Title

Interobserver Agreement in the Assessment of Clinical Findings in Children with Headaches

Authors

Tsze, Daniel S
Cruz, Andrea T
Mistry, Rakesh D
et al.

Publication Date

2020-06-01

DOI

10.1016/j.jpeds.2020.02.018

Peer reviewed

HHS Public Access

Author manuscript

J Pediatr. Author manuscript; available in PMC 2021 June 01.

Published in final edited form as:

J Pediatr. 2020 June ; 221: 207–214. doi:10.1016/j.jpeds.2020.02.018.

Interobserver Agreement in the Assessment of Clinical Findings in Children with Headaches

Daniel S. Tsze, MD, MPH[1], **Andrea T. Cruz, MD, MPH**[2], **Rakesh D. Mistry, MD, MPH**[3], **Ariana E. Gonzalez, BA**[1], **Julie B. Ochs, BA**[1], **Lawrence Richer, MD, MSc**[4], **Nathan Kuppermann, MD, MPH**[5], **Peter S. Dayan, MD, MSc**[1]

[1]Department of Emergency Medicine. Division of Pediatric Emergency Medicine. Columbia University College of Physicians and Surgeons. New York, NY.

[2]Department of Pediatrics, Baylor College of Medicine. Houston, TX.

[3]Department of Pediatrics, University of Colorado School of Medicine. Aurora, CO.

[4]Department of Pediatrics, University of Alberta. Edmonton, Alberta, Canada.

[5]Department of Emergency Medicine, University of California, Davis School of Medicine. Sacramento, CA.

Abstract

Objective: To determine the interobserver agreement of history and physical examination findings in children undergoing emergency department (ED) evaluation for headaches.

Study design: We conducted a prospective cross-sectional study of children aged 2 to 17 years evaluated at 3 tertiary-care pediatric EDs for non-traumatic headaches. Two clinicians independently completed a standardized assessment of each child and documented the presence or absence of history and physical examination variables. Unweighted κ statistics were determined for 68 history and 24 physical examination variables.

Results: We analyzed 191 paired observations; median age was 12 years, with 19 (9.9%) children less than 7 years. Interrater reliability was at least moderate ($\kappa \geq 0.41$) for 41 (60.3%) of history variables. Eleven (61.1%) of 18 physical examination variables for which κ statistics could be calculated had a that was at least moderate.

Conclusions: A substantial number of history and physical examination findings demonstrated at least moderate κ statistic values when assessed in children with headaches in the ED. These variables may be generalizable across different types of clinicians for evaluation of children with headaches. If also found to predict the presence or absence of emergent intracranial abnormalities,

Corresponding author: Daniel S. Tsze, MD, MPH. 3959 Broadway, CHN-1-116, New York, NY, 10032. 212-305-9825, dst2141@columbia.edu.

The authors declare no conflicts of interest.

Portions of this study were presented at the Pediatric Academic Societies annual meeting, April 24-May, 2019, Baltimore, Maryland.

the more reliable clinical findings may be helpful in the development of clinical prediction rules or risk stratification models that could be used across settings for children with headaches.

Keywords

Interrater reliability; pediatric; red flag finding; kappa; emergency medicine; emergency department

Headaches are a common chief complaint in children presenting to emergency departments (EDs) in the US, comprising approximately 1% of visits annually.(1–3) Most of these headaches are primary (eg, migraines) or due to conditions such as respiratory infections.(2–4) However, a small but meaningful proportion (0.5–1%) of headaches are secondary to undiagnosed emergent intracranial abnormalities such as brain tumors, intracranial hemorrhages, or strokes.(2,5–8)

ED clinicians use findings obtained on patient history or physical examination to gauge whether a child with a headache is at high or low risk of having an emergent intracranial abnormality, and to decide whether to obtain emergent neuroimaging.(5,9–14) However, in order for specific clinical findings to guide neuroimaging decision-making appropriately, clinicians must agree on their presence or absence. There are few studies examining interobserver agreement of patient history and physical/neurological examination findings in children with headaches. Clinician agreement is particularly important for children, as patient history and physical examination findings may be more difficult to elicit compared with adults due to differences in developmental capacity, language, and cooperation with the physical examination.(15,16) Findings with adequate agreement can be used across different types of clinicians to guide neuroimaging decision-making and to be considered for incorporation into clinical prediction rules or risk stratification models for identification of children with headaches who are at risk of emergent intracranial abnormalities. These rules or models can help optimize the use of ED neuroimaging, including safely reducing unnecessary sedation and/or exposure to radiation from computed tomography (CT), which is the imaging modality of choice in the ED. Therefore, our objective was to determine the interobserver agreement of patient history and physical examination findings in otherwise healthy children undergoing ED evaluations for headaches.

METHODS

We performed a multicenter, prospective cross-sectional study. This study was approved by the institutional review boards of all participating centers. Written informed consent was waived at all sites, and we were required to provide a study information sheet to each patient’s guardian.

Study Setting and Sample

The study was conducted between April 2016 and November 2018 at three urban tertiary-care academic pediatric EDs, with total annual patient visits of approximately 204,000. In accordance with our other prior work, we maintained the same inclusion and exclusion criteria.(5) A convenience sample of children 2 to 17 years old were eligible if they

presented to the ED with a headache as their chief complaint, or if a headache was identified on review of systems and of sufficient concern such that consultation by a neurologist was obtained or recommended; neuroimaging (cranial CT or magnetic resonance imaging [MRI]) was to be completed in the ED or recommended as an outpatient; or) outpatient follow-up for headache evaluation with a neurologist or primary care physician was recommended. Children were excluded if they had: a temperature of ≥38°C documented in the ED or at home; neuroimaging prior to their index ED visit; an abnormal baseline neurological examination; or any history of, structural cranial/intracranial abnormality, intracranial surgery, or other chronic condition considered to be a risk factor for an intracranial abnormality. Risk factors for an intracranial abnormality included but were not limited to: history of, or current, neoplasm or intracranial lesion; intracranial shunt; history of pseudotumor cerebri; sickle cell or collagen vascular disease; immunocompromised state; coagulopathy or currently receipt of anticoagulant medication; or known pregnancy. Patients were also excluded if previously enrolled in the study or with head trauma within the previous 7 days.

Study Protocol

In accordance with our previous work, we employed similar methodology to determine the interobserver agreement of patient history and physical examination findings.(15) Two clinicians independently evaluated each participant. The first clinician (attending physician, fellow, resident, nurse practitioner, or physician assistant) obtained a patient history and performed a physical examination and recorded his or her findings on a standardized case-report form that included all of the clinical findings reported in this study. The second clinician performed an evaluation independent of the first assessor, recording his or her findings on a similar standardized case-report form, masked to the assessment of the first assessor. None of the findings were collected retrospectively (eg, chart review). At least one of the two clinicians was an attending physician or fellow. To minimize the likelihood of changes in the patient's condition affecting the evaluation, we decided a priori that the two clinician assessments had to be performed within 60 minutes of one another. We did not collect information about whether any intervention to treat the headache was given in between the two assessments. Evaluations were conducted prior to knowledge of the results of any ED imaging studies, if performed.

In order to obtain a practical estimate of agreement, there was no standardized training of clinicians regarding the evaluation of the clinical findings. Clinicians were categorized based on their highest level of training (attending, fellow, resident, nurse practitioner, or physician assistant). Attendings included those certified in pediatric emergency medicine or general pediatrics who practiced in the ED. All fellows were engaged in pediatric emergency medicine training. All residents were trainees in pediatrics, emergency medicine, or family medicine.

Measurements

Clinicians assessed specific clinical findings in five categories: headache descriptors, temporal characteristics (ie, related to time or changes over a period of time), symptoms of increased intracranial pressure, headache-related symptoms, and physical examination

findings (which included general physical and neurological examination findings). Clinical findings from the patient history and physical examination were chosen based on a review of the literature and in consultation with two pediatric neurologists with specific expertise in childhood headaches.(5,9–14,17–24) We expected that developmental immaturity would make assessment difficult in young children and, therefore, included an "unsure" choice on the case-report form for appropriate variables.

Statistical Analyses

To make variables more clinically sensible for decision making in the ED, all patient history and physical examination variables were assessed as dichotomous (ie, presence or absence of a finding). We determined the interobserver agreement for each clinical finding by calculating the unweighted Cohen kappa (κ) statistic with two-sided 95% confidence intervals (CIs) as well as the percent agreement. The interobserver agreement was categorized based on the κ point estimates as slight (0–0.2), fair (0.21–0.4), moderate (0.41–0.6), substantial (0.61–8), and almost perfect (0.81–1.0). (25,26) For each variable, we excluded from the κ analysis any paired observations for which data were missing or at least one assessor marked the variable as "unsure."

To evaluate the interobserver agreement of the overall neurological examination, we created three composite neurological examination findings. The first composite neurological examination finding represented the "standard ED neurological examination." This composite was created by surveying 31 ED physicians from across the three participating sites and combining individual variables for which ≥90% of respondents indicated were evaluated "almost all the time" during their usual practice. The five variables that comprised this composite were altered mental status, abnormal extraocular movements, abnormal pupillary reactivity, motor function abnormality, and abnormal gait. The second composite neurological examination finding was the "five key neurological examination elements" (5KNEE), which represent findings that (of which at least one) were previously identified in 98% of children with brain tumors who had headaches.(27,28) The 5KNEE was comprised of papilledema, abnormal extraocular movements, pronator drift, abnormal tandem gait, and abnormal deep tendon reflexes.(27) The third composite neurological examination finding addressed the evaluation of "balance and coordination" and included abnormal gait, abnormal tandem gait, positive Romberg test, and dysmetria/dysdiadochokinesia.(29) The variables of abnormal gait and abnormal tandem gait were included in more than one composite neurological examination finding because of how each composite was defined.

We based our sample size to obtain a two-sided 95% confidence interval band around the κ point estimate of no greater than 0.30. If the true value of κ was 0.4 or 0.6, and using a conservative standard deviation of κ of 1.0, we could obtain this desired confidence interval with a sample size of at least 171 patients.(30–32) Statistical analyses were conducted using SPSS (version 24; IBM Corporation, Armonk, NY).

RESULTS

We enrolled 230 children and analyzed 191; 39 paired assessments were excluded because the assessments were completed more than 60 minutes apart from each other. The median

age of the patients was 12 years (IQR 9, 15), with 19 (9.9%) patients younger than 7 years. The assessor pairs consisted of 65 (34%) attending or fellow physicians matched with another attending or fellow physician; 110 (57.6%) attending or fellow physicians matched with a resident physician; and 11 (5.8%) attending or fellow physicians matched with a nurse practitioner or physician assistant. The assessor pairs were not documented in 5 (2.6%) cases.

The κ statistics for the patient history and physical examination variables are shown in Tables 1–4. For the headache descriptors, temporal characteristics (ie, related to time or changes over a period of time), symptoms of increased intracranial pressure, and headache-related symptoms, 41 (60.3%) had an interobserver agreement that was at least moderate ($\kappa \geq 0.41$). Twelve variables (17.6%) had substantial agreement ($\kappa \geq 0.61$), which included variables such as severe pain intensity of the headache, occipital location of the headache, vomiting awakening patient from sleep, and neck pain. A number of patient history variables that have been associated with emergent intracranial abnormalities in prior studies were prevalent in this cohort, and were identified in as many as 40 to 50% of evaluable patients. (5,9–14,17–24) Percent agreement for patient history variables was at least 80% for most variables (ie, 45 [66.2%] variables).

Of the individual variables and composite neurological examination findings for which κ statistics could be calculated, 11 (61.1%) had κ values that were at least moderate. There were 7 (38.9%) findings with κ values that showed at least substantial reliability, including the Romberg sign, pronator drift, abnormal gait, nystagmus, and abnormal visual fields. Few assessors identified individual abnormal physical or neurological examination findings (including those variables with substantial reliability). The most frequently identified individual abnormal examination findings were ill appearance, neck stiffness, sensory function abnormality, pronator drift, or abnormal speech. Assessors did not identify any cranial nerve or pupillary reactivity abnormalities. Percent agreement was at least 90% for all but one of the 24 physical examination findings.

Of the three composite neurological examination findings, κ was fair for the standard ED neurological examination, moderate for balance and coordination, and substantial for 5KNEE. However, composites with higher κ values were comprised of fewer patients that were fully evaluable. For example, 5KNEE had the highest κ of all three composites, but fewer than half of the total cohort were fully evaluable for 5KNEE, with a designation of “unsure” for papilledema by at least one assessor being the most common reason for unevaluable paired assessment.

DISCUSSION

In this multicenter study of children presenting to the ED with non-traumatic headaches, we found that interobserver agreement was at least moderate for most patient history variables. A substantial number of variables from the physical examination had at least moderate agreement, although abnormal findings were so infrequent for some variables that the κ statistic could not be calculated. The relative frequency of clinical findings with at least moderate reliability suggests that there are many potentially suitable variables that could be

considered for incorporation into a clinical prediction rule or risk stratification model to identify children with headaches who are at risk for emergent intracranial abnormalities.

Our study provides a novel evaluation of interobserver agreement of clinical findings in children with headaches. Prior studies of interobserver agreement in patients with headaches have been in adult populations, with small sample size (between 24 to 109 patients) and conducted by neurologists.(33,34) These studies typically aimed to determine the interobserver agreement of the headache diagnosis (eg, primary headache type) rather than of specific clinical findings.(33–36) Variables evaluated in these prior studies and the current study included vomiting, physical activity that worsens pain, and severe pain intensity; the κ values from the prior studies ranged from 0.62–0.81, 0.24–0.73, and 0.42, respectively.(33–35) Compared with the current study, these κ values were similar to what we noted for vomiting and physical activity that worsens pain (κ = 0.82 and 0.31, respectively), but lower than what we noted for severe pain intensity (κ = 0.69). In one ED-based study, investigators assessed the interobserver agreement of red flag findings to identify which of 60 adults had subarachnoid hemorrhages.(37) Clinical findings found to have substantial agreement included vomiting (κ = 0.79), headache waking from sleep (κ = 0.77), and worst headache of life (κ = 0.64).(37) The interobserver agreement for these variables was similar to what we noted in the current study (κ = 0.82, 0.58, 0.56, respectively), suggesting generalizability in the agreement of certain patient history findings between adults and children. However, it is unknown if this generalizability holds true for all headache-related patient history findings, as some findings are likely more difficult to determine in children due to differences in developmental capacity and language.

Similar to prior studies of interobserver agreement in children presenting with head trauma, seizures, and abdominal trauma, we found that clinicians were unsure of whether some specific findings were present.(15,16,38) This likely reflects the challenge obtaining an accurate history regarding symptoms in young children and the difficulty in examining them. The choice of “unsure” may also reflect a lack of consistent training or education of assessors regarding specific aspects of the physical and neurological examination, such as funduscopic examination. We did not collect information to ascertain further the reason why clinicians at times chose the “unsure” response. Similarly, the 5KNEE and balance and coordination composite neurological examination findings were frequently not evaluable. This was likely due, in part, to their inclusion of more technical and skilled neurological examination assessments (eg, fundoscopy, Romberg sign) that ED physicians might be less likely to perform as part of their usual physical examination.

There was only slight-to-fair interobserver agreement for certain variables typically used to characterize headaches. For example, the κ for the 2 variables meant to elicit sudden onset of headache (ie, to evaluate for presence of subarachnoid hemorrhage) were only 0.23 and 0.19. The standardized language used to elicit certain variables was adapted from studies evaluating adults, and may not have been easily understood by children or may have been difficult to ascertain by parents providing a history on behalf of younger children.(37) Variables with low κ values may need to be re-evaluated to ensure that the language used to elicit the desired findings is readily comprehensible by parents and across the range of pediatric ages.

The evaluability, relative frequency of abnormal clinical findings, and association of clinical findings with important outcomes must be considered when choosing those factors that should influence clinical decisions-making, such as that to obtain neuroimaging in the ED for children with headaches. For example, papilledema was frequently not fully evaluable or assessed but might be useful to determine the need for ED neuroimaging if it uniquely identified patients at high risk for emergent intracranial abnormalities. The composite neurological examination finding 5KNEE was also frequently unable to be assessed because one of its component variables was often not fully evaluable or assessed. However, 5KNEE demonstrated a substantial proportion of abnormal findings (11.7%) in those that were evaluable. If all the variables that comprised 5KNEE could be completed more frequently, this composite could be useful as part of a clinical prediction rule or risk stratification model.

Our study had limitations. First, it was conducted in tertiary-care academic pediatric EDs, with pediatric emergency medicine faculty, which may limit the generalizability of our findings. However, a substantial proportion of the assessors were residents, which may improve generalizability to clinicians at different levels of training. Second, by allowing assessments to be completed up to 60 minutes of one another, there was still the potential that recollection of patient history or physical examination findings could have changed between the time of the 2 assessments. This might be particularly so if treatment for pain was administered between the two assessments. Third, we did not enroll a sufficient number of patients to determine whether there was a difference in agreement based on the amount of time that had elapsed between the two assessments (eg, whether there was a difference in agreement between patients assessed within the first 30 minutes compared with those assessed during the second 30 minutes). Finally, the infrequency of abnormal physical examination findings potentially contributed to low κ values despite high percent agreement, wide 95% confidence intervals, or an inability to calculate a κ value. A future study with a larger sample size would help address this limitation and provide more accurate estimates of κ values for these infrequent abnormal physical examination findings. It is possible that, even in a larger study, infrequent clinical findings such as those seen on neurological examination may need to be assessed together (eg, any cranial nerve, motor, sensory or gait abnormalities) in order to obtain sufficient reliability between assessors. A larger sample size, which would include a greater number of younger children, may also provide more accurate estimates of κ values across ages.

Clinicians evaluating children with headaches in the ED can achieve moderate to substantial agreement for many findings identified on patient history and physical examination. These findings may generalize across different types of clinicians and suitable for evaluating children with headaches. If also found to predict the presence or absence of emergent intracranial abnormalities, the more reliable clinical findings may be helpful in the development of clinical prediction rules or risk stratification models that could be used across settings for children with headaches.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We thank our colleagues who completed the case-report forms and Drs Robert Fryer and Shannon Babineau for their assistance developing the case-report form. R.F. receives funding from Amgen for the OASIS PEDIATRIC [EM] and [CM] clinical trials.

Funded by CTSA from NCATS/NIH (UL1TR000040).

Abbreviations and acronyms:

ED	emergency department
CT	computed tomography
MRI	magnetic resonance imaging
CI	confidence interval
5KNEE	five key neurological examination elements
IQR	interquartile range

REFERENCES

1. Sheridan DC, Meckler GD, Spiro DM, Koch TK, Hansen ML. Diagnostic Testing and Treatment of Pediatric Headache in the Emergency Department. J Pediatr. 2013;163:1634–7. [PubMed: 23968749]
2. Burton LJ, Quinn B, Pratt-Cheney JL, Pourani M. Headache Etiology in a Pediatric Emergency Department. Pediatr Emerg Care. 1997;13:1–4. [PubMed: 9061724]
3. Kan L, Nagelberg J, Maytal J. Headaches in a Pediatric Emergency Department: Etiology, Imaging, and Treatment. Headache. 2000;40:25–9. [PubMed: 10759899]
4. Lewis DW, Qureshi F. Acute Headache in Children and Adolescents Presenting to the Emergency Department. Headache. 2000;40:200–3. [PubMed: 10759922]
5. Tsze DS, Ochs JB, Gonzalez AE, Dayan PS. Red Flag Findings in Children with Headaches: Prevalence and Association with Emergency Department Neuroimaging. Cephalalgia. 2019;39:185–96. [PubMed: 29874930]
6. Richer L, Graham L, Klassen T, Rowe B. Emergency Department Management of Acute Migraine in Children in Canada: a Practice Variation Study. Headache. 2007;47:703–10. [PubMed: 17501852]
7. Conicella E, Raucci U, Vanacore N, Vigevano F, Reale A, Pirozzi N, et al. The Child With Headache in a Pediatric Emergency Department. Headache J Head Face Pain. 2008;48:1005–11.
8. Scagni P, Pagliero R. Headache in an Italian Pediatric Emergency Department. J Headache Pain. 2008;9:83–7. [PubMed: 18250964]
9. Lewis DW, Ashwal S, Dahl G, Dorbad D, Hirtz D, Prensky A, et al. Practice Parameter: Evaluation of Children and Adolescents with Recurrent Headaches: Report of the Quality Standards Subcommittee of the American Academy of Neurology and the Practice Committee of the Child Neurology Society. Neurology. 2002;59:490–8. [PubMed: 12196640]
10. Hayes LL, Palasis S, Bartel TB, Booth TN, Iyer RS, Jones JY, et al. ACR Appropriateness Criteria ® Headache–Child. J Am Coll Radiol. 2018;15:S78–90. [PubMed: 29724429]
11. Lewis DW, Koch T. Headache Evaluation in Children and Adolescents: When to Worry? When to Scan? Pediatr Ann. 2010;39:399–406. [PubMed: 20666345]
12. Gofshteyn JS, Stephenson DJ. Diagnosis and Management of Childhood Headache. Curr Probl Pediatr Adolesc Health Care. 2016;46:36–51. [PubMed: 26750538]
13. Kabbouche MA, Cleves C. Evaluation and Management of Children and Adolescents Presenting with an Acute Setting. Semin Pediatr Neurol. 2010;17:105–8. [PubMed: 20541102]

14. Detsky ME, McDonald DR, Baerlocher MO, Tomlinson GA, McCrory DC, Booth CM. Does this patient with headache have a migraine or need neuroimaging? JAMA. 2006;296:1274–83. [PubMed: 16968852]
15. Dayan PS, Lillis K, Bennett J, Conners G, Bailey P, Callahan J, et al. Interobserver Agreement in the Assessment of Clinical Findings in Children With First Unprovoked Seizures. Pediatrics. 2011;127:e1266–71. [PubMed: 21482615]
16. Gorelick MH, Atabaki SM, Hoyle J, Dayan PS, Holmes JF, Holubkov R, et al. Interobserver Agreement in Assessment of Clinical Variables in Children with Blunt Head Trauma. Acad Emerg Med. 2008;15:812–8. [PubMed: 19244631]
17. Blume HK. Childhood Headache: A Brief Review. Pediatr Ann. 2017;46:e155–65. [PubMed: 28414397]
18. Medina LS, Pinter JD, Zurakowski D, Davis RG, Kuban K, Barnes PD. Children with Headache: Clinical Predictors of Surgical Space-occupying Lesions and the Role of Neuroimaging. Radiology. 1997;202:819–24. [PubMed: 9051039]
19. Lanphear J, Sarnaik S. Presenting Symptoms of Pediatric Brain Tumors Diagnosed in the Emergency Department: Pediatr Emerg Care. 2014;30:77–80. [PubMed: 24457493]
20. Ansell P, Johnston T, Simpson J, Crouch S, Roman E, Picton S. Brain Tumor Signs and Symptoms: Analysis of Primary Health Care Records From the UKCCS. Pediatrics. 2010;125:112–9. [PubMed: 20026498]
21. Reulecke BC, Erker CG, Fiedler BJ, Niederstadt T-U, Kurlemann G. Brain Tumors in Children: Initial Symptoms and Their Influence on the Time Span Between Symptom Onset and Diagnosis. J Child Neurol. 2008;23:178–83. [PubMed: 18160552]
22. Maytal J, Bienkowski RS, Patel M, Eviatar L. The Value of Brain Imaging in Children with Headaches. Pediatrics. 1995;96:413–6. [PubMed: 7651770]
23. Sheridan DC, Waites B, Lezak B, Coryell RJ, Nazemi KJ, Lin AL, et al. Clinical Factors Associated With Pediatric Brain Neoplasms Versus Primary Headache: A Case-Control Analysis. Pediatr Emerg Care. 2017; [Epub ahead of print].
24. Honig PJ. Children With Brain Tumor Headaches: Distinguishing Features. Am J Dis Child. 1982 2 1;136(2):121. [PubMed: 7064925]
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 3;33(1):159–74. [PubMed: 843571]
26. Cicchetti D, Bronen R, Spencer S, Haut S, Berg A, Oliver P, et al. Rating scales, scales of measurement, issues of reliability: resolving some critical issues for clinicians and researchers. J Nerv Ment Dis. 2006 8;194(8):557–64. [PubMed: 16909062]
27. Lewis DW, Rothner AD, Linder SL. Evaluation of Headache In: Winner P, Lewis DW, Rothner AD, editors. Headache in Children and Adolescents. 2nd ed Hamilton, Ontario: BC Decker Inc.; 2008 p. 19–36.
28. Gilles FH, The Childhood Brain Tumor Consortium. The Epidemiology of Headache Among Children with Brain Tumor. J Neurooncol. 1991 2 1;10(1):31–46. [PubMed: 2022972]
29. Bickley LS. Bate's Guide to Physical Examination. 12th ed Szilagyi PG, Hoffman RM, editors. Philadelphia: Wolters Kluwer; 2017.
30. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20(1):37–46.
31. Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of kappa and weighted kappa. Psychol Bull. 1969;72(5):323–7.
32. Flack VF, Afifi A, Lachenbruch P, Schouten H. Sample size determinations for the two rater kappa statistic. Psychometrika. 1988;53(3):321–325.
33. Granella F, D'Alessandro R, Manzoni Gc al, Cerbo R, D'Amato CC, Pini LA, et al. International Headache Society classification: interobserver reliability in the diagnosis of primary headaches. Cephalalgia. 1994;14(1):16–20. [PubMed: 8200017]
34. van Suijlekom JA, de Vet HC, van den Berg SG, Weber WE. Interobserver reliability of diagnostic criteria for cervicogenic headache. Cephalalgia Int J Headache. 1999 11;19(9):817–23.
35. Leone M, Filippini G, D'Amico D, Farinotti M, Bussone G. Assessment of International Headache Society diagnostic criteria: a reliability study. Cephalalgia Int J Headache. 1994 8;14(4):280–4.

36. Friedman BW, Hochberg ML, Esses D, Grosberg B, Corbo J, Toosi B, et al. Applying the International Classification of Headache Disorders to the emergency department: an assessment of reproducibility and the frequency with which a unique diagnosis can be assigned to every acute headache presentation. Ann Emerg Med. 2007 4;49(4):409–19, 419.e1–9. [PubMed: 17210203]

37. Perry JJ, Stiell IG, Sivilotti MLA, Bullard MJ, Hohl CM, Sutherland J, et al. Clinical Decision Rules to Rule Out Subarachnoid Hemorrhage for Acute Headache. JAMA. 2013 9 25;310(12):1248. [PubMed: 24065011]

38. Yen K, Kuppermann N, Lillis K, Monroe D, Borgialli D, Kerrey BT, et al. Interobserver Agreement in the Clinical Assessment of Children With Blunt Abdominal Trauma. Zehtabchi S, editor. Acad Emerg Med. 2013 5;20(5):426–32. [PubMed: 23672355]

Table 1:

Interobserver agreement for headache descriptors and temporal characteristics (n=191)

Finding	Number of evaluable patients (N)[1]	Number of patients for whom at least 1 assessor marked "unsure"	Characteristic present per first assessor, n/N (%)	Percent agreement (95% CI)	κ (95% CI)
Headache Descriptors					
Location of headache[2]					
Frontal	184	7	123/184 (66.8)	78.8 (72,2, 84.5)	0.54 (0.42, 0.67)
Temporal or parietal	184	7	65/184 (35.3)	79.3 (72.8, 85)	0.57 (0.45, 0.69)
Occipital	184	7	38/184 (20.7)	92.4 (87.6, 95.8)	0.77 (0.65, 0.88)
Top of head	184	7	24/184 (13.0)	92.4 (87.6, 95.8)	0.63 (0.45, 0.80)
Laterality/distribution of headache					
Unilateral/focal	175	16	68/175 (38.9)	80.0 (73.3, 85.7)	0.58 (0.45, 0.70)
Bilateral	175	16	81/175 (46.3)	74.9 (67.8, 81.1)	0.49 (0.36, 0.62)
Band around head	175	16	6/175 (3.4)	95.4 (91.2, 98)	0.18 (−0.17, 0.51)
Diffuse/whole head/"all over"	175	16	20/175 (11.4)	90.3 (84.9, 94.2)	0.58 (0.41, 0.76)
Quality of headache[2]					
Pulsating/pounding/ throbbing	190	0	94/190 (74.2)	74.2 (67.4, 80.3)	0.48 (0.36, 0.61)
Pressing/squeezing/ tightening	190	0	47/190 (24.7)	74.2 (67.4, 80.3)	0.29 (0.14, 0.45)
Sharp/stabbing	190	0	37/190 (19.5)	78.9 (72.5, 84.5)	0.30 (0.13, 0.47)
Other	190	0	7/190 (3.7)	96.3 (92.6, 98.5)	0.44 (0.10, 0.79)
Patient cannot describe	190	0	22/190 (11.6)	88.9 (83.6, 93)	0.53 (0.35, 0.70)
Headache reached maximal intensity in…					
… < 1 minute but > 1 second	141	41	27/141 (19.1)	76.6 (68.7, 83.3)	0.23 (0.04, 0.43)
… < 1 second	141	41	10/141 (7.1)	90.8 (84.8, 95)	0.19 (−0.09, 0.46)
Pain intensity at ED presentation					
None (or pain score=0)	176	10	25/176 (14.2)	89.4 (83.7, 93.4)	0.58 (0.41, 0.75)
Mild (or pain score=1–3)	176	10	27/176 (15.3)	85.0 (79.1, 90.1)	0.42 (0.24, 0.60)
Moderate (or pain score=4–6)	176	10	56/176 (31.8)	83.3 (77.2, 88.7)	0.62 (0.49, 0.74)
Severe (or pain score=7–10)	176	10	68/176 (38.6)	85.6 (79.7, 90.6)	0.69 (0.58, 0.80)
Worst headache of patient's life[3]	122	15	43/122 (35.2)	80.0 (72.2, 87)	0.56 (0.41, 0.71)
Duration of headache					
1 day or less	172	14	51/172 (29.7)	85.4 (79.3, 90.4)	0.68 (0.56, 0.80)
2 to 3 days	172	14	39/172 (22.7)	90.1 (84.6, 94.1)	0.70 (0.57, 0.83)
4 to 7 days	172	14	20/172 (11.6)	90.7 (85.3, 94.6)	0.60 (0.42, 0.78)

Finding	Number of evaluable patients (N)[1]	Number of patients for whom at least 1 assessor marked "unsure"	Characteristic present per first assessor, n/N (%)	Percent agreement (95% CI)	κ (95% CI)
>7 days	172	14	62/172 (36.0)	86.6 (80.6, 91.3)	0.71 (0.59, 0.82)
Headache Temporal Characteristics[4]					
First headache of patient's life	186	1	37/186 (19.9)	87.1 (81.4, 91.6)	0.60 (0.46, 0.75)
Increasing headache…[3]					
-frequency	124	12	47/124 (38.8)	79.0 (70.8, 85.8)	0.58 (0.43, 0.72)
-duration	119	18	46/119 (38.7)	74.8 (66, 82.3)	0.48 (0.31, 0.64)
-severity	120	16	54/120 (45.0)	68.3 (59.2, 76.5)	0.36 (0.20, 0.53)
Change in headache…[3]					
-location	127	11	28/127 (22)	79.5 (71.5, 86.2)	0.40 (0.21, 0.59)
-quality/character	117	19	26/117 (22.2)	75.2 (66.4, 82.7)	0.18 (−0.02, 0.39)
Temporal pattern of headaches[3]					
Acute recurrent[5]	95	20	65/95 (68.4)	67.4 (57, 76.6)	0.32 (0.14, 0.51)
Chronic progressive[6]	95	20	11/95 (11.6)	82.1 (72.9, 89.2)	0.38 (0.14, 0.62)
Chronic non-progressive[7]	95	20	14/95 (14.7)	83.2 (74.1, 90.1)	0.29 (0.03, 54.9)
Mixed but not progressive[8]	95	20	5/95 (5.3)	87.3 (79, 93.3)	0.08 (−0.18, 0.34)
Duration of history of headaches[3]					
<1 month	131	0	14/131 (10.7)	84.7 (77.4, 90.4)	0.39 (0.18, 0.60)
1 month to <6 months	131	0	20/131 (15.3)	84.7 (77.4, 90.4)	0.48 (0.28, 0.67)
6 months to <1 year	131	0	20/131 (15.3)	86.3 (79.2, 91.6)	0.37 (0.14, 0.59)
1 year or longer	131	0	77/131 (58.8)	75.6 (67.3, 82.7)	0.51 (0.36, 0.65)

[1] Patients remaining after removing those with "missing" or "unsure" responses

[2] Patients allowed to choose more than one option

[3] Only includes patients who have had more than one headache in their life

[4] Temporal = related to time or changes over a period of time

[5] Episodic headaches with symptom-free intervals[11]

[6] Headaches that gradually increase in frequency and severity over time[11]

[7] Non-progressive daily or near-daily headaches[11]

[8] Mixed pattern of daily headaches with superimposed more intense attacks[11]

Table 2:

Interobserver agreement for symptoms of increased intracranial pressure (n=191)

	Number of evaluable patients (N)[1]	Number of patients for whom at least 1 assessor marked "unsure"	Characteristic present per first assessor, n/N (%)	Percent agreement (95% CI)	κ (95% CI)
Headache wakes from sleep	179	10	72/179 (40.2)	79.9 (73.3, 85.5)	0.58 (0.45, 0.70)
Headache on or soon after waking in the morning	170	15	84/170 (49.4)	69.4 (61.9, 76.2)	0.39 (0.25, 0.53)
Headache precipitated/caused or worsened by lying down	159	30	25/159 (16.4)	85.5 (79.1, 90.6)	0.48 (0.30, 0.66)
Headache precipitated/caused or worsened by coughing, sneezing, straining, or any Valsalva maneuver	144	45	34/144 (23.6)	77.1 (69.3, 83.7)	0.36 (0.18, 0.53)
Headache precipitated/caused by exertion or physical activity	144	45	34/144 (23.6)	77.1 (69.3, 83.7)	0.32 (0.13, 0.50)
Vomiting	188	0	48/188 (25.5)	93.1 (88.5, 96.3)	0.82 (0.73, 0.92)
Vomiting with, or soon after, waking in the morning	181	5	21/181 (11.6)	90.6 (85.4, 94.4)	0.49 (0.28, 0.70)
Vomiting awakens patient from sleep	180	6	11/180 (6.1)	96.1 (92.2, 98.4)	0.65 (0.40, 0.89)

[1] Patients remaining after removing those with "missing" or "unsure" responses

Table 3:

Interobserver agreement for headache-related symptoms (n=191)

Finding	Number of evaluable patients (N)[1]	Number of patients for whom at least 1 assessor marked "unsure"	Characteristic present per first assessor, n/N (%)	Percent agreement (95% CI)	κ (95% CI)
Migraine symptoms[2]					
Nausea	189	0	95/189 (50.3)	80.4 (74, 85.8)	0.61 (0.49, 0.72)
Photophobia	189	0	101/189 (42.1)	75.1 (68.3, 81.1)	0.50 (0.37, 0.62)
Phonophobia	189	0	90/189 (47.6)	75.1 (68.3, 81.1)	0.50 (0.38, 0.62)
Abdominal pain	189	0	34/189 (18.0)	83.1 (76.9, 88.1)	0.43 (0.26, 0.59)
Seeing abnormal patterns (e.g. spots or splotches, zig zag lines, flashes of light)	189	0	37/189 (19.6)	82.0 (75.8, 87.2)	0.43 (0.27, 0.59)
Headache worsens with exertion	189	0	56/189 (29.6)	72.5 (65.5, 78.7)	0.31 (0.17, 0.46)
Headache improves with rest	189	0	88/189 (46.6)	61.3 (54, 68.4)	0.22 (0.08, 0.36)
Family history of migraine headaches	147	21	79/147 (53.7)	83.0 (75.9, 88.7)	0.66 (0.54, 0.78)
Neurological symptoms					
Abnormal movements, non-seizure	189	0	12/189 (6.3)	91.5 (86.6, 95.1)	0.29 (0.04, 0.54)
Seizures	189	0	4/189 (2.1)	97.9 (94.7, 99.4)	0.32 (−0.17, 0.81)
Episodes of confusion	189	0	15/189 (7.9)	90.4 (85.4, 94.3)	0.26 (0.02, 0.50)
Declining school performance	189	0	10/189 (5.3)	93.7 (89.2, 96.7)	0.37 (0.19, 0.65)
Behavior, mood, or personality change	189	0	19/189(10.1)	86.2 (80.5, 90.8)	0.12 (0–0.08, 0.19)
Hearing problems	189	0	4/189 (2.1)	94.2 (89.8, 97.1)	−0.03 (−0.05, −0.01)
Double vision	171	14	16/171 (9.4)	92.4 (87.4, 95.9)	0.51 (0.28, 0.74)
Brief or transient ("a few seconds") episodes of vision loss	180	7	16/180 (3.3)	92.8 (88, 96.1)	0.57 (0.35, 0.78)
Focal motor weakness	189	0	19/189(10.1)	91.0 (86, 94.7)	0.37 (0.14, 0.60)
Sensory changes	189	0	24/189 (12.7)	89.9 (84.7, 93.8)	0.52 (0.33, 0.71)
Unsteadiness	175	10	43/175 (24.6)	79.4 (72.7, 85.2)	0.41 (0.25, 0.57)
Other					
Neck pain	188	1	24/188 (12.8)	92.6 (87.8, 95.9)	0.67 (0.50, 0.83)
Neck stiffness	187	0	6/187 (3.2)	96.7 (93.1, 98.8)	−0.033 (−0.05, −0.01)
Associated upper respiratory infection symptoms	189	0	41/189 (21.7)	83.6 (77.5, 88.6)	0.53 (0.38, 0.68)

[1]Patients remaining after removing those with "missing" or "unsure" response

[2] Interobserver agreement for the variable “vomiting” is described in Table 2.

Table 4:

Interobserver agreement for general physical and neurological examination findings (n=191)

Finding	Number of evaluable patients (N)[1]	Number of patients for whom at least 1 assessor marked "unsure"	Characteristic present per first assessor, n/N (%)	Percent agreement (95% CI)	κ (95% CI)
General physical examination findings					
Not well or ill-appearing general appearance	179	0	19/179 (10.6)	88.8 (83.3, 93)	0.41 (0.20, 0.62)
Neck stiffness	179	0	4/179 (2.2)	98.9 (96, 99.9)	0.66 (0.22, 1)
Neurocutaneous findings	170	11	2/170 (1.2)	98.8 (95.8, 99.9)	Unable to calculate
Individual neurological examination findings					
Altered mental status	182	0	1/182 (0.5)	99.5 (97, 100)	Unable to calculate
Head tilt	170	7	2/170 (1.2)	98.8 (95.8, 99.9)	Unable to calculate
Abnormal speech	180	0	4/180 (2.2)	96.7 (92.9, 98.8)	−0.02 (−0.03, 0)
Cranial nerve abnormality	177	5	0	98.9 (96, 99.9)	Unable to calculate
Extraocular movement abnormal	174	2	2/174 (1.1)	98.9 (95.9, 99.9)	Unable to calculate
Pupillary reactivity abnormal	180	1	0	100 (98, 100)	Unable to calculate
Nystagmus	176	5	1/176 (0.6)	99.4 (96.9, 100)	0.66 (0.05, 1.28)
Papilledema	94	81	3/94 (3.2)	95.7 (89.5, 98.8)	0.48 (0.05, 0.91)
Visual fields abnormal	106	65	3/106 (2.8)	98.1 (93.3, 99.8)	0.66 (0.21, 1.10)
Motor function abnormality	179	2	1/179 (0.6)	98.9 (96, 99.9)	0.50 (−0.10, 1.10)
Sensory function abnormality	174	7	5/174 (2.9)	96.6 (92.6, 98.7)	0.38 (−0.01, 0.77)
Deep tendon reflexes abnormal	146	32	2/146 (1.4)	97.9 (94.1, 99.6)	−0.01 (−0.02, 0.00)
Babinski reflex present	100	69	1/100 (1.0)	98.0 (93, 99.8)	−0.01 (−0.02, 0.00)
Gait abnormal	167	9	3/167 (1.8)	98.8 (95.7, 99.9)	0.66 (0.22, 1.10)
Tandem gait abnormal	132	43	2/132 (1.5)	97.7 (93.5, 99.5)	0.39 (−0.16, 0.94)
Pronator drift	139	39	4/139 (2.9)	98.6 (94.9, 99.8)	0.66 (0.22, 1.10)
Romberg sign present	135	40	3/135 (2.2)	99.3 (95.9, 100)	0.85 (0.57, 1.14)
Dysmetria or dysdiadochokinesia	155	24	2/155 (1.3)	97.4 (93.5, 99.3)	0.32 (−0.17, 0.81)
Composite neurological examination findings					
Standard ED neurological exam[2]	154	13	5/154 (3.2)	96.1 (91.7, 98.6)	0.38 (0, 0.77)
Balance and coordination[3]	115	55	6/115 (5.2)	93.9 (87.9, 97.5)	0.50 (0.18, 0.82)
Five key neurological examination elements (5KNEE)[4]	77	88	6/77 (11.7)	92.2 (83.8, 97.1)	0.62 (0.35, 0.90)

[1]Patients remaining after removing those with "missing" or "unsure" responses

[2]Abnormal mental status, extraocular movement abnormal, pupillary reactivity abnormal, motor function abnormality, and gait abnormal

[3] Gait abnormal, tandem gait abnormal, Romberg test positive, and dysmetria/dysdiadochokinesia present. Most common reasons for unevaluable patients were designations of “unsure” by at least one assessor for tandem gait (n = 43) and Romberg test (n = 40).

[4] Extraocular movements abnormal, papilledema present, deep tendon reflexes abnormal, tandem gait abnormal, and pronator drift present.[20] Most common reason for unevaluable patients was a designation of “unsure” by at least one assessor for papilledema (n = 81).