healthcare providers typically collaborate in caring for patients with complex neurological injuries. Any variability in technique or other inconsistency between examiners may lead to delayed detection, and therefore delayed treatment, of a deterioration in neurologic condition.

Given these limitations of the traditional pupil assessment, an emerging alternative includes the utilization of an automated pupillometer device. The pupillometer is a handheld device with a built-in infrared light and camera, whose sole function is to provide a pupil assessment with minimal inter-rater variability. Raw data captured by the pupillometer includes measurements of pupillary size, response latency, constriction

velocity, and dilation velocity (Chen et al., 2011). Each of these pupillometer measurements is then compared to previously-validated normal reference ranges and the resultant data is converted to a proprietary value called the neurological pupil index (NPi). The NPi was developed to provide easy interpretation of results and a high degree of objectivity. Neurological pupil index scores fall between 0 and 5, with a score of 3 or greater representing normal pupil activity, while scores less than 3 suggest a sluggish pupillary response (Chen et al.).

The foundation of the care and management of a patient with an acute neurologic injury continues to reside with the neurologic assessment. However, with advancements guided by medical technology, secondary deterioration in neurologic condition after acute brain injury may potentially be avoidable (Meeker et al., 2005). Pupillometry may be used to more accurately assess the PLR in patients at greatest risk for life-threatening neurologic complications.

The purpose of this project is to conduct a systematic review to comprehensively examine the impact of an automated pupillometer device on the accuracy and reliability of the pupillary assessment when compared to the traditional application of a penlight and the naked eye assessment.

Next, a review of the literature will be presented.

Literature Review

To construct a comprehensive review of the literature searches through both PubMed and CIHAHL databases were performed between September 2017 and January 2019. Key words used to direct the search include ‘neurologic exam,’ ‘pupil exam,’ ‘pupillary response,’ and ‘pupillometry.’ The articles included in the review are described in the section below.

Neurologic Assessment

The neurologic assessment is comprised of multiple components aimed at evaluating a host of neurologic domains, including a patient’s level of consciousness, cognitive ability, and cranial nerves, in addition to motor and sensory function (Olson & Fishel, 2016). Tools have been developed to facilitate the gathering, organization, and communication of assessment data. For instance, the Glasgow Coma Scale (GCS) is a tool widely used by doctors and nurses around the world to grade the severity of acute brain injury using alterations in level of consciousness, motor, and verbal ability (Hickey, 2009). In the context of global cerebral injuries and herniation syndromes, the cranial nerve assessment allows localization of lesions involving the brainstem (Olson & Fishel, 2016). In all, there are 12 pairs of cranial nerves. The PLR allows an examiner to test the function of both CN II and CN III (Olson & Fishel).

Insight into a patient’s neurologic condition is gained at the time the assessment is performed. To place newly obtained assessment findings into a greater context, the examiner must compare these findings to those previously obtained. Hence, the neurologic assessment is performed in a serial fashion, for example, hourly, and the

results are trended over time. In this way, the assessor may infer an illness' trajectory, devise an appropriate treatment plan, and evaluate the effectiveness of prior treatments.

Anatomy, Physiology and Factors that Affect the Pupillary Light Reflex

When light enters the pupil, rods and cones in the retina signal CN II to convert the stimulus to an electrical impulse. This impulse travels along the afferent tract of CN II toward an area in the midbrain called the Edinger-Westphal nucleus (Wilson-Pauwels & Akesson, 2001). Once there, the signal is taken up by the efferent pathway of CN III. It continues to travel toward motor fibers in the eye which then generate pupillary constriction (Olson & Fishel).

Anatomically, fibers from CN III pass anteriorly from the CN III nucleus and emerge from the ventral surface of the midbrain. The medial temporal lobe lies directly adjacent to CN III at this point and represents a potential source of external compression (Figure 1.) (Haines, 2008).

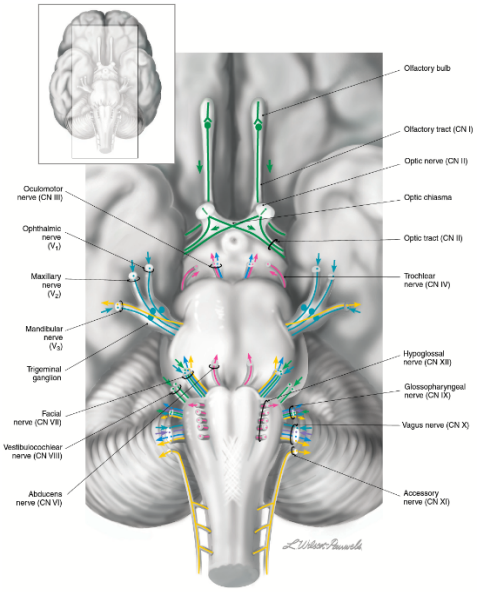

Figure 1. Basal view of the brain and brainstem (Wilson-Pauwels & Akesson, 2001)

The iris dilator and sphincter muscles are controlled by the sympathetic and parasympathetic nervous systems, respectively, and work antagonistically to one another to regulate pupil size (Wilson-Pauwels & Akesson, 2001). Muscular dysfunction may lead to changes in the PLR.

The PLR may be altered due to a number of causes, necessitating that providers place an abnormal PLR within the context of the other components of the neurological exam, as well as other clinical data. External compression of CN III may be due to a focal space-occupying lesion (e.g. tumor, aneurysm, abscess, or hemorrhage) or a more global elevation in ICP causing uncal herniation (Portran et al., 2017; Ritter et al., 1999). A direct injury such as ischemia or infarction involving CN III or its nucleus in the midbrain may also manifest as PLR changes (Portran et al.; Ritter et al.). Meanwhile, though toxic or metabolic processes may carry with them widespread systemic manifestations, they may also depress the nervous system's metabolism sufficiently to also cause impairment of the PLR. This category may also include iatrogenic causes such as anesthetics, opiates, and neuromuscular blocking agents (Posner, Saper, Schiff, & Plum, 2008). Similar findings may also be seen with cardiac arrest and other shock-like states leading to a loss of cerebral perfusion (Posner et al., 2008).

Though less commonly encountered in the acute care setting, abnormal PLR findings may also be due to defects in afferent conduction, and such conditions should also be considered in respect to the differential diagnosis depending on clinical context. These include optic neuropathy (e.g. ischemic or demyelinating), cataracts, retinal abnormalities, or injuries to the globe itself (Safa, 2010; Yoo, 2017). The PLR may also

be misleading in patients who have received various forms of eye surgery, such as iridectomy or cataract removal (Safa, 2010).

Additional Prognostic Value of Pupil Assessment

Although the PLR provides a valuable piece of the neurologic assessment, it has also been shown to carry some merit in terms of outcome prognostication. In a single-center retrospective study of 272 patients admitted following out-of-hospital cardiac arrest between January 1, 2005 and March 31, 2009. Rittenberger et al. (2010) found that among all components of the neurological assessment, absent pupillary response perfectly predicted poor outcome. Data collected in this study included GCS, motor examination, in addition to pupil and corneal responses. Each of these assessment findings were available upon admission, after 24 hours, and at 72 hours. Data were only collected during "sedation vacation," periods to control for possible changes in pupillary responsiveness due to the use of sedatives or paralyzing agents.

Of patients lacking a pupillary response on admission, 7/65 or 11% survived (95% CI 2.4-19%) while those without a pupillary response at 72 hours, 0/17 (95% CI 0-2.9%) survived. Although survival rates were poor, GCS motor exam score less than 3 (characterized by abnormal flexor posturing) at 24 hours (17% survival) or 72 hours (20% survival), was universally predictive of mortality. Based on their findings, Rittenberger et al. (2010) suggested that the lack of a pupil response at 72 hours is indicative of poor outcomes; however, they also noted the limitation that their study was retrospective, and that there may be a component of self-fulfilling prophecy.

Hoffman et al. (2012) performed a retrospective cohort analysis of data from 51,425 patients with severe TBI registered in the Trauma Registry of the German Society for Trauma Surgery from 1993 to 2009. Data were analyzed in an effort to determine the utility of the pupil assessment as a predictor of mortality. Other variables included the components of the GCS such as eye opening, verbal and motor response, which they used in addition to pupil size and reactivity to create receiver operating characteristic (ROC) models. Hoffman et al. used area under the ROC curve analysis (AUROC) based on multivariate logistic regression to determine prognostic accuracy of their models, with an AUROC of 1.0 indicating perfect separation of survivors from non-survivors.

The model with the highest accuracy was the "full model," where GCS, pupil reactivity, and pupil size were considered (AUROC 0.830, 95% CI, 0.822-0.838). However, their findings also suggested that pupil reactivity alone was highly sensitive for mortality after severe TBI (AUROC 0.770, 95% CI, 0.761-0.779). The group's findings validate the prediction that the presence of sluggish or non-reactive pupils is a poor prognostic sign for patients with severe TBI.

Pupillometry: Strengths and Practical Advantage

The pupillometer device was originally introduced in the 1960's. However, it did not gain popularity in the acute care setting until the 2000's (Larson & Singh, 2016). Chen et al. (2011) were the first to introduce the NPi as a measure of pupil reactivity.

Measurements obtained via pupillometry are consistently achieved by several key strengths in design: the detachable headrest, which optimizes and maintains consistency of device placement and the fixed intensity and duration of its flash of light (Figure 2.)

(Chen et al., 2011). The whole measurement is taken over approximately 3.2 seconds, during which the camera records over 30 frames per second. The visual data is automatically converted to numeric values representing the size, latency, constriction velocity, and dilation velocity, all of which ultimately contribute to the final NPi (Figure 3). This precise measurement can then be used by providers to reliably trend the PLR between examiners and over time.

Figure 2. NPi®-200 Pupillometer System ("NeurOptics, Inc," 2019)

Figure 3. NPi®-200 Pupillometer – Results Screen 2 ("NeurOptics, Inc," 2019)

In their study evaluating pupillometry and its' associations with ICP, Chen et al. (2011) used data from 134 patients admitted to eight different neurological and critical care units. All patients had at least one reactive pupil and evidence of cerebral edema or substantial mass lesion on computed tomography (CT) scan imaging obtained on admission. Pupillary exams were performed using a pupillometer every 30 minutes for 72 hours and patients were separated into three groups based on NPi values. Patients with

normal pupil reactivity indicated by NPi values > 3 (n=98) had a mean ICP 19.6 mmHg; those with abnormal pupil reactivity or at least one occurrence of NPi ≤ 3 (n=28) had a mean ICP 30.5 mmHg (p=0.0014); and those with non-reactive pupils (n=8) had a mean ICP 33.8 mmHg (p=0.0046). This trend suggests an inverse correlation between NPi and ICP. Additionally, further analysis demonstrated abnormal NPi results preceded spikes in ICP by an average of 15.9 hours prior to the time of maximum ICP, with a range varying widely, from 0 to 60 hours.

Another prospective cohort study by McNett et al. (2017) was built on previous studies by concretely investigating the relationship between pupillometry and ICP. Patients included in this study were admitted to a neurological intensive care unit where hourly pupillometer and ICP values were recorded for 72 hours after admission. The study included 76 subjects and more than 2100 pupillometer results with corresponding ICP values. Pupillometer values consisted of pupil size, constriction velocity, and the NPi. Statistical analyses included Pearson correlation between pupillometer and ICP values.

Data suggested that as pupil size increased there was an accompanying rise in ICP (right eye, $r = 0.166$, $p<0.001$; left eye, $r= 0.133$, $p=0.001$) and second, as pupil reactivity decreased, evidenced by a decrease in NPi, there was an opposing elevation in ICP [right eye, $r = -0.126$, $p=0.001$; left eye, $r=-0.225$, $p<0.001$] (McNett et al., 2017). This study was limited by its lack of generalizability due to its small, single-site design. However, it lays the groundwork for future, larger prospective studies to determine guidelines for monitoring parameters and the use of pupillometry in patients with severe brain injury.

Both studies (Chen et al., 2011; McNett et al., 2017) reported an inverse relationship between pupil reactivity and ICP.

Next, the theoretical framework guiding this project will be presented.

Theoretical Framework

Evidence-based practice (EBP) propels quality of care and patient safety. Scientific reporting via systematic review or meta-analysis promotes the development and expansion of EBP. Before research findings can be translated to clinical practice, data must be derived from the literature, critically appraised, and succinctly reported. As a result, the reviewer must have considerable knowledge of the topic to be studied and the ability to sufficiently appraise the evidence (Polit & Beck, 2017). To ensure findings are translated to the highest quality EBP, it is important for clinicians to review all literature and appraise the strength and validity of each study. Once a thorough critique is completed, recommendations to drive clinical practice changes and foster improvements in patient care may be developed.

During this project, Polit and Beck's (2017) guidelines for developing a research question and performing a literature review in addition to the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) statement were followed. According to the Polit and Beck framework (2017), once a clinical problem is identified it should be followed by the development of a research question that contains the elements population, intervention, comparison, and outcome. The question must be focused and specific, yet broad enough to elicit an ample amount of literature in the search. Key terms are identified based on the subject matter and used to search electronic databases for relevant literature. In the preliminary search, articles are gathered and set aside based on the title and content in the abstract (Polit & Beck). Once this phase is complete, the search must be narrowed by carefully screening articles based on inclusion and exclusion criteria.

In 1996, the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement was developed to standardize and improve the quality of scientific reporting (Liberati et al., 2009). The PRISMA framework contains two key components to guide the systematic reviewer to a satisfactory completion. First, a 27-item checklist (Figure 4.) outlines specific elements such as title, abstract, introduction, methods, results, discussion, and funding that must be included (Liberati et al.). The checklist provides the reviewer with a reference to ensure each element is included in the final report.

Similar to the process described by Polit and Beck (2017), the PRISMA framework requires the systematic reviewer to develop a research question using the Population, Intervention, Comparison, and Outcomes (PICO) approach (Liberati et al., 2009). Key terms are identified, a search of the literature is performed, and articles are selected based on inclusion and exclusion criteria. The second component of the PRISMA framework is a four-phase flow diagram (Figure 5.) that visually depicts the process of article selection (Liberati et al.).

Next, the methods guiding this study will be presented.

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	

Section/topic	#		Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	

Figure 4. PRISMA Checklist (Moher, Liberati, Tetzlaff, & Altman, 2009).

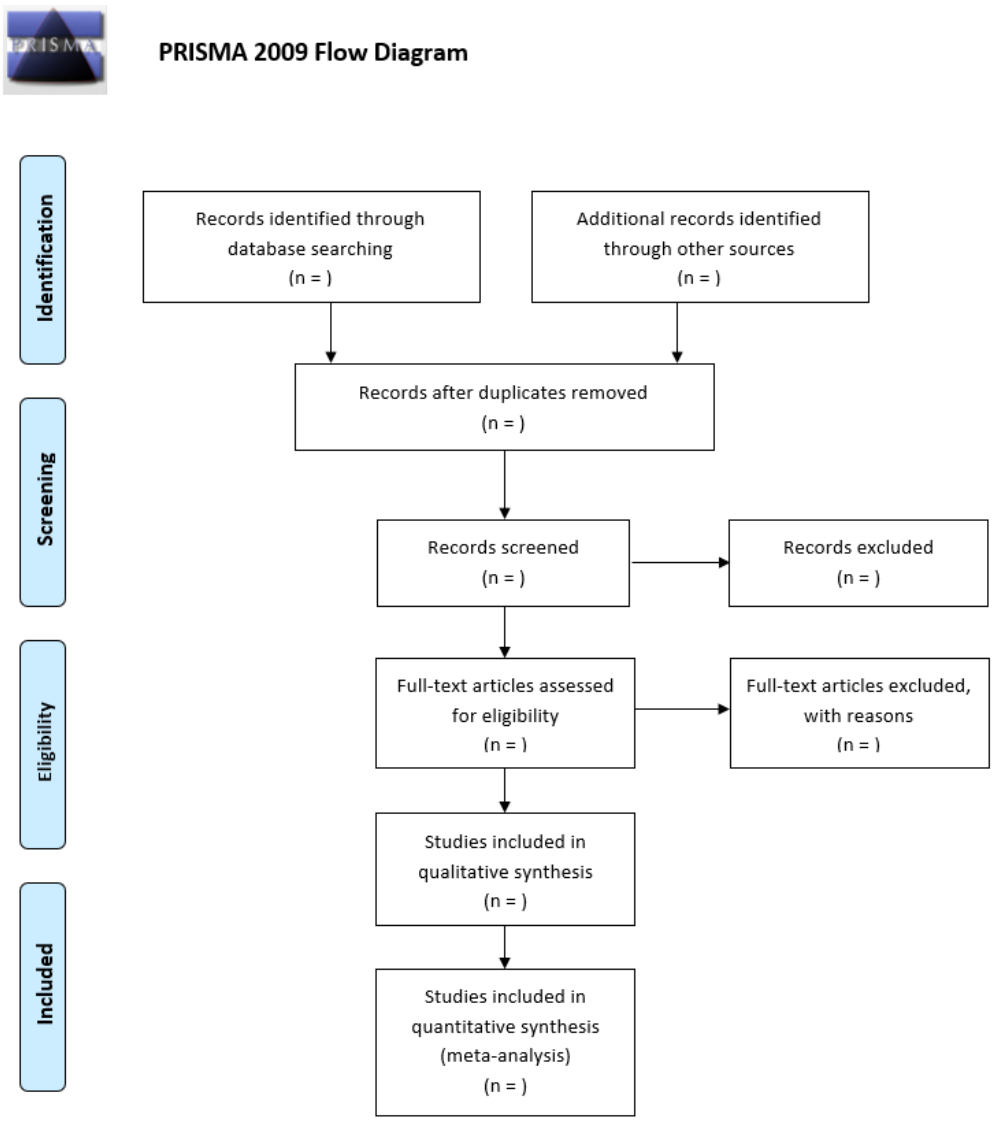

Figure 5. PRISMA flow diagram (Moher et al., 2009)

Method

Purpose and Design

The purpose of this study was to conduct a systematic review which examined the impact of an automated pupillometer device on the accuracy and reliability of the pupillary assessment when compared to the traditional application of a penlight and the naked eye assessment.

Inclusion and Exclusion Criteria

Inclusion criteria consisted of studies (1) on humans; (2) on adults over the age of 18; (3) with quasi-experimental or randomized control designs; (4) that sought to evaluate the reliability of the traditional pupillary assessment; and (5) that compared the traditional pupil assessment with that obtained by a pupillometer.

Exclusion criteria consisted of studies not performed on human subjects, those not reported in the English language, and the pediatric population.

Search Strategy

A literature search was performed with PubMed and CINAHL databases using the key words ‘pupillometer,’ ‘pupil assessment,’ ‘pupil response,’ ‘pupil,’ ‘interrater reliability.’

An initial search on the search engine PubMed using the key term ‘pupillometer,’ yielded 337 search results and another 41 results were found on the CINAHL database. The search was narrowed using Boolean operators and MESH terminology. Five articles met inclusion and exclusion criteria and were chosen to be included in the systematic review. The PRISMA flow chart in Figure 6 displays this process.

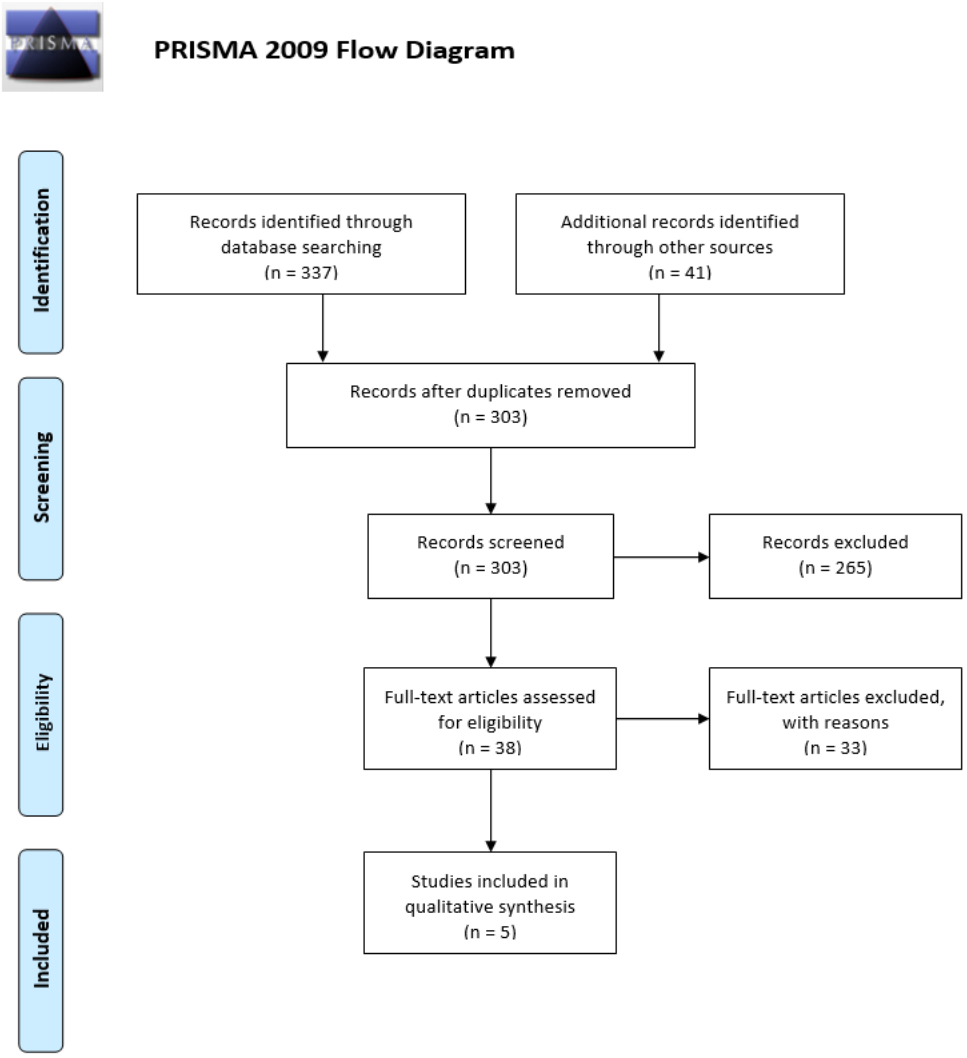

Figure 6. PRISMA flow diagram (Moher et al., 2009).

Data Collection

Pertinent information from each article was organized in two tables. The first table lists the aim of the study, sample/setting, and design method (Table 1). The second table lists the variables measured, data analysis, study findings and limitations (Table 2).

Table 1. Aim, Sample/Setting, Design/Methods.

Aim	Sample/Setting	Design/Methods

Table 2. Variables measured, Data analysis, Study findings, Limitations.

Variables measured	Data analysis	Study findings	Limitations

Critical Appraisal

Critical Appraisal Skills Programme (CASP) developed a series of checklists to guide readers of research through the appraisal process and thereby foster a greater understanding of scientific reporting ("CASP Cohort Study Checklist," 2018). Their checklists highlight key concepts such as validity, results, and clinical relevance (2018). The studies included in this review were critically appraised using the CASP Cohort Study Checklist, which allowed the reviewer to employ a standardized, consistent, and reliable method to evaluate the strength and validity of each study's findings. The findings from the CASP inquiry are detailed in Appendix C.

Data Synthesis and Cross Study Analysis

Facilitated by data collected in the tables, the studies were analyzed for emerging trends, patterns, and themes. A cross study analysis was performed that evaluated the reliability and variability of the manual pupil assessment compared to the pupillometer assessment. The findings from the cross study analysis are found in Appendix D. Next, the results will be presented.

Results

Articles included in this systematic review were selected based on a process of elimination utilizing the PRISMA flow diagram (Figure 3). A total of 378 articles were initially screened out of which 5 were chosen to be reviewed. Below is a description of each study.

In a prospective observational single-blind study, Meeker et al. (2005) (Appendix A-1) used serial pupillometer assessments to examine the accuracy and inter-rater reliability of the manual pupil assessment. A combined 452 manual and pupillometer assessments were performed on 20 patients in an intensive care unit. Successive manual assessments were completed by two groups of examiners; each group contained one neurosurgical attending physician, two neurosurgical interns, and four advanced practice nurses. Each examiner recorded bilateral pupil size in millimeters and categorized pupil reactivity as non-reactive, sluggish, or brisk. Before and after each manual assessment, pupillometer measurements were obtained by a trained member of the investigative team. To minimize bias, the examiners were blinded to each other's results and to those of the pupillometer. Interpolated pupillometer measurements for pupil size and reactivity provided the reference data for which each manual assessment was compared. Dimmed ambient light conditions were consistent for all assessments but pupil gauges were not used by the examiner groups.

Meeker et al. determined that neither examiner group was able to reliably estimate pupil reactivity, with inter-examiner disagreement occurring 39% of the time (95% CI 28-52%). Examiners missed non-reactive pupils in three patients whose mean pupil size was 3.2 mm (SD=1.2mm). They also inaccurately labeled briskly reactive pupils as non-

reactive in 27 patient assessments. Between examiners, the median standard deviation for manual pupil size measurements was 0.58 mm (95% CI 0.50-0.58 mm). The pupillometer group generated smaller rates of absolute error and less variability than the manual assessment. The median absolute error was 0.23 (95% CI 0.20-0.31 mm) versus 0.5 mm (95% CI 0.47-0.60) for the manual assessment (Appendix B-1). Meeker et al. concluded that the pupillometer provided nurses with an accurate and reliable measure of both pupil size and reactivity.

Meeker et al. recruited their cohort from a random sampling of patients admitted to the ICU (Appendix C-1). Of the 20 patients included in their study, 10 were admitted with acute neurologic diagnoses including hemorrhagic stroke and traumatic brain injury while the other 10 were admitted with medical problems such as pancreatitis and pneumonia. Seventeen patients required continuous intravenous sedation while three patients had Glasgow Coma Scale scores between 3-5 and did not receive any sedatives. It is unclear if the delivery of sedating medications skewed the results in any way; this factor was not discussed. Although the study results are more generalizable to critical care medicine and less specific to the neurological patient population, outcomes were clearly measured with minimal bias and the analysis was thorough and valid. The concept that the manual pupillary assessment is unreliable was clearly addressed by the authors.

In a single-blinded observational study, Olsen et al. (2016) (Appendix A-2) examined the interrater reliability of two methods of pupil assessment: a comparison of manual pupil exams performed by two practitioners, and a comparison between a practitioner performing conventional manual assessments and assessments using an automated pupillometer device. Patients with a neurological or neurosurgical diagnosis

and pre-existing orders for serial pupil examination were eligible for inclusion in the study. Practitioners consisted of registered nurses (RN), nurse practitioners, neurologists, neurosurgeons, and resident physicians, all of whom who routinely perform pupil assessments as part of their usual practice. For each study participant, a convenience pairing designated two practitioners who performed the manual assessment, while a trained member of the investigative team performed the pupillometer assessment. Size and reactivity measures derived by the pupillometer were considered the reference results and provided the basis for which assessments obtained by the practitioner groups were compared. Practitioners used a light source of their choosing to evaluate the size, shape, and reactivity of patients' left and right pupils. Both practitioners performed their assessments independently and were blinded to the other's findings, while all members of the team were blinded to the pupillometer results. To minimize variation in testing conditions, investigators ensured all pupil exams were performed under ambient light and that patients were assessed in identical physiologic states, with no greater than five minutes between assessments.

A total of 2329 paired manual pupil assessments and 2192 pupillometer assessments were obtained from a total of 127 patients. Two hundred and twenty-two practitioners participated in the study, including 194 RNs and 28 physicians, while three trained research assistants performed the pupillometer assessments. A Cohen's kappa coefficient (k) was calculated to measure interrater agreement for pupil size, shape and reactivity, both as itemized component scores and as a composite score aggregating all three components. Lower kappa results indicated lower interrater agreement.

The investigators found little agreement between practitioners' manual pupil size, shape and reactivity scores, which supported their hypothesis that the traditional pupil assessment may be an unreliable assessment technique. This was especially the case for patients with abnormal pupillary assessment findings. For pupil size, practitioner agreement was fair (k = 0.54; 95% CI 0.50-0.57) and agreement with the pupillometer was low (k = 0.29; 95% CI 0.27-0.32). Practitioner agreement on pupil reactivity was also fair (k = 0.40; 95% CI 0.36-0.44); meanwhile, when compared to the pupillometer results, practitioners agreed 95.7% (2135/2230) of the time when pupils reacted normally. However, when the device detected a fixed pupil, practitioner agreement decreased to 49.7% (94/189). The pupillometer device scored 83 pupils as non-reactive, with only 58/83 (69.9%) and 46/83 (55.4%) of the paired practitioner observations correctly identified—a potentially life-threatening change in pupillary responsiveness that would otherwise have gone unrecognized (Appendix B-2). As a result, Olsen et al. concluded in their study that measures should be taken to standardize and improve the reliability of the pupil assessment, and that the pupillometer may do this.

Although the manual assessors were not blinded to the aims of the study, they were blinded to each other and the pupillometer assessment results. The design sought to minimize bias and the outcomes were measured thoroughly. The results are both valid and relevant to the clinical problem (Appendix C-2).

In France, Couret et al. (2016) (Appendix A-3) conducted a two-phase study to compare the reliability of pupillometer assessments with that of traditional manual pupil assessments. In the first phase, they sought to determine inter-observer agreement using the pupillometer device. To achieve this, junior and senior practitioners (residents and

neurointensivists, respectively) performed repeated measures on 200 healthy volunteers, yielding a total of 400 paired measurements for both right and left eyes in a variety of ambient light conditions. They determined that there was a high level of agreement among providers of varying experience levels. The intra-class correlation coefficient for maximum resting pupil size and minimum pupil size after light stimulation was 0.95 (95% confidence interval [CI]: 0.93-0.97) and 0.87 (95% CI: 0.83-0.89), respectively (Appendix B-3).

In the second phase of their study, Couret et al. used a prospective, observational, double-blind study design to compare the traditional pupil assessment as performed by a nurse with results from a pupillometer used by a trained physician (Appendix A-3). The study was performed in two neurocritical care units between January and December 2012. During that time, 406 pupillary measurements were obtained on 59 patients who were included in the study cohort. Nurses who participated in the study had an average of 10 years of experience performing pupil assessments in neurologically impaired patients. The nurses were asked to estimate the pupil size, identify anisocoria, and manually assess pupillary light reflex using a penlight. As in standard practice, pathologic anisocoria was defined by the researchers as a difference in pupil size greater than or equal to 1mm. The research team controlled for ambient lighting conditions by ensuring all environmental lights were dimmed prior to the exam. The team also controlled for the effects of any unintentional consensual light reflex by ensuring nurses closed the patient's opposite eye during the assessment.

For each patient, four measurements were obtained every 24 hours with a rest period of approximately five minutes between the nurse and physician measurements.

Spearman's rank correlation coefficients were calculated for three test groups: pupils <2mm (n=61); pupils 2-4mm (n=232); and pupils > 4mm (n=113). The results suggested overall low agreement between the device and manually obtained measurements, with reported Spearman's rho values ranging from 0.39 (95% CI: 0.15-0.59; p = 0.002) for pupils <2mm in size to 0.44 (95% CI: 0.33-0.54; p < 0.001) for pupils that were 2-4mm, and 0.37 (95% CI: 0.19-0.51; p = 0.001) for pupils >4mm. In the 2-4mm group, which was most frequent, nurses disagreed with the pupillometer in 19% of cases (43/223), including 41 cases in which nurses missed non-reactive pupils. Additionally, anisocoria was detected by nurses only 50% (15/30) of the time (Appendix B-3). Couret et al. therefore concluded that the use of a pupillometer improves the reliability of the pupil assessment, and as a result encouraged the use of pupillometers in the routine care of brain-injured patients.

Their study design sought to minimize bias and their outcomes were measured thoroughly. The results are both valid and relevant to the clinical problem. No major limitations were identified in the appraisal process (Appendix C-3).

In a three-phase prospective observational cohort study, Kerr et al. (2016) (Appendix A-4) sought to evaluate neuroscience nurses' abilities to accurately measure pupil size and detect anisocoria. In the first phase, a group of 30 critical care and neurosurgical nurses graded pupil size from a simple black-and-white drawing. In the second phase that followed two weeks later, 27 nurses from the first phase graded pupil size from a color photograph of a human face. In the third phase, a total of 489 pupillary assessments were conducted on 93 patients admitted to either the intensive care unit or a neurosurgical inpatient floor. In this part of the study, the results of bedside nurses'

manual pupillary assessments were compared to those obtained by a trained research coordinator using a pupillometer.

As a result of their study, Kerr et al. discovered that nurses consistently underestimated pupil size, were unable to reliably detect anisocoria, and inaccurately measured pupil reactivity. In both phases that involved pictorial assessments, there was a commensurate decrease in accuracy of nurses' pupil size estimates as actual pupil diameters increased, with poor accuracy for the largest pupil sizes. Results from the first phase revealed 54% measurement accuracy for pupils with diameters >5mm, while in the second phase nurse estimates of pupil diameter were accurate only 37% of the time for pupils >4.5mm. Additionally, when nurses in the second phase were shown the same color photograph twice, they consistently and accurately measured the duplicate pupil diameter only 11.7% of the time. Meanwhile, in the third phase's survey of real-world clinical pupillary assessments, 82.4% of nurses were able to accurately assess pupil reactivity despite one-third of examiners not dimming the room lights prior to their assessment. However, nurses failed to detect sluggish readings 21% of the time, inaccurately reported normally reactive pupils as sluggish 17% of the time, and correctly identified anisocoria just 58.1% of the time (Appendix B-4).

Results from each phase in this study reinforce the variability inherent in pupillary assessment, specifically pertaining to pupil size and the misidentification of anisocoria. Kerr et al. concluded by hypothesizing that improving the reliability of the pupil assessment may promote early detection of pupillary changes and potentially improve patient outcomes.

Although pupil reactivity could not be assessed in the pictorial assessments of the first two phases, Kerr et al. clearly addressed the clinical problem. The study design minimized bias and the outcomes were measured thoroughly. The results are both valid and relevant to the clinical problem (Appendix C-4).

Marshall et al. (2018) (Appendix A-5) conducted a two-part prospective observational single blind study to evaluate the feasibility and variability of automated pupillometry for use in the care of stroke patients. In the first part, a sample of 12 acute stroke patients and their nurses evaluated three qualities of the automated pupillometer: feasibility, acceptability, and safety for use. The nurses were asked to obtain pupillometer measurements along with their usual neurologic assessment; feasibility was determined by a compliance rate of 80% or greater. Between shifts, feasibility varied widely. A total of 92.7% of day shift (9am – 7pm) assessments were completed compared to only 30.8% of those from the night shift (9pm – 7am). Following the assessment period, a survey was distributed to both participating nurses and their patients to evaluate how well each accepted the pupillometer for use in practice. Both groups were asked to rate the device's comfort and ease of use using a 5-point Likert scale. Only scores between 1 and 3 were considered acceptable. Both groups rated the device favorably with an average rating of 2.4 from nurses and 1.4 from patients. During the study period there were no reported adverse events related to the device and it was considered safe to use (Appendix B-5).

In the second part of the study, Marshall et al. (2018) sought to compare the interrater reliability and variability of the manual pupillary assessment to that of the pupillometer (Appendix A-5). A total of 132 paired measurements of individual pupils were obtained from 52 participants; 42 were performed on stroke patients and the

remaining 90 were performed on healthy staff volunteers. The assessments were completed within 15 minutes of each other. Agreement for size was defined as a difference of less than 1 mm between the observers' measurements and anisocoria was defined as a difference greater than or equal to 1 mm between pupils.

For the measurement of pupil size, interrater agreement for the pupillometer was 99.2% with a Spearman correlation coefficient of 0.949 (95% CI, 0.929-0.969). In contrast, manual examiners agreed on pupil size just 61.4% of the time with a Spearman correlation coefficient of 0.633 (95% CI, 0.531-0.735). The majority of manual observers (84.1%) graded pupil size as 3 or 4 mm while the corresponding pupillometer values varied between 1.9 and 6.1 mm. Within the cohort, there was a total of 14 identified cases of anisocoria for which interrater agreement with the pupillometer was substantial (98.5%, k = 0.660; 95% CI 0.039-1.00) and agreement between manual observers was fair (89.4%, k = 0.306; 95% CI -0.078-0.690). Agreement between manual observers and the pupillometer was poor (87.9%, k = -0.027; 95% CI -0.074-0.020) (Appendix B-5).

In the study, both the Neurological Pupil Index (NPi) and constriction velocity (CV) values were used to interpret pupil responsiveness. An NPi value less than 3.0 and CV less than 0.8 m/s represented a sluggish response. Of the manual assessments graded as sluggish, none of them were associated with an NPi value less than 3.0. Additionally, for every pupillometer assessment with an NPi value less than 3.0, a brisk reaction was reported by the manual observers (k = -0.026; 95% CI -0.042 to -0.010). Of the pupils with a sluggish CV (n=10), only two were graded as sluggish by manual observers (20.0%, k = 0.006; 95% CI, -0.004–0.016) (Appendix B-5).

Marshall et al. (2018) concluded that the integration of pupillometry in routine neurologic monitoring of stroke patients may improve the detection of early neurological deterioration and thereby hasten the delivery of time sensitive, life-saving treatments. They also suggested the discrepancies in compliance rates between day and night shift may be remedied by improved staff education however, did not discuss potential limitations due to staffing ratios.

Although there were flaws in participant recruitment, the majority of the cohort from the second part of the study consisted of healthy volunteers. The study design sought to minimize bias and the outcomes were measured thoroughly. The results are both valid and relevant to the clinical problem (Appendix C-5).

Cross Study Analysis

As explained above, each of the five studies investigated the accuracy and reliability of manual pupil assessments compared to pupillometer assessments. In addition, three of the five studies also sought to establish inter-rater agreement for the pupillometer devices (Couret et al., 2016; Kerr et al., 2016; Marshall et al., 2018). The table found in Appendix D was created to organize data from each study which was then used to conduct the cross study analysis.

Between the studies, there were notable similarities for the detection of anisocoria as well as sluggish and non-reactive pupils. Four of the five studies reported that examiners frequently missed anisocoria. Couret et al. (2016), reported that nurses accurately detected anisocoria 50% of the time while Kerr et al. (2016), found a similar rate of 58.1% by nursing staff. Likewise, Olson et al. (2016), reported moderate

agreement (k = 0.60, 0.54 to 0.64) among providers. Lastly, Marshall et al. (2018), found when identifying anisocoria, nurses agreed with each other only 36.4% of the time and had poor agreement with the pupillometer (87.9%, k = -0.027; 95% CI, -0.074 to 0.020) (2018) (Appendix D).

In their study, Meeker et al. (2005), discovered nurses missed non-reactive pupils in 3/20 patients but they also noted a degree of false positive reporting with a total of 27 pupil assessments that were incorrectly graded as non-reactive. In the study by Olsen et al. (2016), there was a total of 189 non-reactive pupils reported between the two manual groups . The providers agreed with each other 49.7% of the time (94/189); however, only 33.3% (58/189) were confirmed as non-reactive by pupillometry. Couret et al. (2016), reported 41/406 missed cases of non-reactive pupils (2016). Kerr et al. (2016), found nurses accurately reported sluggish pupils in 7/33 (21%) cases but inaccurately reported sluggish pupils 77/444 times (17%) of the time. Finally, Marshall et al. (2018) noted that every sluggish pupil graded by the manual assessment group was found to be brisk according to the pupillometer measurements. Conversely, every assessment corresponding to an abnormal NPi score was graded as brisk by the manual group (k = -0.026; 95% CI, -0.042 to -0.010) (Appendix D).

Across the studies, findings varied regarding accuracy of manual assessments of small versus large pupils. Kerr et al. (2016), reported that manual estimations were less accurate for pupils >4.0 mm (mean 0.6 mm, SD 1.32 mm) (2016). In contrast, Couret et al. (2016), found the rate of error was greatest for pupils between 2 and 4 mm (0.44, 95% CI, 0.33-0.54). However, Olson et al. (2016), compared agreement for pupil size (<3 mm

or >3 mm) in four subsets of practitioners, all yielded moderate results (k = 0.54, 0.53, 0.63 and 0.54) (Appendix D).

Next, the summary and conclusions will be presented.

Summary and Conclusions

The purpose of this paper was to determine the effects of automated pupillometer devices on the accuracy and reliability of the pupil assessment in comparison to the manual assessment technique. A literature review using key words ‘neurologic exam,’ ‘pupil exam,’ ‘pupillary response,’ and ‘pupillometry’ was performed and articles were selected based on chosen inclusion and exclusion criteria. The PRISMA framework was utilized to organize and guide article selection; this process is visually depicted in the four-phase flow diagram (Figure 6). Five prospective observational blinded studies were selected for the review and data from each study including aim, sample/setting, design, variables measured, data analysis, study findings, and limitations were organized in tables (Appendix A and B). The studies were then critiqued using the CASP Cohort Study Checklist. This allowed the reviewer to employ a standardized, consistent, and reliable method to evaluate the strength and validity of findings from each (Appendix C). Using data collected in the tables, the studies were analyzed for emerging trends, patterns, and themes. A cross study analysis was performed that evaluated the reliability and variability of the manual pupil assessment compared to that of the pupillometer assessment (Appendix D).

A feasible alternative to the manual pupil assessment includes the use of an automated pupillometer device. The pupillometer is a handheld device with a built-in infrared light and camera. Its’ sole function is to provide a pupil assessment with minimal inter-rater variability. Raw data captured by the pupillometer includes measurements of pupillary size, response latency, constriction velocity, and dilation velocity (Chen et al., 2011). Each of these pupillometer measurements is then compared to previously-

validated normal reference ranges and the resultant data is converted to a proprietary value called the NPi. The NPi was developed to provide easy interpretation of results and a high degree of objectivity. Neurological pupil index scores fall between 0 and 5, with a score of 3 or greater representing normal pupil activity, while scores less than 3 suggest a sluggish pupillary response (Chen et al.). Precise measurements may be trended over time and used to more accurately assess the PLR in neurologically injured patients.

Each study detailed in this review compared the reliability of manual assessments to those of pupillometer devices. Across the studies, levels of interrater agreement in the manual assessment groupings for pupil reactivity and anisocoria were similar. Overall, providers frequently missed non-reactive pupils (Couret et al., 2015; Kerr et al., 2016; Meeker et al., 2005). False positive reporting was another common finding among providers who graded pupils as non-reactive when there was in fact a degree of reactivity detected by the pupillometer (Kerr et al., 2016; Meeker et al., 2005; Olson et al., 2016). When it came to identifying anisocoria, examiners missed this assessment finding approximately 50% the time (Couret et al., 2016; Kerr et al., 2016; Marshall et al., 2018; Olson et al., 2016).

In addition, findings varied regarding accuracy of the manual assessment of small versus large pupils. Both Meeker et al. (2005) and Kerr et al. (2016) reported that manual estimations were less accurate for large pupils (greater than 4 mm) while Couret et al. (2016) found the rate of error was greatest for mid-sized pupils (2-4 mm). Olson et al. (2016) compared four subsets of practitioners and found similar rates of agreement for pupils less than 3 mm and greater than 3 mm. Finally, three of the five studies reported

significantly less variability between pupillometer devices compared to the manual groups (Couret et al., 2016; Marshall et al., 2018; Meeker et al., 2005).

There were several limitations within the studies that should be noted. Meeker et al. (2005) included patients with a variety of non-neurological diagnoses, potentially limiting the availability of abnormal pupil measurements. Similarly, Marshall et al. (2018) increased their sample size by including healthy volunteers, which also limited the availability of abnormal pupil measurements in their cohort. Several studies reported that providers were unable to obtain some pupil measurements using the pupillometer due to periorbital edema and patient movement (Marshall et al., 2018; Meeker et al., 2005; Olson et al., 2016). Olson et al. (2016) noted this occurred more frequently in the first part of their study and proposed this may have been due to an operator learning curve. Penlights do not provide a consistent amount of illumination and are a potential source of variability noted by Couret et al (2016). However, the authors concluded that their study design represented real world practice.

In addition, there are several noteworthy limitations concerning this review. Although all five studies followed similar study design and methods, there were differences among them that may weaken their collective strength. Only the study by Olson et al. (2016) was statistically powered. Sample sizes of the remaining four studies were either small or enhanced using assessment data from healthy volunteers. All of the studies controlled for ambient light conditions, but others allowed the manual assessors to use a light source and pupil gauges of their choosing, which likely influenced their reliability and accuracy.

Furthermore, each of the studies implemented differing stabilization periods, ranging from five to 15 minutes to two hours, between manual and pupillometer assessments. Greater intervals between assessments may have increased the possibility that a change in patient condition occurred sometime in between the two, potentially resulting in a disagreement regarding pupil size, reactivity, or both. While agreement for pupil size was universally defined as a difference in reported pupil size greater than 1 mm, agreement for pupil reactivity was more inconsistent. For example, Couret et al. (2016) reported pupils as either reactive or non-reactive, Meeker et al. (2005) and Olson et al. (2016) reported pupils as brisk, sluggish, or non-reactive, and Marshall (2018) used a reference range of NPi values to compare their assessors' observations. Kerr et al. (2016) only looked at nurses' estimation of pupil size and therefore did not report pupil reactivity. Only Meeker et al. (2005) measured sedative use, a known precursor of pupillary changes. It is unclear what impact this variable may have had on assessment findings from the remaining studies. Lastly, each of the studies used data derived by the pupillometer to justify its use in practice.

In conclusion, five studies were reviewed to determine the effect of a pupillometer on the accuracy and reliability of the pupillary assessment. For pupil size, agreement between manual assessors was fair while agreement between manual assessors and the pupillometer was low. For the manual assessment, agreement of pupil reactivity was also poor. Overwhelmingly, the studies supported the use of the pupillometer device which appears to reduce variability and improve the reliability of the pupil assessment.

Recommendations and Implications for Advanced Nursing Practice

The foundation of the care and management of patients with acute neurologic injuries is the neurologic assessment and the PLR is one of its key components. Any change in pupil size or reactivity, however subtle, may indicate an imminent decline in neurologic condition. Changes in pupil size or reactivity may be due to several causes; however, when the change is manifested by unilateral, dilated pupils, the etiology is usually compression of CN III due to impending herniation and increased ICP (Meeker et al., 2005; Singhal & Josephson, 2014). If left undetected and untreated, the neurologic damage that occurs may become irreversible and can rapidly progress to coma and death (Hoffmann et al., 2012). Therefore, accurate and reliable pupil assessments are critical in detecting life-threatening conditions for which pupillary changes may represent an early sign.

The traditional pupil assessment inherently carries with it a significant degree of assessor subjectivity. This can lead to inconsistencies in findings between examiners and may lead to delayed detection, and therefore delayed treatment, of a deterioration in neurologic condition (Olson & Fishel, 2016). Health care providers should be aware of several factors that may contribute to disagreement between examiners, including varying clinical experience and skill levels, variations in assessment technique (e.g. validation with an adjunctive pupil gauge), differences in a given light source's illumination intensity, and differences in ambient light conditions (Olson & Fishel). The pupillometer's NPi algorithm provides a precise and objective measurement of pupil size and reactivity which may be trended from shift to shift and between examiners of varying experience levels. With advancements guided by medical technology, secondary

deterioration in neurologic condition after acute brain injury may potentially be avoidable (Meeker et al., 2005).

Advanced practice registered nurses (APRNs) are at the forefront of planning and implementing change to improve patient safety and the quality of care that is delivered. Practice change comes with its' own set of challenges, especially in fast paced, high stress environments like intensive care units. The essential first step toward obtaining staff support is to share information. The problem, unreliability of the manual pupillary assessment, may be discussed during staff meetings or morning huddles. At this time the pupillometer device can be introduced. The benefits of its' use should be clearly stated and conveyed to all the key stakeholders, including both nurses, APRNs, and physicians, because ultimately, their support and acceptance is vital to the successful integration of any new policy or technology into practice. In-services should be arranged to formally educate staff on the correct set-up and use of the device.

With proper training, the device is simple to use. Marshall et al. (2018) administered a survey to their nurses who reported the pupillometer was both acceptable and feasible for use in practice. Results from a quality improvement project aimed to implement pupillometry in a neurotrauma intensive care unit suggest the pupillometer was considered by nurses to be both easy to operate and a useful assessment tool (Anderson, Elmer, Shutter, Puccio, & Alexander, 2018). Anderson et al. (2018) also reported that nurses preference for the pupillometer over a flashlight increased significantly over the course of the study period and continued to increase after the completion of the project .

It takes time for new policies or technology to be accepted into routine practice. During the implementation phase, it is important to provide staff with consistent feedback and reinforce all necessary information until the change is adopted. For the full benefit to be realized, staff must be able to interpret the data derived by the pupillometer and clear assessment parameters must be established. Although it is expensive, the list price of one device is approximately five thousand dollars, the pupillometer may be used to more accurately assess the PLR in patients at greatest risk for life-threatening neurologic complications (P. Lane, personal communication, April 18, 2019). Such improvements in care may ultimately produce better patient outcomes and long-term cost savings for the health care system as a whole. However, more research is needed to determine if dynamic changes in pupillary function detected by the pupillometer precede other clinical signs of neurologic deterioration and whether the use of a pupillometer device improves patient care and health outcomes.

References

Anderson, M., Elmer, J., Shutter, L., Puccio, A., & Alexander, S. (2018). Integrating quantitative pupillometry into regular care in a neurotrauma intensive care unit. *J Neurosci Nurs, 50*(1), 30-36. doi:10.1097/jnn.0000000000000333

CASP Cohort Study Checklist. (2018). *Critical appraisal skills programme.* Retrieved from https://casp-uk.net/wp-content/uploads/2018/03/CASP-Systematic-Review-Checklist-Download.pdf

Chen, J. W., Gombart, Z. J., Rogers, S., Gardiner, S. K., Cecil, S., & Bullock, R. M. (2011). Pupillary reactivity as an early indicator of increased intracranial pressure: the introduction of the neurological pupil index. *Surgical Neurology International, 82(2)*, 1-12.

Couret, D., Boumaza, D., Grisotto, C., Triglia, T., Pellegrini, L., Ocquidant, P., . . . Velly, L. J. (2016). Reliability of standard pupillometry practice in neurocritical care: an observational, double-blinded study. *Critical Care, 20*(99), 1-9. doi:10.1186/s13054-061-1239-z

Haines, D. E. (2008). *Neuroanatomy: an atlas of structures, sections, and systems* (7th ed. ed.). Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Hickey, J. V. (2009). *The clinical practice of neurological and neurosurgical nursing (6th ed.).* (6th ed. ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Hoffmann, M., Lefering, R., Rueger, J. M., Kolb, J. P., Izbicki, J. R., Ruecker, A. H., . . . Lehmann, W. (2012). Pupil evaluation in addition to glasgow coma scale components in prediction of traumatic brain injury and mortality. *Br J Surg, 99 Suppl 1*, 122-130. doi:10.1002/bjs.7707

Kerr, R., Bacon, A. M., Baker, L. L., Gehrke, J. S., Hahn, D., Lillegraven, C. L., . . . Spilman, S. K. (2016). Underestimation of pupil size by critical care and neurosurgical nurses. *American Journal of Critical Care Nurses, 25*(3), 213-219.

Larson, M. D., & Singh, V. (2016). Portable infrared pupillometry in critical care. *Crit Care, 20*(1), 161. doi:10.1186/s13054-016-1349-7

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P., . . . Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol, 62*(10), e1-34. doi:10.1016/j.jclinepi.2009.06.006

Marshall, M., Deo, R., Childs, C., & Ali, A. (2018). Feasibility and variability of automated pupillometry among stroke patients and healthy participants: potential implications for clinical practice. *J Neurosci Nurs*. doi:10.1097/jnn.0000000000000416

McNett, M., Moran, C., Janki, C., & Gianakis, A. (2017). Correlations between hourly pupillometer readings and intracranial pressure values. *Journal of Neuroscience Nursing, 49*(4), 229-234.

Meeker, M., Du, R., Bacchetti, P., Privitera, C. M., Larson, M. D., Holland, M. C., & Manley, G. (2005). Pupil examination: validity and clinical utility of an automated pupillometer. *J Neurosci Nurs, 37*(1), 34-40.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med, 6*(7), e1000097. doi:10.1371/journal.pmed.1000097

NeurOptics, Inc. (2019). *Pupillometry education - critical care.* Retrieved from https://neuroptics.com/resource-center-neurocritical-care/

Olson, D. M., & Fishel, M. (2016). The use of automated pupillometry in critical care. *Crit Care Nurs Clin North Am, 28*(1), 101-107. doi:10.1016/j.cnc.2015.09.003

Olson, D. M., Stutzman, S., Saju, C., Wilson, M., Zhao, W., & Aiyagari, V. (2016). Interrater reliability of pupillary assessments. *Neruocritical Care, 24*, 251-257.

Polit, D. F., & Beck, C. T. (2017). *Nursing research : generating and assessing evidence for nursing practice* (10th ed.). Philadelphia: Wolters Kluwer Health.

Portran, P., Cour, M., Hernu, R., de la Salle, S., & Argaud, L. (2017). Pupillary abnormalities in non-selected critically ill patients: an observational study. *J Thorac Dis, 9*(8), 2528-2533. doi:10.21037/jtd.2017.07.58

Posner, J. B., Saper, C. B., Schiff, N. D., & Plum, F. (2008). Plum and posner's diagnosis of stupor and coma: 4th edition. *Archives of Neurology, 65*(5), 670-671. doi:10.1001/archneur.65.5.671

Rittenberger, J., Sangl, J., Wheeler, M., Guyette, F. X., & Callaway, C. W. (2010). Association between clinical examination and outcome after cardiac arrest. *Resuscitation, 81*(9), 1128-1132.

Ritter, A. M., Muizelaar, J. P., Barnes, T., Choi, S., Fatouros, P., Ward, J., & Bullock, M. R. (1999). Brain stem blood flow, pupillary response, and outcome in patients with severe head injuries. *Neurosurgery, 44*(5), 941-948.

Safa, R., Cuthbertson, F. M., Wulff, K. S., Downes, M., Foster, R. G., & Peirson, S. N. (2010). Changes in pupil area and dynamics following cataract surgery. *Investigative Ophthalmology and Visual Science, 51*(13), 5397.

Singhal, N. S., & Josephson, S. A. (2014). A practical approach to neurologic evaluation in the intensive care unit. *J Crit Care, 29*(4), 627-633. doi:10.1016/j.jcrc.2014.02.014

Wilson-Pauwels, L., & Akesson, E. J. (2001). *Cranial nerves in health and disease* (2nd ed.). Shelton, United States: PMPH USA, Ltd.

Yoo, Y. J., Hwang, J. M., & Yang, H. K. . (2017). Differences in pupillary light reflex between optic neuritis and ischemic optic neuropathy. *PloS one, 12*(10).

Appendix A

Table A-1. Study 1 (Meeker et al., 2005)

Aim	Sample/Setting	Design/methods
To examine the accuracy and reliability of an automated pupillometer compared to the manual pupil assessment.	Twenty randomly selected patients aged 4-87 years admitted to an intensive care unit at San Francisco General Hospital. Ten patients were admitted with acute neurologic diagnoses including hemorrhagic stroke and traumatic brain injury while the other ten were admitted with medical problems such as pancreatitis and pneumonia. Seventeen patients required continuous intravenous sedation while three patients had Glasgow Coma Scale scores between 3-5 and did not receive any sedating medications.	Prospective, observational, single-blind. Successive manual assessments were performed by two groups of examiners, each with one neurosurgical attending physician, two neurosurgical interns, and four advanced practice nurses. Each examiner recorded bilateral pupil size in millimeters and categorized pupil reactivity as non-reactive, sluggish, or brisk. A trained member of the investigative team obtained pupillometer measurements before and after each manual assessment. Ambient lighting conditions were consistent for all assessments which were performed no greater than five minutes apart. Pupil gauges were not used.

Table A-2. Study 2 (Olson et al., 2016)

Aim	Sample/Setting	Design method
To examine the interrater reliability of two methods of pupillary assessment: a comparison of manual pupil exams performed by two practitioners; and a comparison between a practitioner performing a conventional manual assessment and an automated pupillometer device.	194 RNs and 28 MDs at the University of Texas Southwestern Medical Center 127 patients with a neurological or neurosurgical diagnosis and pre-existing orders for serial pupil examinations	Prospective, observational, single-blind Study assessments were obtained using a convenience pairing of two practitioners from available registered nurses (RN), nurse practitioners, neurologists, neurosurgeons, and resident physicians who routinely perform pupil assessments. A trained member of the investigative team performed the pupillometer assessment. All practitioners that performed manual assessments were blind to each other's and the pupillometer results. Pupillometer values for size and reactivity were considered the reference results and provided the basis for which assessments obtained by the practitioner groups were compared.

Table A-3. Study 3 (Couret et al., 2016)

Aim	Sample/Setting	Design method
Part I – validation study, to determine inter-observer agreement of the pupillo-meter device. Part II – to evaluate agreement between the manual pupil assessment and the pupillometer.	Part I – 200 healthy volunteers age 21-58 Part II – 59 patients aged 18 years and older who were admitted within 48 hours of an acute brain injury to one of two neurocritical care units in Marseille and Saint-Pierre la Reunion, France between Jan 2012 and Dec 2012. Exclusion criteria included eye trauma, opalescent cataract, iris surgery, blindness, third cranial nerve damage.	Two part prospective, observational, double-blind. Part I – paired pupillometer measurements were obtained under a variety of ambient light conditions. Part II – Ten nurses with an average of ten years of experience in neurological nursing assessed their patient's pupil size, reactivity and for the presence of anisocoria. Each manual assessment was followed by a pupillometer assessment performed by a physician with no more than 5 minutes passing between measurements. For each manual assessment, the room lights were dimmed, nurses used pen lights and pupil gauges to estimate pupil size in millimeters (mm), and the opposite eye was covered. Anisocoria was defined as the difference of greater than 1 mm between eyes and pupils were graded as either reactive or nonreactive. Physicians and residents were trained in the use of the pupillometer. The patient's opposite eye was kept closed for the assessment. Every 24 hours, four measurements were obtained for each patient.

Table A-4. Study 4 (Kerr et al., 2016)

Aim	Sample/Setting	Design method
To evaluate neuroscience nurses' abilities to accurately measure pupil size and detect anisocoria	Phases I & II – A group of 30 critical care and neurosurgical nurses participated in phase 1; 27 nurses from the first phase participated in the second. Nurses had an average of 13.4 years of experience in nursing and 9.7 years in critical care or neurosurgical nursing. Data was collected in the spring 2012. Phase III – 93 patients aged 18 or older, admitted to the intensive care unit or a neurosurgical inpatient floor with a diagnosis of a subdural, subarachnoid, epidural or intracerebral hemorrhage, or another head injury; with at least one reactive pupil. Data was collected from February 2013 through February 2014 All phases were instituted at Iowa Methodist Medical Center in Des, Moines, Iowa.	Three-phase prospective observational cohort study. Phase I – nurses graded pupil size in millimeters (mm) from 12 randomly ordered black-and-white drawing. To evaluate interrater agreement, 2/10 (20%) drawings were duplicates. Phase II – nurses graded pupil size in mm from 24 color photographs of a human face. To evaluate interrater agreement, 4/20 (20%) were duplicates. To evaluate nurses' ability to identify anisocoria, 5 photographs depicted unequal pupils with a difference of 0.5 mm to 1.0 mm between the right and left pupil. Phase III – results of bedside nurses' manual pupillary assessments were compared to those obtained by a trained research coordinator using a pupillometer. The pupillometer assessments were done immediately following the manual assessment under the same lighting conditions.

Table A-5. Study 5 (Marshall et al., 2018)

Aim	Sample/Setting	Design method
To evaluate the feasibility and variability of automated pupillometry for use in the care of stroke patients.	Data was collected on the Hyperacute Stroke Unit (HASU) at Sheffield Teaching Hospitals, United Kingdom. Patients were admitted with a stroke and any of the following that increases the risk for sustaining a secondary neurologic injury: National institute of Health Stroke Scale score greater than 5, large vessel occlusion, intracerebral hemorrhage, diabetes, atrial fibrillation, hemorrhagic conversion of infarct or cerebral edema on initial computed tomography. Part I - 12 patients and their nurses (both day and night shift) Part II - 52 participants, both patients and healthy volunteers	Two-part prospective observational single blind study. Part I – participating patients and their nurses evaluated three qualities of the automated pupillometer: feasibility, acceptability, and safety for use. Feasibility: the nurses were asked to obtain pupillometer measurements along with their usual neurologic assessment. Acceptability: following the assessment period, a survey was distributed to both patients and nurses to evaluate how well each accepted the pupillometer for use in practice. Both groups were asked to rate the device's comfort and ease of use using a 5-point Likert scale. Safety: monitored for device-related adverse events. Part II – paired manual and pupillometer assessments were obtained by two blinded examiners within 15 minutes of each other. Assessments were performed on participating patients in the HASU and on healthy volunteers. Agreement for size was defined as a difference less than 1 mm between the observers' measurements and anisocoria was defined as the difference greater than or equal to 1 mm between pupils.

		For pupil reactivity, examiners agreed if they reported the same manual or NPi category: Nonreactive or NPi 0.0 Sluggish or NPi 0.1-2.9; CV less than 0.8 m/s Briskly reactive or NPi 3.0-5.0.

Appendix B

Table B-1. Study 1 (Meeker et al., 2005)

Variables measured	Data analysis	Study findings	Limitations
452 manual and pupillometer assessments Manual: pupil size in mm, pupil reactivity as non-reactive, sluggish or brisk Pupillometer: constriction velocity Sedatives infusing during assessment period	The before and after pupillometer results were interpolated to represent the true pupil size. Between examiner standard deviation (SD) was calculated for each assessment method and was summarized by medians. Statistical error for each method was measured by Spearman rank correlations. A nonparametric Wilcoxon signed-rank test measured the accuracy of the manual measurements against the interpolated true measurement.	For pupil size: 1. The median absolute error for the pupillometer was 0.23 (95% CI 0.20-0.31 mm) versus 0.50 mm (95% CI 0.47-0.60) for the manual measurement. 2. When compared to the manual method, the pupillometer's median improvement in accuracy was 0.27 mm (95% CI 0.20-0.30). 3. The median standard deviation in manual measurements was 0.58 mm (95% CI 0.50-0.58 mm) compared to that of the pupillometer which was 0.15 mm (95% CI 0.12-0.25 mm). 4. Variability between examiners increased as the size of the pupil increased. For pupil reactivity: 1. Inter-examiner disagreement for the pupillometer was 1.4% (95% CI 0%-7.6%) compared to the manual assessment of 39% (95% CI 28%-52%). 2. Manual examiners missed non-reactive pupils in three patients with a mean pupil size of 3.2 mm (SD=1.2mm). 3. Manual examiners inaccurately labeled briskly reactive pupils as non-reactive in 27 patient assessments with a mean pupil size of 2.6 mm (SD 0.7 mm).	Only half of the patient sample (n=20) was admitted with an acute neurologic process. Two different pupillo-meter devices were used. Pupillo-meter was unable to detect pupil response in 5 paired assessments due to periorbital edema.

Table B-2. Study 2 (Olson et al., 2016)

Variables measured	Data analysis	Study findings	Limitations
2329 paired manual pupil assessments: pupil size in mm, pupil reactivity as non-reactive, sluggish or brisk 2192 pupillometer assessments: size and reactivity measures	Cohen's kappa coefficient (*k*) measured interrater agreement for pupil size (≤3 or ≥3), shape and reactivity (non-reactive or reactive), both as itemized component scores and as a composite score aggregating all three components	1. Agreement between providers for the composite score was low (*k* = 0.26; 95% CI 0.23-0.29). 2. Agreement between providers for pupil size was fair (*k* = 0.54; 95% CI 0.50-0.57). 3. Agreement between providers for pupil reactivity was fair (*k* = 0.40; 95% CI 0.36-0.44). 4. Provider agreement with the pupillometer was low [(*k* = 0.29; 95% CI 0.27-0.32) and (*k* = 0.31; 95% CI 0.28-0.34)] for the first and second provider respectively. 5. When pupils reacted normally, providers agreed 95.7% of the time (2135/2230). When the pupillometer detected a fixed pupil, provider agreement decreased to 49.7% (94/189). 6. Of 83 non-reactive pupils detected by pupillometry, only 58/83 (69.9%) and 46/83 (55.4%) providers correctly identified this abnormal finding. Reactive pupils were reported in cases of cataracts and a prosthetic eye. 7. Variability in agreement between providers was evaluated in four groupings: (1) the whole cohort, (2) RN and RN, (3) MD and MD, and (4) RN and MD. Agreement for pupil size was similar within all four groups (*k* = 0.54, 0.53, 0.63, and 0.54, respectively). Agreement for pupil reactivity was also similar within all four groups (*k* = 0.64, 0.67, 0.55, and 0.54), respectively.	Internal validity may have been limited by including a variety of practitioners from nurses to junior residents to attending physicians. However, the authors conclude that the diversity of their study design strengthens its general-izability. Unable to obtain pupillo-meter measure-ments 5.9%, no data collected on history of glaucoma or iridectomy. The most common reasons that was cited by the authors were periorbital edema,

		8. For pupil size, agreement between both RN's and the pupillometer and MD's and the pupillometer were fair [(k = 0.30; 95% CI 0.27-0.32) and (k = 0.38; 95% CI 0.31-0.45)] respectively. 9. For pupil reactivity, agreement between both RN's and the pupillometer and MD's and the pupillometer were moderate [(k = 0.47; 95% CI 0.40-0.53) and (k = 0.42; 95% CI 0.22-0.61)] respectively.	patient movement, and cataracts or prosthetic eye. There were more unable to assess readings in the first half of the study compared to the second. The authors suggest this may have been due to an operator learning curve.

Table B-3. Study 3 (Couret et al., 2016)

Variables measured	Data analysis	Study findings	Limitations
400 paired pupillometer measure-ments Part II – patients yielded 406 pupillary measure-ments Manual: pupil size in mm, anisocoria, pupil reactivity as brisk, sluggish, or nonreactive Pupillometer: pupil size, anisocoria, and percent of pupillary light reflex	Intraclass correlations for two-way mixed-effects models was used to describe inter-rater agreement. The variation between operators relative to its mean was measured using the median coefficient of variation (CoV) and interquartile range (IQR). The pupil size CoV between senior and junior physicians was measured using Wilcoxon signed-rank test. Receiver operating characteristic analyses were used for three pupil size groups, less	Part I 1. Intraclass correlation coefficient for the maximum resting pupil size was 0.95 (95% CI, 0.93-0.97) and the minimum pupil size following light stimulation was 0.87 (95% CI, 0.83-0.89). 2. The mean difference between senior and junior practitioners was -0.06 ± 0.35 with a median CoV of 23.3% (IQR 23.26-23.32%). Part II 1. When compared to measurements obtained using the pupillometer, the nurses' manual pupil assessments were less accurate and less reliable. 2. Spearman's rank correlation coefficients were calculated for three test groups: pupils <2mm (n=61); pupils 2-4mm (n=232); and pupils > 4mm (n=113). The results suggested overall low agreement between the device and manually obtained measurements, with reported Spearman's rho values ranging from 0.39 (95% CI: 0.15-0.59; p = 0.002) for pupils <2mm in size to 0.44 (95% CI: 0.33-0.54; p < 0.001) for pupils that were 2-4mm, and 0.37 (95% CI: 0.19-0.51; p = 0.001) for pupils >4mm. 3. The pupillometer detected 30 cases of anisocoria of which 12 pupils were nonreactive. Nurses accurately reported 15/30 of them. They also inaccurately reported 16 cases of anisocoria.	Penlights do not provide a consistent amount of illumin-ation and is a potential bias. However, the authors conclude that their study design represents real world practice. Measure-ments may have been affected by the limit of 5 minutes between assess-ments.

	than 2 mm, 2-4 mm, and greater than 4 mm. Spearman's rank correlation coefficients were calculated to test the association between the manual and pupillometer measurements.	4. Nurses disagreed with the pupillometer 18% of the time (72/406). The rate of disagreement increased with pupils smaller than 2 mm to 39% (24/61) and decreased to 4% in pupils greater than 4 mm (5/122). For pupils 2-4 mm, the rate of disagreement was 19% (42/223). 5. Nurses inaccurately reported non-reactive pupils as reactive in 41 cases.	

Table B-4. Study 4 (Kerr et al., 2016)

Variables measured	Data analysis	Study findings	Limitations
Phases I & II – nurses' estimates of pupil size in mm. Phase III – nurses recorded pupil size in mm and graded reactivity as brisk, sluggish, or absent. 489 pupillo-meter assess-ments included minimum pupil size, maximum pupil size, and the neurologic al pupil index (NPi). Up to 20 assess-ments were performed on each patient.	Standard deviations and percentages Pearson correlation coefficients measured the association between the manual and pupillo-meter assess-ments.	1. Nurses consistently underestimated pupil size, were unable to reliably detect anisocoria, and inaccurately measured pupil reactivity. 2. There was a commensurate decrease in accuracy of nurses' pupil size estimates as actual pupil diameters increased, with poor accuracy for the largest pupil sizes. 3. Results from the first phase revealed 100% accuracy for pupils less than 4.0 mm but accuracy decreased to 54% for pupils with diameters greater than 5.0 mm. 4. In the second phase nurse estimates of pupil diameter were accurate 98.4% of the time for pupils less than 4.0 mm but only 37% of the time for pupils >4.5mm. 5. In the second phase nurses were consistent in their measurement for duplicate photographs only 11.7% of the time. 6. In the third phase, the mean pupil size reported by nurses was 2.92 (SD, 0.97) and the mean pupil size according to the pupillometer recording was 2.85 (SD, 0.90). When pupils were greater than 4.0 mm, the difference between the two was slightly greater at 0.6 mm (SD, 1.32). 7. In 85% of cases, nurses' assessments were within 1.0 mm of the pupillometer. 8. Nurses failed to detect sluggish readings 21% of the time. 9. Nurses inaccurately reported normally reactive pupils as sluggish 17% of the time. 10. Nurses correctly identified anisocoria just 58.1% of the time.	Small and homogenous sample size for phases I & II. In phase II the 2-dimensional images may limit validity. Images from pictures cannot convey pupil reactivity and therefore was not measured in the first two phases.

Table B-5. Study 5 (Marshall et al., 2018)

Variables measured	Data analysis	Study findings	Limitations
Part I – Feasibility was deter-mined by a compliance rate of 80% or greater. The device was considered acceptable only if every participant rated the device between 1 and 3 out of 5. Safety for use was determined by the absence of serious device-related adverse events. Part II – manual assessment: pupil size in mm, anisocoria and pupil reactivity as brisk, sluggish, or	Turkey boxplots illustrated the range of measureme nts for pupil size and reactivity for both the manual and pupillomet er groups. Interrater agreement was reported as percentages for pupil size, anisocoria, and reactivity. Interrater agreement was calculated using Cohen's kappa coefficient for anisocoria and pupil reactivity. Spearman correlation coefficient	Part I 1. Feasibility varied widely: a total of 92.7% of day shift (9am – 7pm) assessments were completed compared to only 30.8% of those from the night shift (9pm – 7am). 2. Acceptability: the average Likert score from patients was 1.4/5 and the average score from nurses was 2.4/5. 3. Safety: no events were reported thus the device was deemed safe for use. Part II 1. For pupil size, interrater agreement for the pupillometer was 99.2% with a Spearman correlation coefficient of 0.949 (95% CI, 0.929-0.969). 2. For pupil size, manual examiners agreed on pupil size just 61.4% of the time with a Spearman correlation coefficient of 0.633 (95% CI, 0.531-0.735). 3. The majority of manual observers (84.1%) graded pupil size as 3 or 4 mm while the corresponding pupillometer values varied between 1.9 and 6.1 mm. 4. Of 14 identified cases of anisocoria, interrater agreement for the pupillometer was substantial (98.5%, k = 0.660; 95% CI 0.039-1.00) and agreement between manual observer's was fair (89.4%, k = 0.306; 95% CI -0.078-0.690). 5. Agreement between manual observers and the pupillometer was poor (87.9%, k = -0.027; 95% CI -0.074-0.020). 6. Of the manual assessments graded as sluggish, none of them were associated with an NPi value less than 3.0. 7. For every pupillometer assessment with an NPi value less than 3.0, a brisk reaction was reported by the manual	The high number of healthy participants limited availability of abnormal measuremen ts. Low statistical power. Potential for observer bias.

non-reactive. Pupillo-meter: NPi, constriction velocity (CV). 132 paired measure-ments, 42 of which were from stroke patients and 90 were from healthy volunteers	was calculated for pupil size to test the associate-ion between the manual and pupillo-meter measure-ments.	observers (k = -0.026; 95% CI -0.042 to -0.010). 8. The mean CV for sluggish pupils was 1.60 (SD 1.08) m/s while it was significantly higher for brisk pupils 2.51 (SD 0.84) m/s (P = .001). 9. Of the pupils with a sluggish CV (n=10), only two were graded as sluggish by manual observers (20.0%, k = 0.006; 95% CI, -0.004–0.016).	

Appendix C

Table C-1. Study 1 (Meeker et al., 2005)
CASP Cohort Study Checklist ("CASP Cohort Study Checklist," 2018)

	Study 1 (Meeker et al., 2005)
1. Did the study address a clearly focused issue?	Yes. See Table A-1.
2. Was the cohort recruited in an acceptable way?	Yes. See Table A-1.
3. Was the exposure accurately measured to minimize bias?	Yes. See Table A-1.
4. Was the outcome accurately measured to minimize bias?	Yes. See Table B-1.
5. (a) Have the authors identified all important confounding factors? (b) Have they taken account of the confounding factors in the design and/or analysis?	Only half of the patient sample (n=20) was admitted with an acute neurologic process. Two different pupillometer devices were used. Pupillometer was unable to detect pupil response in 5 paired assessments due to periorbital edema. Confounding factors are not discussed.
6. (a) Was the follow up of subjects complete enough? (b) Was the follow up of subjects long enough?	There was no follow up as part of the study.
7. What are the results of the study?	See Table B-1.
8. How precise are the results?	95% confidence intervals
9. Do you believe the results?	Yes.
10. Can the results be applied to the local population?	Yes.
11. Do the results of this study fit with the results of other available evidence?	Yes.
12. What are the implications of this study for practice?	The pupillometer provided nurses with an accurate and reliable measure of both pupil size and reactivity.

Table C-2. Study 2 (Olson et al., 2016)
CASP Cohort Study Checklist ("CASP Cohort Study Checklist," 2018)

	Study 2 (Olson et al., 2016)
1. Did the study address a clearly focused issue?	Yes. See Table A-2.
2. Was the cohort recruited in an acceptable way?	Yes. See Table A-2.
3. Was the exposure accurately measured to minimize bias?	Yes. See Table A-2.
4. Was the outcome accurately measured to minimize bias?	Yes, however assessors were not blind to the aims of the study. See Table B-2.
5. (a) Have the authors identified all important confounding factors? (b) Have they taken account of the confounding factors in the design and/or analysis?	Internal validity may have been limited by including a variety of practitioners from nurses to junior residents to attending physicians. The authors conclude that the diversity of their study design strengthens its generalizability.
6. (a) Was the follow up of subjects complete enough? (b) Was the follow up of subjects long enough?	There was no follow up as part of the study.
7. What are the results of the study?	See Table B-2.
8. How precise are the results?	95% confidence intervals
9. Do you believe the results?	Yes.
10. Can the results be applied to the local population?	Yes.
11. Do the results of this study fit with the results of other available evidence?	Yes.
12. What are the implications of this study for practice?	Accurate and reliable pupil assessments are a clinical necessity. The pupillometer may provide a means of overcoming the inherent faults of the manual assessment technique.

Table C-3. Study 3 (Couret et al., 2016)
CASP Cohort Study Checklist ("CASP Cohort Study Checklist," 2018)

	Study 3 (Couret et al., 2016)
1. Did the study address a clearly focused issue?	Yes. See Table A-3.
2. Was the cohort recruited in an acceptable way?	Yes. See Table A-3.
3. Was the exposure accurately measured to minimize bias?	Yes. See Table A-3.
4. Was the outcome accurately measured to minimize bias?	Yes. See Table B-3.
5. (a) Have the authors identified all important confounding factors? (b) Have they taken account of the confounding factors in the design and/or analysis?	Penlights do not provide a consistent amount of illumination and is a potential bias. However, the authors conclude that their study design represents real world practice. Measurements may have been affected by the limit of 5 minutes between assessments.
6. (a) Was the follow up of subjects complete enough? (b) Was the follow up of subjects long enough?	There was no follow up as part of the study.
7. What are the results of the study?	See Table B-3.
8. How precise are the results?	95% confidence intervals
9. Do you believe the results?	Yes.
10. Can the results be applied to the local population?	Yes.
11. Do the results of this study fit with the results of other available evidence?	Yes.
12. What are the implications of this study for practice?	Nurses frequently disagreed or were inaccurate in their manual pupillary assessment. Improving the reliability of the pupil assessment may promote early detection of pupillary changes and potentially improve patient outcomes.

Table C-4. Study 4 (Kerr et al., 2016)
CASP Cohort Study Checklist ("CASP Cohort Study Checklist," 2018)

	Study 4 (Kerr et al., 2016)
1. Did the study address a clearly focused issue?	Yes. See Table A-4.
2. Was the cohort recruited in an acceptable way?	Yes. See Table A-4.
3. Was the exposure accurately measured to minimize bias?	Yes. See Table A-4.
4. Was the outcome accurately measured to minimize bias?	Yes. See Table B-4.
5. (a) Have the authors identified all important confounding factors? (b) Have they taken account of the confounding factors in the design and/or analysis?	Small and homogenous sample size for phases I & II. Phase II the 2-dimensional images may limit validity. Images from pictures can not convey pupil reactivity and therefore was not measured in the first two phases.
6. (a) Was the follow up of subjects complete enough? (b) Was the follow up of subjects long enough?	There was no follow up as part of the study.
7. What are the results of the study?	See Table B-4.
8. How precise are the results?	95% confidence intervals
9. Do you believe the results?	Yes.
10. Can the results be applied to the local population?	Yes.
11. Do the results of this study fit with the results of other available evidence?	Yes.
12. What are the implications of this study for practice?	The pupillometer provided a more accurate and reliable measure of pupil size and reactivity compared to the manual assessment performed by nurses. Improving the reliability of the pupil assessment may promote early detection of

	pupillary changes and potentially improve patient outcomes.

Table C-5. Study 5 (Marshall et al., 2018)
CASP Cohort Study Checklist ("CASP Cohort Study Checklist," 2018)

	Study 5 (Marshall et al., 2018)
1. Did the study address a clearly focused issue?	Yes. See Table A-5.
2. Was the cohort recruited in an acceptable way?	Yes. See Table A-5.
3. Was the exposure accurately measured to minimize bias?	Yes. See Table A-5.
4. Was the outcome accurately measured to minimize bias?	Yes, See Table B-5.
5. (a) Have the authors identified all important confounding factors? (b) Have they taken account of the confounding factors in the design and/or analysis?	The high number of healthy participants limited availability of abnormal measurements. Low statistical power. Potential for observer bias.
6. (a) Was the follow up of subjects complete enough? (b) Was the follow up of subjects long enough?	There was no follow up as part of the study.
7. What are the results of the study?	See Table B-5.
8. How precise are the results?	95% confidence intervals
9. Do you believe the results?	Yes.
10. Can the results be applied to the local population?	Yes.
11. Do the results of this study fit with the results of other available evidence?	Yes.
12. What are the implications of this study for practice?	The pupillometer is feasible, accepted by staff and patients alike, and is safe for use in clinical practice. It provides a more reliable measure of pupil size and reactivity. The integration of pupillometry in routine neurologic monitoring of stroke patients may improve the detection of early

	neurological deterioration and thereby hasten the delivery of time sensitive, life-saving treatments.

Appendix D

	Inter-examiner agreement manual		Agreement between manual & pupillometer		Inter-examiner agreement pupillometer	
	Pupil size	Pupil reactivity	Pupil size	Pupil reactivity	Pupil size	Pupil reactivity
1	Median SD 0.58 mm (0.50 to 0.58 mm). The degree of error/ between examiner SD increased along with an increase in pupil size.	Disagree-ment was 39% (28 to 52%).	The median absolute error of the manual assessment 0.50 mm (0.47 to 0.60); 0.27 mm (p=0.0001) greater than the pupillo-meter group.	Pupillo-metry consistently found no pupillary reflexes in 3 patient assessments that were reported as present by manual examiners. Pupillo-metry reported brisk pupils for 27 patients that were reported as non-reactive by the manual examiners.	Median SD 0.15 mm (95% CI 0.12 to 0.23 mm). Median absolute error of the pupillo-meter assessment 0.23 (0.20 to 0.31 mm). Degree of error increased along with an increase in pupil size.	Disagree-ment was 1.4% (0% to 7.6%).
2	Fair k = 0.54; (0.50 to 0.57). Practitioner agreement for anisocoria was moderate k = 0.60; (0.54 to 0.64).	Reactive vs fixed: moderate k = 0.64; (0.58 to 0.71). Reactive vs sluggish vs fixed: fair k = 0.40; (0.36 to 0.44).	Provider 1 fair k = 0.29; (0.27 to 0.32). Provider 2 fair k = 0.31; (0.28 to 0.34).	Reactive vs fixed: Provider 1 moderate k = 0.52; (0.44 to 0.60). Provider 2 fair k = 0.40; (0.32 to 0.49).	Not reported	Not reported

		Agreement for fixed pupils only (k = 0.28) for the right eye and k = 0.47) for the left eye.		Accurately reported fixed pupils: Provider 1 58/83 (69.9%) Provider 2 46/83 (55.4%)		
3	Global area under the ROC curve was 0.75 (95% CI: 0.70 to 0.79). For pupils <2 mm 0.89 (0.85 to 0.92); pupils 2-4 mm 0.59 (0.54 to 0.64) and pupils >4 mm 0.86 (0.82 to 0.89).	Not reported	Spearman's rho overall suggests low agreement for pupils in three test groups: Pupils <2mm – 0.39 (95% CI: 0.15 to 0.59; p = 0.002) Pupils 2-4mm – 0.44 (95% CI: 0.33 to 0.54; p < 0.001), Pupils >4mm – 0.37 (95% CI: 0.19 to 0.51; p = 0.001). Nurses inaccurately reported 16 cases of anisocoria.	The pupillo-meter detected 30 cases of anisocoria of which 12 pupils were nonreactive. Nurses accurately reported 15/30 of them. Nurses inaccurately reported non-reactive pupils as reactive in 41 cases. Nurses disagreed with the pupillo-meter 18% of the time (72/406). Rates of disagree-ement were greatest for pupils	Intraclass correlation coefficient for the maximum resting pupil size was 0.95 (95% CI, 0.93 to 0.97) and after light stimulation the minimum pupil size was 0.87 (0.83 to 0.89).	Mean percent reduction in pupil size for healthy volunteers was 40 ±7%.

				<2mm, 39% (24/61).		
4	Phase 1: For pupils >5 mm, measurement accuracy was 54%. Phase 2: nurses were accurate 37% of the time for pupils >4.5mm When nurses were shown the same color photograph twice, they consistently and accurately measured the duplicate pupil diameter only 11.7% of the time. Phase 3: Mean pupil size 2.92 mm (SD 0.97).	Not reported	Agreement was close for pupils <4.0 mm but when they were >4.0 mm, the mean difference was 0.6 mm (SD 1.32 mm). Nurses assessments were within 1.0 mm of the pupillo-meter 85% of the time. Nurses correctly identified anisocoria 58.1% of the time.	Nurses failed to detect sluggish readings 21% of the time. Nurses inaccurately reported normally reactive pupils as sluggish 17% of the time.	Not reported	Not reported
5	Manual examiners agreed 61.4% of the time;	Agreement was poor 92.4% k = -0.039; (-	84.1% of observers graded pupils 3-4 mm while	None of the manual assessments graded as sluggish	Agreement for anisocoria, 98.5%, k = 0.660;	Agreement was fair 97.7% k = 0.389;

	Spearman correlation coefficient 0.633 (95% CI; 0.531 to 0.735). Agreement for anisocoria, 89.4%, *k* = 0.306; (-0.078 to 0.690).	0.063 to -0.015).	the corres-ponding pupillo-meter values varied between 1.9 and 6.1 mm. Agreement for anisocoria 87.9%, *k* = -0.027; (95% CI -0.074 to 0.020).	were associated with an NPi value less than 3.0. For every pupillo-meter assessment with an NPi value less than 3.0, a brisk reaction was reported by the manual observers *k* = -0.026; (-0.042 to -0.010). 2 of 10 manually reported sluggish pupils had abnormal CV results, 20.0%, *k* = 0.006, (-0.004 to 0.016).	0.039 to 1.00. Agreement for CV was perfect 100% (*k* = 1.00, 1.00 to 1.00).	(-0.160 to 0.938).