Background  Early detection of pupillary changes in patients with head injuries can alert the care team to increasing intracranial pressure. Previous research has shown inconsistencies in pupil measurement that are most likely due to the subjective nature of measuring pupils without the assistance of technology.

Objectives  To evaluate nurses’ abilities to assess pupil diameter accurately and detect unequal pupils.

Methods  In a 3-part study, the accuracy of critical care and neurosurgical nurses’ assessments of pupils was determined. The study included assessment of drawings of eyes with an iris and pupil, examination of photographs of human eyes, and bedside examination of patients with a head injury.

Results  Subjective assessments of pupil diameter and symmetry were not accurate. Across all phases of the study, pupil diameters were underestimated and the rate of error increased as pupil size increased. Nurses also failed to detect anisocoria and misidentified pupil reactivity. In addition, nearly all nurses relied on subjective estimation, even when tools were available.

Conclusions  Critical care and neurosurgical nurses underestimated pupil size, were unable to detect anisocoria, and incorrectly assessed pupil reactivity. Standardized use of pupil assessment tools such as a pupillometer is necessary to increase accuracy and consistency in pupil measurement and to potentially contribute to earlier detection of subtle changes in pupils. If pupillary changes are identified early, diagnostic and treatment intervention can be delivered in a more timely and effective manner. (American Journal of Critical Care. 2016;25:213-219)
Regular assessments of pupils are important for monitoring and assessing neurological function of patients with head injuries. Changes in pupil size may signal neurological deterioration and a need for a change in clinical management. Serial assessments are vital for early identification of subtle changes in patients’ neurological status. Prior research, however, has documented that health care professionals are inconsistent and inaccurate in measuring pupil diameter. Although protocols specify the conditions under which pupillary examinations should occur (e.g., room lighting, angle of light shone in the eye), health care professionals are rarely compliant with these recommendations. This lack of compliance is problematic if inaccuracies and inconsistencies prevent detection of pupillary change and delay clinical intervention.

Clinical evaluation of pupils focuses on 4 characteristics: size or diameter, reactivity to light, shape, and presence of anisocoria. Of these, changes in pupil diameter or development of anisocoria may be most important. Pupil size is measured in millimeters, and the mean pupil diameter is from 2 to 6 mm. Although both pupils are typically the same diameter, a discrepancy of less than 1.0 mm is considered to be within the normal range. In a trauma population, unequal pupils are one indicator of traumatic brain injury; thus detection of anisocoria may be clinically relevant.

Technology exists to diminish measurement discrepancies. A pupillometer is an infrared system that analyzes pupil dynamics over a brief time period, during which time the angle and intensity of light is controlled and multiple assessments are recorded. The instrument has higher reliability and greater precision than subjective estimates and can detect changes in pupil reactivity hours before changes in intracranial pressure are noted. Few studies have compared subjective measurement by nursing staff with objective measurement by using a pupillometer.

The purpose of this 3-phase study was to evaluate nurses’ accuracy in assessing pupil diameter and symmetry. Two phases assessed accuracy of estimation with drawings and pictures, and the final phase compared standard bedside assessment with assessments with a pupillometer. The research objectives were 3-fold: determine the accuracy of current practice, specify the thresholds at which the quality of subjective pupil measurement degrades, and examine nurses’ ability to detect sluggish and unequal pupils.

Methods

The following study has 3 subcomponents, which we refer to as phase 1, phase 2, and phase 3. All phases were approved by the institutional review board at Iowa Methodist Medical Center, part of the UnityPoint Health System, in Des Moines, Iowa.

Procedure: Phase 1 and Phase 2

Critical care and neurosurgical nurses at Iowa Methodist Medical Center, a level 1 trauma facility in Des Moines, were asked to participate in a study assessing variability in pupil estimation. Nurses were recruited until a sample size of 30 was reached. They were asked to participate in phase 1 and phase 2, with 2 weeks between phases. All data were collected in the spring of 2012. Nurses were given study forms and told that filling out the forms indicated their consent. It took about 5 or 10 minutes to go through the study forms. Demographic variables were collected at each phase, including sex (male/female), age (20-35 years old, 36-50 years old, or ≥51 years old), years of experience in nursing, and whether or not the nurse wore corrective lenses (yes/no).

For the first phase, each nurse was given a packet of 12 randomly ordered black and white drawings of an eye with an iris and pupil. The size of the pupil ranged from 1 mm to 10 mm (Figure 1). Two of 10 drawings (20%) in each packet were duplicated to...
check for intrarater agreement. Each sheet of paper contained a single eye figure. In the second phase, nurses were presented with 24 color pictures of eyes; each picture included both the right and left eyes (Figure 2). Four of 20 photos in each packet (20%) were duplicated to check for intrarater agreement. Half of the photos were altered in photo editing software to expand the size of 1 pupil to create anisocoria. All pupils were between 2.5 mm and 8 mm in diameter, and each packet included 5 pictures with unequal pupils that differed by 0.5 to 1.0 mm. All nurses who participated in phase 2 also participated in phase 1.

Procedure: Phase 3

In the third phase, bedside nurse assessments were compared with measurements made with the pupillometer. Data were collected prospectively from February 2013 through February 2014 as patients were admitted to the intensive care unit or a neurosurgical inpatient floor with qualified diagnoses. If a patient had repeat hospitalizations during the study period, only the first encounter was included in analyses. The study phase had approval from the institutional review board, but the requirement of informed consent was waived because the risk to patients was no more than minimal.

Patients were identified in the electronic medical record (EMR) on a daily basis, and the study team was alerted when a patient met criteria. Inclusion criteria were age 18 years or older; diagnosis of subdural, subarachnoid, epidural, or intracerebral hemorrhage, or other head injury; and at least 1 reactive pupil. Patients were excluded from the study if they had ocular injury or malformation in both eyes or if they were discharged from the hospital before the study team could complete the initial assessment. Each eye was considered a measurement, so in most cases, a single visit to the bedside resulted in 2 measurements. Up to 20 measurements were collected from each patient.

Bedside examinations using the pupillometer were conducted by study coordinators. Initial training was conducted via a web conference with the manufacturer of the pupillometer. The instrument used was the NPI-100 (NeuroOptics Inc), which is a noninvasive, hand-held device that stored all assessments on the instrument through the duration of the study. This study was investigator-initiated, and no payments or in-kind donations were received from the manufacturer.

For each measurement, the study coordinator recorded the date and time of the assessment, the location in the hospital (intensive care unit or general inpatient floor), whether or not the bedside nurse used a pupil card or dimmed the lights, the pupil diameter (in millimeters), and whether the pupil reaction was noted to be absent, sluggish, or brisk.

The pupillometer provided additional data, including minimum and maximum diameter readings and a neurological pupil index (NPI). The NPI is an algorithm for describing a pupil’s reactivity to light; a reading between 3 and 5 is considered brisk (normal), readings below 3 are considered sluggish, and the pupil is considered nonreactive as NPI.

Figure 2 Example of photo that nurses were presented with in phase 2 of study.
The accuracy of nurses’ assessment decreased as pupil diameter increased. In 7 instances, the pupillometer failed to record a minimum diameter. In these cases, we imputed a minimum by using the mean difference between minimum and maximum values (0.41), subtracted the mean from the recorded maximum measurement, and determined the median diameter. Also in 7 instances, the pupillometer did not record NPI; these readings were considered missing and no imputation method was used. If a patient blinked during the assessment, the study coordinator was prompted to rescann the pupil.

**Statistical Methods**

All analyses were conducted in IBM SPSS Basic Statistics for Windows, version 20.0 (IBM Corp., 2011). Descriptive statistics are presented as means (standard deviations) for continuous variables and as counts (percentages) for categorical variables. Correlations were assessed as Pearson correlation coefficients with a 95% confidence interval.

**Results**

**Phase 1**

In phase 1, 30 nurses assessed pupil size by using the method they most typically use for patient care. Two nurses used a pupil card and the other 28 nurses (93%) made subjective estimates. The nurses had a mean of 13.4 years of experience in nursing and 9.7 years of experience in critical care or neurosurgical nursing.

The accuracy of nurses’ measurements decreased as pupil diameter increased (Table 1). When the pupil was 1 mm in diameter, the nurses underestimated the pupil size by a mean of 1.2 mm. Nurses were internally inconsistent, with only 49% of pupils measured identically in the duplicated drawings. Accuracy was not correlated with the nurses’ use of corrective lenses, sex, age, or years of experience in a critical care or neurosurgical setting.

To determine the threshold at which accuracy was impaired, the accuracy of measurement was compared by using a cutoff of 4 mm. When the actual measurement in the drawing was 4.0 mm or less, 100% of nurses gave a measurement in that range. However, when the actual measurement was 5.0 mm or greater, only 54% of nurses indicated a measurement greater than 5.0 mm.

**Phase 2**

The second phase included 27 nurses. Two nurses used a pupil card and the other 25 nurses (93%) made subjective estimates. The accuracy of the nurses’ measurements again decreased as pupil diameter increased (Table 1). When the pupil was 2.5 mm in diameter, the diameter was underestimated by a mean of 0.4 mm, but when the pupil was 8.0 mm in diameter, the size was underestimated by a mean of 1.4 mm. When assessing the pictures with unequal pupils, one-third of the photographs were correctly identified as unequal.

Intrarater agreement was assessed. Overall, nurses measured the duplicated photographs consistently only 54.8% of the time; the consistency level varied widely as 2 nurses were consistent with their own measurements 0% of the time and 2 nurses were consistent 100% of the time. However, when the nurses saw the exact same photographs at 2 different time points, they measured the pupils consistently and correctly only 11.7% of the time. The rate of internal agreement did not correlate

---

**Table 1**

<table>
<thead>
<tr>
<th>Nurses’ measurement</th>
<th>Objective measurement, mm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Phase 1: Drawings (n = 234)</strong></td>
<td></td>
</tr>
<tr>
<td>Nurse, mm</td>
<td>1.1</td>
</tr>
<tr>
<td>Difference, mm</td>
<td>0.1</td>
</tr>
<tr>
<td>Number of assessments</td>
<td>28</td>
</tr>
<tr>
<td><strong>Phase 2: Photographs (n = 540)</strong></td>
<td></td>
</tr>
<tr>
<td>Nurse, mm</td>
<td>2.1</td>
</tr>
<tr>
<td>Difference, mm</td>
<td>-0.4</td>
</tr>
<tr>
<td>Number of assessments</td>
<td>11</td>
</tr>
<tr>
<td><strong>Phase 3: Pupillometer (n = 489)</strong></td>
<td></td>
</tr>
<tr>
<td>Nurse, mm</td>
<td>1.9</td>
</tr>
<tr>
<td>Difference, mm</td>
<td>0.9</td>
</tr>
<tr>
<td>Number of assessments</td>
<td>8</td>
</tr>
</tbody>
</table>

*Objective measurement indicates the true measurement, as assessed with a ruler in phases 1 and 2 and with a pupillometer in phase 3. For each objective measurement, the first number indicates the beginning of the range. For example, 1.0 indicates the measurement was 1.0 or greater and less than 1.5. Dashes indicate that no drawing, photograph, or patient was available at that measurement increment.*
with any of the demographic variables in the study, and the rate of internal consistency did not correlate with the overall rates of accuracy. 

Accuracy was once again compared with the cutoff of 4.0 mm. When the actual measurement was 4.0 mm or less, 98.4% of nurses reported a diameter of 4.0 mm or less; when the actual measurement was 4.5 mm or greater, only 37% reported a diameter of 4.5 mm or greater.

Phase 3
In the third phase, 489 assessments were conducted on 93 patients. The mean age of patients in the study was 61.0 (SD, 19.1) years and 60% of patients were male (Table 2). The most common diagnoses were subdural hemorrhages and subarachnoid hemorrhages, and mean Injury Severity Score for patients with traumatic injuries was 25 (SD, 9.8). Sixty-one percent of the assessments were conducted in the intensive care unit (Table 3). None of the assessments were conducted with the assistance of a pupil card, but room lights were dimmed in more than two-thirds of the assessments.

A mean pupil size of 2.92 (SD, 0.97) mm was obtained from the measurements recorded by the nurse, and the pupillometer recorded a mean pupil size of 2.85 (SD, 0.90) mm. The differences between nurse and pupillometer assessments were very close when the actual diameter was less than 4.0 mm (Table 1). However, assessments were less accurate when the diameter was greater than 4.0 mm (mean, 0.6 mm; SD, 1.32 mm). Overall, nurses’ bedside assessments were within 1.0 mm of the pupillometer reading in 85% of the assessments. Figure 3 shows a scatterplot of the pupillometer and nurse assessments, indicating that convergence was high but that out-of-range values were present. Accuracy was positively correlated with patient’s age ($r = 0.09$, $P = .04$), initial GCS score ($r = 0.17$, $P < .001$), and if the mechanism of injury was a fall ($r = 0.16$, $P < .001$).

Nurses accurately assessed pupil reactivity in 82.4% of the assessments. However, they had relatively high rates of false-positives and false-negatives. Specifically, in 33 assessments, the pupillometer provided an NPi reading considered to be sluggish (NPi < 3.0), and nurses missed this sluggishness in 7 instances (21%). In 444 assessments, the pupillometer provided an NPi reading considered to be normal, and the nurses reported a sluggish pupil in 77 of those instances (17%).

The sample included 242 paired sets (right and left eyes). The pupillometer determined that 31 (12.8%) of the paired sets were unequal (diameters differed by 1.0 mm or greater), but nurses correctly identified only 58.1% of the unequal sets. Conversely, nurses reported 40 sets as unequal; of these, only 17 (42%) were unequal, as determined by the pupillometer.

**Discussion**
Study results indicate that accurate assessment of pupil diameter and reactivity is difficult when measured subjectively. When evaluating drawings, photographs, and patients, nurses had lower rates of accuracy as pupil diameter increased. Specifically, accuracy declined when the objective measurement exceeded 4.0 mm. In addition, nurses did not identify unequal pupils correctly, with accurate identification in only 33% of the pictures and 58% of the patients. Results suggest that subjective assessment of
If pupillary changes are noted early, treatment can be timely and effective.

pupil diameter and anisocoria is fraught with problems and could lead to delays or failures in detecting important signs of neurologic deterioration.

The study was designed in 3 phases in order to progress from the most basic assessment (simple black-and-white drawings) to the more complex assessments on real patients. Findings reveal that manual pupil assessment is susceptible to inaccuracy at even the most basic level (phase 1). In other words, manual assessment of a static, simple drawing was prone to error; assessment is inherently more complicated when evaluating a real patient.

Results are consistent with prior research but extend knowledge on this topic in several important ways. First, in all phases the trend was toward underestimates of the true diameters, and the rate of error increased as pupil size increased (Figure 3). Previous work is discrepant about whether the accuracy of assessment varies as pupil size increases. For example, Hults and colleagues\(^\text{11}\) reported that visual assessment was more difficult as pupil diameter decreased, but Meeker et al\(^\text{4}\) and Taylor et al\(^\text{10}\) found more error as pupil diameter increased. Arguably, in the neurologic critical care setting, important evaluation occurs when pupils are large and enlarging as pupil dilation may indicate cerebral ischemia or herniation.\(^\text{12}\) Findings of this study indicate that pupils will be underestimated by as much as 1.5 mm, which could affect clinical decisions. In addition, we quantify the threshold where accuracy deteriorates (between 4 and 4.5 mm). Other studies\(^\text{6}\) have specified an error rate but failed to note the diameters where subjective estimates degrade significantly. This threshold is important because it could specify the diameters at which more frequent and accurate measurement is warranted.

Second, we show that nurses were inconsistent with their own measurements of duplicated images. Previous studies\(^\text{13}\) have looked at interrater reliability, but no study of which we are aware has specified such low rates of intrarater reliability. This finding is crucial because 1 nurse may conduct all the serial assessments on a patient during a shift, and inaccurate measurements could result in failure to detect pupillary changes.

Third, we found that subjective estimates were inadequate in detecting anisocoria that could be imperative to early detection of neurological impairment, which confirms the findings of other studies.\(^\text{15,14}\) Nurses also had high rates of false-positives and false-negatives when assessing pupil reactivity. More than 20% of sluggish pupils were misjudged as normal in reactivity, and 17% of normal pupils were misjudged by the nurses as sluggish. This finding reaffirms the difficulty in accurately assessing pupillary change.

Limitations

This study has several limitations. First, the sample size of nurses in phase 1 and phase 2 was fairly small and relatively homogeneous, although there is no reason to believe that nurse demographics affect study outcomes. Second, in the first 2 phases, measurements were taken from 2-dimensional images on paper and not from real patients, which may limit the validity of the results. Images from drawings and photographs cannot show pupil reactivity, which is an important aspect of a pupil examination. Third, published reports have established that dark brown irises may be more difficult to assess,\(^\text{15}\) and we did not control for the color of the iris in our analyses. Prior researchers have found no diminished accuracy with the pupillometer based on the color of the iris,\(^\text{16}\) but more work is warranted to examine the relationship between iris color and pupil reactivity in manual assessment. Fourth, only 2 nurses used a pupil card in the first 2 phases, and no nurse used a measurement tool in phase 3. This finding is consistent with results of other studies,\(^\text{4,8}\) yet it should be noted that accuracy might increase if nurses were required to consistently use a tool for pupil measurement. Finally, published reports state that pupils become smaller as people age.\(^\text{17,18}\) The mean age of patients in phase 3 was 61.0 (SD, 19.06) years, which potentially biases the results and limits the generalizability of our findings.
Summary

If pupillary changes are identified early, diagnostic and treatment intervention can be delivered in a timely and effective manner. Study results demonstrate that nurses underestimate pupil diameter when using subjective methods and are inconsistent with their own readings. Further work is warranted to determine if accuracy improves through the use of a pupillometer, leading to earlier detection of pupillary changes and improved outcomes for patients.

ACKNOWLEDGMENTS

We extend our gratitude to Garret Lechtenberg, Sheryl Sahr, Kit VanderPloeg, and all the nurses who participated in the study.

FINANCIAL DISCLOSURES

None reported.

REFERENCES


To purchase electronic or print reprints, contact American Association of Critical-Care Nurses, 101 Columbia, Aliso Viejo, CA 92656. Phone, (800) 899-1712 or (949) 362-2050 (ext 532); fax, (949) 362-2049; e-mail, reprints@aacn.org.
Underestimation of Pupil Size by Critical Care and Neurosurgical Nurses
Robert G. Kerr, Andrea M. Bacon, Laura L. Baker, Janelle S. Gehrke, K. Danielle Hahn, Cheryl L. Lillegraven, Catherine Hackett Renner and Sarah K. Spilman
Am J Crit Care 2016;25 213-219 10.4037/ajcc2016554
©2016 American Association of Critical-Care Nurses
Published online http://ajcc.aacnjournals.org/

Personal use only. For copyright permission information:
http://ajcc.aacnjournals.org/cgi/external_ref?link_type=PERMISSIONDIRECT

Subscription Information
http://ajcc.aacnjournals.org/subscriptions/

Information for authors
http://ajcc.aacnjournals.org/misc/ifora.xhtml

Submit a manuscript
http://www.editorialmanager.com/ajcc

Email alerts
http://ajcc.aacnjournals.org/subscriptions/etoc.xhtml