Academic Emergency Medicine
Official Journal of the Society for Academic Emergency Medicine

BRIEF REPORT

Interobserver Agreement in Retrospective Chart Reviews for Factors Associated With Cervical Spine Injuries in Children

Cody S. Olsen, MS, Nathan Kuppermann, MD, MPH, David M. Jaffe, MD, Kathleen Brown, MD, Lynn Babcock, MD, MS, Prashant V. Mahajan, MD, MBA, MPH, Julie C. Leonard, MD, MPH, for The Pediatric Emergency Care Applied Research Network (PECARN) Cervical Spine Injury Study Group

Abstract

Objectives: The objective was to describe the interobserver agreement between trained chart reviewers and physician reviewers in a multicenter retrospective chart review study of children with cervical spine injuries (CSIs).

Methods: Medical records of children younger than 16 years old with cervical spine radiography from 17 Pediatric Emergency Care Applied Research Network (PECARN) hospitals from years 2000 through 2004 were abstracted by trained reviewers for a study aimed to identify predictors of CSIs in children. Independent physician-reviewers abstracted patient history and clinical findings from a random sample of study patient medical records at each hospital. Interobserver agreement was assessed using percent agreement and the weighted kappa (κ) statistic, with lower 95% confidence intervals.

Results: Moderate or better agreement ($\kappa > 0.4$) was achieved for most candidate CSI predictors, including altered mental status ($\kappa = 0.87$); focal neurologic findings ($\kappa = 0.74$); posterior midline neck tenderness ($\kappa = 0.74$); any neck tenderness ($\kappa = 0.89$); torticollis ($\kappa = 0.79$); complaint of neck pain ($\kappa = 0.83$); history of loss of consciousness ($\kappa = 0.89$); nonambulatory status ($\kappa = 0.74$); and substantial injuries to the head ($\kappa = 0.50$), torso/trunk ($\kappa = 0.48$), and extremities ($\kappa = 0.59$). High-risk mechanisms showed near-perfect agreement (diving, $\kappa = 1.0$; struck by car, $\kappa = 0.93$; other motorized vehicle crash, $\kappa = 0.93$; fall, $\kappa = 0.92$; high-risk motor vehicle collision, $\kappa = 0.89$; hanging, $\kappa = 0.80$). Fair agreement was found for clotheslining mechanisms ($\kappa = 0.36$) and substantial face injuries ($\kappa = 0.40$).

Conclusions: Most retrospectively assessed variables thought to be predictive of CSIs in blunt trauma–injured children had at least moderate interobserver agreement, suggesting that these data are sufficiently valid for use in identifying potential predictors of CSI.

ACADEMIC EMERGENCY MEDICINE 2015;22:487–491 © 2015 by the Society for Academic Emergency Medicine

From the Division of Critical Care, Department of Pediatrics, University of Utah School of Medicine (CSO), Salt Lake City, UT; the University of California, Davis, School of Medicine (NK), CA; the Department of Pediatrics, Washington University and St. Louis Children's Hospital (DMJ, JCL), St. Louis, MO; the Division of Emergency Medicine, Department of Pediatrics, George Washington University School of Medicine (KB), Washington, DC; the Division of Emergency Medicine, Department of Pediatrics, University of Cincinnati and Cincinnati Children's Hospital Medical Center (LB), Cincinnati, OH; Children's Hospital of Michigan and Division of Emergency Medicine, Department of Pediatrics, Wayne State University School of Medicine (PVM), Detroit, MI; and Nationwide Children's Hospital and the Ohio State University (JCL), Columbus, OH.

Received August 19, 2014; revision received November 4, 2014; accepted November 11, 2014.

Pediatric Emergency Care Applied Research Network (PECARN) Cervical Spine Injury Study Group members are listed in Appendix A.

Partially presented at the annual meeting of the Pediatric Academic Societies, Baltimore, MD, May 2009, and the annual meeting of the Society for Academic Emergency Medicine, New Orleans, LA, May 2009.

This work was supported by a grant from the Health Resources and Services Administration/Maternal and Child Health Bureau (HRSA/MCHB) Emergency Medical Services of Children (EMSC) Program (H34 MC04372). PECARN is supported by the HRSA/MCHB/EMSC Program through the following cooperative agreements: U03MC00001, U03MC00003, U03MC00006, U03MC00007, and U03MC00008.

The authors have no relevant financial information or potential conflicts to disclose.

Supervising Editor: Michelle L. Macy, MD, MS.

Address for correspondence and reprints: Cody S. Olsen, MS; e-mail: Cody.Olsen@hsc.utah.edu.

doi: 10.1111/acem.12630
ISSN 1069-6563
PII ISSN 1069-6563583

Cervical spine injuries (CSIs) are serious but rare events in children.[1] A retrospective chart review involving 3,314 children experiencing blunt trauma was undertaken by the Pediatric Emergency Care Applied Research Network (PECARN) and identified potential predictors of pediatric CSI.[2] Although the PECARN CSI study employed methods of performing high-quality medical chart reviews described by Gilbert et al.,[3] retrospective data extraction requires a degree of interpretation of medical records by data abstractors. The results of the PECARN CSI study depend on the validity of the abstracted data. The objective of this planned secondary analysis was to describe interobserver agreement between trained non–physician-reviewers and physician-reviewers, when retrospectively abstracting patient history and physical examination findings for the PECARN CSI study.

METHODS

Study Design

We independently abstracted study data for a sample of patients included in the PECARN CSI study and assessed interobserver agreement for candidate CSI predictors. Institutional review boards for each hospital approved this study.

Study Setting and Population

Details of the main study design were described previously.[2] Briefly, children (<16 years old) presenting to 17 PECARN hospitals between January 1, 2000, and December 31, 2004, who underwent cervical spine imaging were identified retrospectively through query of hospital billing databases. Cases were confirmed as CSI and control patients were confirmed to be free of CSI.

Study Protocol

Primary reviewers were clinical research coordinators trained in medical record abstraction. The work of primary reviewers at each hospital was overseen by a trained physician study investigator who verified chart abstraction of clinical data. Each hospital may have had multiple primary reviewers. Reviewers followed a hospital-specific source hierarchy to abstract each variable from existing electronic and paper-based prehospital documentation, transferring hospital medical records, emergency department (ED) medical records, and hospital admission records.

Secondary reviewers used the same procedures to independently abstract a subset of data, including variables to be used individually, or in composite, as candidate predictors of CSI. The secondary reviewer was a physician study investigator from a different hospital (DMJ, KB, LB, PM, JCL) with access to the same data sources as the primary reviewers during in-person data abstraction sessions. Secondary reviewers abstracted data from multiple hospitals and were familiar with those hospitals' medical record systems.

Secondary reviewer data were kept separate from primary reviewer data and were not used in any comparison or analysis besides for this study. The PECARN data coordinating center worked with primary reviewers to resolve data discrepancies within patient records, but did not compare data between primary and secondary reviews.

CSI Predictors. CSI predictors derived from abstracted data are described previously and included patient history findings, injury mechanisms, and physical examination findings.[2] "Not applicable" options (including pre-/nonverbal and pre-/nonambulatory) were grouped with not-present findings. Patients with missing data were not included in predictor-level comparisons.

Sample Size and Sampling Method

We estimated that 359 patients would be required to demonstrate at least a moderate kappa statistic (95% lower confidence limit (LCL) > 0.40) for a predictor, assuming kappa = 0.50 and a prevalence of 10%.[4,5] Near the end of the PECARN CSI study's data collection phase, 10% of cases and controls were randomly sampled from each hospital's available study patients. When 10% represented less than 10 patients, 10 were randomly selected. This resulted in a sample of 365 patients.

Data Analysis

We compared our patient sample to patients in the PECARN CSI study using means and relative frequencies. We described the prevalence of each candidate CSI predictor according to the primary and secondary reviewers and the percentage of patients for whom reviewers agreed. Interobserver agreement was assessed using the Fleiss-Cohen weighted kappa (κ) statistic with lower confidence limits estimated using normal approximation methods.[4] Interpretation of the kappa statistic is explained by several sources.[4–6] The kappa statistic ranges from –1 to 1, with a positive kappa indicating agreement better than expected by chance.[6] In accordance with suggested kappa interpretation guidelines, we considered agreement fair if $0.2 < \kappa \leq 0.4$, moderate if $0.4 < \kappa \leq 0.6$, substantial if $0.6 < \kappa \leq 0.8$, and near perfect if $0.8 < \kappa \leq 1.0$.[5,6] Data analyses were performed using SAS software (version 9.3).

RESULTS

Characteristics of the Study Patients

Information was reabstracted for 365 (11%) of the 3,314 PECARN CSI study patients. Characteristics of our sample were similar to the full sample, with mean (±SD) age of 10.1 (±4.7) years (10.0 [±4.7] years in the full sample); 59% male (61%), 47% white (49%), and 18% with CSIs (16%).

Interobserver Agreement

Interobserver agreement is summarized in Table 1. Of the 25 predictors, 23 (92%) had moderate or better agreement, of which nine (36%) had near-perfect agreement. Two remaining predictors had fair agreement, including substantial face injury ($\kappa = 0.40$) and clotheslining mechanism ($\kappa = 0.36$). Clotheslining mechanism was identified in only nine children, on whom two reviewers agreed.

Table 1
Interobserver Agreement Results for Candidate Predictors of Cervical Spine Injury*

Predictor	Prevalence, %		Percent Agreement	Kappa (95% LCL)
	Primary Reviewer (n = 365)	Secondary Reviewer (n = 365)		
Altered mental status†	20	20	96	0.87 (0.82)
GCS	76, GCS = 15	76, GCS = 15	92	0.69 (0.46)
AVPU	83, AVPU = A	81, AVPU = A	90	0.68 (0.56)
Focal neurologic findings‡	12	13	94	0.74 (0.64)
Paresthesias	5	6	96	0.68 (0.54)
Sensory loss	4	4	96	0.50 (0.30)
Motor weakness	6	5	96	0.61 (0.45)
Other neurologic deficit	3	3	98	0.66 (0.46)
Neck tenderness	37	38	95	0.89 (0.85)
Posterior/midline neck tenderness§	28	27	90	0.74 (0.68)
Torticollis or pain with neck movement	7	6	97	0.79 (0.67)
Complaint of neck pain	38	36	92	0.83 (0.77)
History of loss of consciousness	37	38	95	0.89 (0.84)
History of ambulation	23	24	91	0.74 (0.67)
Substantial head injury‖	14	13	89	0.50 (0.39)
Substantial face injury‖	6	5	94	0.40 (0.23)
Substantial torso or trunk injury‖	5	7	94	0.48 (0.32)
Substantial extremity injury‖	8	8	94	0.59 (0.46)
High risk fall¶	5	6	99	0.92 (0.84)
High risk motor vehicle crash**	16	14	98	0.89 (0.83)
Other motor vehicle crash††	2	2	100	0.93 (0.82)
Struck by car	17	16	98	0.93 (0.89)
Diving mechanism	1	1	100	1.00 (1.00)
Hanging mechanism	1	1	100	0.80 (0.47)
Clotheslining mechanism‡‡	2	1	98	0.36 (0.05)

AVPU = alert, voice, pain, unresponsive; GCS = Glasgow Coma Scale; LCL = lower confidence limit.
*Prevalence according to the primary and secondary reviewers, percent agreement, and the Fleiss-Cohen weighted kappa with 95% LCLs shown for each candidate predictor.
†GCS < 15, AVPU scale less than alert, evidence of intoxication, or mental status descriptions deemed by consensus panel to represent altered level of consciousness.
‡Paresthesias, loss of sensation, motor weakness, or other neurological finding deemed consistent with spine injury by consensus panel (e.g., priapism).
§Physical examination notes neck tenderness as posterior, midline, or at a designated cervical level or a descriptor that consensus panel deemed consistent with posterior midline neck tenderness.
‖Observable injuries that are life-threatening, warrant surgical intervention, or inpatient observation.
¶Fall from a height ≥ 10 feet.
**Head-on collision, rollover, ejected from the vehicle, death in the same crash, or speed > 55 mph.
††Nonautomobile motor vehicle crash (e.g., motorcycle).
‡‡Injury resulting from rope, cable, or similar item exerting traction on the neck while the child is in motion.

Percent agreement ranged from 89% (substantial head injury) to 100% (diving mechanism). The prevalence of each predictor was similar between reviewers. Some predictors were rare, resulting in reduced κ when reviewers did not agree perfectly (clotheslining 2% in primary sample, other motor vehicle crash mechanism 2%, and hanging mechanism 1%).

When minor injuries were combined with substantial injuries, the prevalence increased, percent agreement decreased slightly, and κ improved compared to substantial injuries alone. Kappa estimates for minor or substantial injuries to body regions were as follows: head (κ = 0.60, LCL = 0.53), face (κ = 0.70, LCL = 0.64), torso/trunk (κ = 0.61, LCL = 0.51), and extremity (κ = 0.70, LCL = 0.64) injuries.

Rates of missing data according to either the primary or the secondary review varied: Glasgow Coma Scale (GCS) score (missing 28%), focal neurologic findings components (15%), and motor vehicle crash details (12%). All other predictors were missing less than 10%.

DISCUSSION

Our interobserver agreement analysis demonstrated moderate to near-perfect agreement for almost all patient history and physical examination findings used as candidate CSI predictors in the PECARN CSI study. The high rates of agreement demonstrated in this study, despite differences in training and experience between primary and secondary reviewers, support the validity of these variables and their use as potential components of a pediatric CSI prediction tool.

Of 25 predictors analyzed, 92% had moderate or better interobserver agreement, with 36% having near-perfect agreement. Composite variables showed higher agreement than their components likely because

reviewers had more opportunities to codify presence of the condition and therefore increase the prevalence of the finding. Increasing the prevalence of a rare finding leads to an increased kappa when reviewers agree.[6] The reverse is also true, as exemplified by clotheslining, for which kappa was less than moderate due to infrequency, despite 98% agreement.

Consistent classification of variables by reviewers suggests that the variables can be objectively abstracted on retrospective chart review. Injury mechanism, patient history, neck examination, and neurologic examination variables were consistently classified on chart review for this study and are likely to be objectively ascertained on chart review by future investigators using similar methods. Further, although secondary reviewers were physicians with a higher level of training and clinical expertise, agreement for most variables was high. This suggests that these candidate CSI predictors may be identified consistently by a diverse community of clinicians.

Our findings are consistent with prospective investigations of interobserver agreement in trauma populations. The NEXUS study prospectively enrolled children and adults at risk of CSI and showed altered neurologic function $\kappa = 0.58$, midline neck tenderness $\kappa = 0.77$, and distracting injury $\kappa = 0.77$.[7] A prospective study of traumatic brain injuries found high agreement for injury mechanism and history of loss of consciousness and moderate agreement for altered mental status.[8] A prospective study of blunt abdominal trauma found almost perfect agreement for injury mechanism and moderate agreement for painful injury, difficulty breathing, and tenderness.[9] A study of adults with dystonia showed moderate agreement ($\kappa = 0.52$) for cervical dystonia.[10]

The agreement of retrospectively obtained data is fundamentally different from agreement between prospective observers due to bias introduced by the abstractor's interpretation of the medical record. Combining results from previous prospective reliability analyses with results from this study suggests that injury mechanism and composite variables of altered mental status, neurological deficits, and painful distracting injuries are reliable. Findings such as torticollis, history of ambulation, and history of loss of consciousness may also be considered reliable due to their substantial agreement in this study.

LIMITATIONS

Due to the retrospective nature of the study, we were unable to calculate rates of misclassification for chart abstraction relative to prospective evaluation. Some data were missing according to one or both reviewers, which resulted in some loss of precision. Secondary reviewers may have been less familiar with hospital data sources compared to primary reviewers, and intrareviewer variability may have been present. These issues, however, would bias results toward worse interobserver agreement. Two variables used as candidate predictors in the PECARN CSI study were not reabstracted due to a high amount of missing information (axial load mechanism) and low prevalence (predisposing conditions) in the primary sample.

CONCLUSIONS

Most candidate cervical spine injury predictors included in the Pediatric Emergency Care Applied Research Network cervical spine injury study had moderate to near-perfect interobserver agreement and were consistently abstracted from medical records. This suggests that these data are sufficiently valid for use in identifying potential predictors of cervical spine injury.

We thank the site PIs and research coordinators in PECARN, whose dedication and hard work made this study possible. We also thank Jeffrey R. Leonard, MD, for his expertise in reviewing cases of cervical spine injury.

References

1. Viccellio P, Simon H, Pressman BD, Shah MN, Mower WR, Hoffman JR. A prospective multicenter study of cervical spine injury in children. Pediatrics 2001;108:E20.
2. Leonard JC, Kuppermann N, Olsen C, et al. Factors associated with cervical spine injury in children after blunt trauma. Ann Emerg Med 2011;58:145–55.
3. Gilbert EH, Lowenstein SR, Koziol-McLain J, Barta DC, Steiner J. Chart reviews in emergency medicine research: where are the methods? Ann Emerg Med 1996;27:305–8.
4. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Meas 1973;33:613–9.
5. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
6. Gisev N, Bell JS, Chen TF. Interrater agreement and interrater reliability: key concepts, approaches, and applications. Res Soc Adm Pharm 2013;9:330–8.
7. Mahadevan S, Mower WR, Hoffman JR, Peeples N, Goldberg W, Sonner R. Interrater reliability of cervical spine injury criteria in patients with blunt trauma. Ann Emerg Med 1998;31:197–201.
8. Gorelick MH, Atabaki SM, Hoyle J, et al. Interobserver agreement in assessment of clinical variables in children with blunt head trauma. Acad Emerg Med 2008;15:812–8.
9. Yen K, Kuppermann N, Lillis K, et al. Interobserver agreement in the clinical assessment of children with blunt abdominal trauma. Acad Emerg Med 2013;20:426–32.
10. Logroscino G, Livrea P, Anaclerio D, et al. Agreement among neurologists on the clinical diagnosis of dystonia at different body sites. J Neurol Neurosurg Psychiatry 2003;74:348–50.

APPENDIX A

Pediatric Emergency Care Applied Research Network (PECARN) Cervical Spine Injury Study Group Members

Participating Centers and Investigators (Listed in Alphabetical Order). Boston Children's Hospital, Boston, MA—Lise E. Nigrovic, MD, MPH; State University of New York, Buffalo, Buffalo, NY—Kathleen Lillis, MD;

Children's Hospital of Michigan, Detroit, MI—Curt Stankovic, MD; Children's Hospital of Philadelphia, Philadelphia, PA—Aaron Donoghue, MD, MSCE; Children's National Medical Center, Washington, DC—Kathleen Brown, MD; Cincinnati Children's Hospital Medical Center, Cincinnati, OH—Scott D. Reeves, MD; DeVos Children's Hospital/Spectrum Health, Grand Rapids, MI—John D. Hoyle, Jr., MD; Hurley Medical Center, Flint, MI—Dominic Borgialli, DO, MPH; Johns Hopkins Medical Center, Baltimore, MD—Jennifer Anders, MD; Medical College of Wisconsin and Children's Hospital of Wisconsin, Milwaukee, WI—Greg Rebella, MD; Chicago Memorial/Northwestern, Chicago, IL—Elizabeth Powell, MD, MPH; Primary Children's Medical Center, Salt Lake City, UT—Kathleen Adelgais, MD; UC Davis Medical Center, Sacramento, CA—Nathan Kuppermann, MD, MPH; University of Michigan, Ann Arbor, MI—Alexander J. Rogers, MD; University of Rochester Medical Center, Rochester, NY—Lynn Babcock, MD, MS; University of Maryland, Baltimore, MD—Getachew Teshome, MD, MPH; Washington University and St. Louis Children's Hospital, St. Louis, MO—David M. Jaffe, MD, Jeffrey R. Leonard, MD, and Julie C. Leonard, MD, MPH.

PECARN Steering Committee: N. Kuppermann, Chair; E. Alpern, D. Borgialli, K. Brown, J. Chamberlain, J. M. Dean, G. Foltin, M. Gerardi, M. Gorelick, J. Hoyle, D. Jaffe, C. Johns, K. Lillis, P. Mahajan, R. Maio, S. Miller*, D. Monroe, R. Ruddy, R. Stanley, M. Tunik, and A. Walker.

MCHB/EMSC Liaisons: D. Kavanaugh and H. Park.

Central Data Management and Coordinating Center: M. Dean, R. Holubkov, S. Knight, A. Donaldson, and S. Zuspan.

Feasibility and Budget Subcommittee (FABS): T. Singh, Chair; A. Drongowski, L. Fukushima, M. Shults, J. Suhajda, M. Tunik, and S. Zuspan.

Grants and Publications Subcommittee (GAPS): M. Gorelick, Chair; E. Alpern, G. Foltin, R. Holubkov, J. Joseph, S. Miller*, F. Moler, O. Soldes, and S. Teach.

Protocol Concept Review and Development Subcommittee (PCRADS)—D. Jaffe, Chair; A. Cooper, J. M. Dean, C. Johns, R. Kanter, R. Maio, N.C. Mann, D. Monroe, K. Shaw, and D. Treloar.

Quality Assurance Subcommittee (QAS): R. Stanley, Chair; D. Alexander, J. Burr, M. Gerardi, R. Holubkov, K. Lillis, R. Ruddy, M. Shults, and A. Walker.

Safety and Regulatory Affairs Subcommittee (SRAS): W. Schalick, Chair; J. Brennan, J. Burr, J.M. Dean, J. Hoyle, R. Ruddy, T. Singh, D. Snowdon, and J. Wright.

*Deceased.