**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 (κ > 0.4) was achieved for most candidate CSI predictors, including altered mental status (κ = 0.87); focal neurologic findings (κ = 0.74); posterior midline neck tenderness (κ = 0.74); any neck tenderness (κ = 0.89); torticollis (κ = 0.79); complaint of neck pain (κ = 0.83); history of loss of consciousness (κ = 0.89); nonambulatory status (κ = 0.74); and substantial injuries to the head (κ = 0.50), torso/trunk (κ = 0.48), and extremities (κ = 0.59). High-risk mechanisms showed near-perfect agreement (diving, κ = 1.0; struck by car, κ = 0.93; other motorized vehicle crash, κ = 0.93; fall, κ = 0.92; high-risk motor vehicle collision, κ = 0.89; hanging, κ = 0.80). Fair agreement was found for clotheslining mechanisms (κ = 0.36) and substantial face injuries (κ = 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.
Cervical spine injuries (CSIs) are serious but rare events in children. 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. Although the PECARN CSI study employed methods of performing high-quality medical chart reviews described by Gilbert et al., 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. 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. “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%. 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 ($\kappa$) statistic with lower confidence limits estimated using normal approximation methods. Interpretation of the kappa statistic is explained by several sources. The kappa statistic ranges from −1 to 1, with a positive kappa indicating agreement better than expected by chance. 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 $0.8 < \kappa \leq 1.0$. 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.

488 Olsen et al. • INTEROBSERVER AGREEMENT FOR PEDIATRIC CSI
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 $\kappa$ 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 $\kappa$ improved compared to substantial injuries alone. Kappa estimates for minor or substantial injuries to body regions were as follows: head ($\kappa = 0.60$, LCL = 0.53), face ($\kappa = 0.70$, LCL = 0.64), torso/trunk ($\kappa = 0.61$, LCL = 0.51), and extremity ($\kappa = 0.70$, 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 (95%), 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. 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 \( \kappa = 0.77 \). 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. A prospective study of blunt abdominal trauma found almost perfect agreement for injury mechanism 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. 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


APPENDIX A

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

Participating Centers and Investigators (Listed in Alphabetic 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.


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


*Deceased.