

## BRIEF REPORT

# Completeness of cause of injury coding in healthcare administrative databases in the United States, 2001

J H Coben, C A Steiner, M Barrett, C T Merrill, D Adamson

*Injury Prevention* 2006;12:199–201. doi: 10.1136/ip.2005.010512

**Objectives:** To determine the completeness of external cause of injury coding (E-coding) within healthcare administrative databases in the United States and to identify factors that contribute to variations in E-code reporting across states.

**Design:** Cross sectional analysis of the 2001 Healthcare Cost and Utilization Project (HCUP), including 33 State Inpatient Databases (SID), a Nationwide Inpatient Sample (NIS), and nine State Emergency Department Databases (SEDD). To assess state reporting practices, structured telephone interviews were conducted with the data organizations that participate in HCUP.

**Results:** The percent of injury records with an injury E-code was 86% in HCUP's nationally representative database, the NIS. For the 33 states represented in the SID, completeness averaged 87%, with more than half of the states reporting E-codes on at least 90% of injuries. In the nine states also represented in the SEDD, completeness averaged 93%. Twenty two states had mandates for E-code reporting, but only eight had provisions for enforcing the mandates. These eight states had the highest rates of E-code completeness.

**Conclusions:** E-code reporting in administrative databases is relatively complete, but there is significant variation in completeness across the states. States with mandates for the collection of E-codes and with a mechanism to enforce those mandates had the highest rates of E-code reporting. Nine statewide ED data systems demonstrate consistently high E-coding completeness.

External cause of injury codes (E-codes) are an integral component of injury research efforts because they describe the mechanism and intent of the injury. E-coded hospital discharge data systems are potentially one of the most effective and efficient means available to collect data needed to prevent and control injuries<sup>1</sup> and they have been used extensively in the United States and several other countries, including Australia,<sup>2</sup> New Zealand,<sup>3</sup> Canada,<sup>4</sup> and Hong Kong.<sup>5</sup> Concerns about E-coding in hospital discharge data emerged from a 1997 survey that revealed wide variation in rules and practices for the collection of state-level E codes.<sup>6</sup> An update of this study, published in 2005, found that despite evidence of overall improvement, wide variations in state E-coding practices continued to exist within the United States.<sup>7</sup>

This paper examines the completeness of E-coding within administrative databases in the United States, including inpatient and emergency department (ED) data. The objectives were to: (1) determine the completeness of E-code information at the national level, (2) determine the completeness of E-code information at the state level, and (3) identify factors that account for variations in E-code reporting across states.

## METHODS

### Data source

We examined the largest collection of longitudinal, all-payer, encounter-level healthcare data available—the Healthcare Cost and Utilization Project (HCUP) databases.<sup>8</sup> The data in HCUP are derived from discharge summaries and abstracts created by hospitals for billing and payment purposes. In 45 states, hospital discharge data systems now exist based upon hospitals providing these discharge summaries to the state government, a hospital association, or similar data organizations.<sup>7</sup>

The HCUP is built through a partnership between the state-level data organizations and the Agency for Healthcare Research and Quality (AHRQ).<sup>9</sup> The state data organizations provide their unique statewide database to HCUP. The data are then subjected to internal consistency and edit checks. Data elements that are similar across states are recoded into a uniform coding scheme. These uniformly formatted data sets are the core of the HCUP databases.<sup>10</sup> Each state inpatient database (SID) contains the universe of that state's community hospital inpatient discharge records. Using a stratified probability sample of hospitals included in the SID, AHRQ produces the Nationwide Inpatient Sample (NIS), which is designed to approximate a 20% sample of all community hospitals in the United States.<sup>10</sup>

A comprehensive analysis of E-coding within HCUP was undertaken, and the results have been published elsewhere (<http://www.hcup-us.ahrq.gov/reports/methods.jsp>). That report included HCUP inpatient and outpatient data and separately examined E-coding for injury diagnoses and adverse reactions/medical misadventures. In this paper we summarize our findings relating to injury E-coding in three HCUP databases of greatest interest to injury control researchers and practitioners:

- The 2001 Nationwide Inpatient Sample (NIS): A database that contains a nationally stratified sample of hospitals from 33 states with a census of inpatient discharges from sampled hospitals.<sup>8</sup>
- The 2001 Statewide Inpatient Databases (SID): State-specific databases that contain all inpatient discharges from hospitals in 33 participating states.<sup>8</sup>
- The 2001 State Emergency Department Databases (SEDD): State-specific databases containing all ED encounters that do not result in an admission from nine participating states.<sup>8</sup>

**Abbreviations:** AHRQ, Agency for Healthcare Research and Quality; CSTE, Council of State and Territorial Epidemiologists; HCUP, Healthcare Cost and Utilization Project; NHAMCS, National Hospital Ambulatory Medical Care Survey; NIS, Nationwide Inpatient Sample; SEDD, State Emergency Department Databases; SID, State Inpatient Databases.

**Table 1** E-code completeness rates in HCUP and comparable databases

Hospital setting	Data source	Percentage of records with injury diagnosis	Percentage of injury records with E-code
Inpatient	HCUP NIS	4.6%	85.7%
	HCUP SID	4.1% to 6%	52.6%–99.4% (avg = 87.2%)
	Comparative data source: NHDS	5.0%	68.0%
Emergency department	HCUP SEDD	26.4% to 34.8%	71.9%–99.4% (avg = 92.5%)
	Comparative data source: NHAMCS	N/A	80.8%

### Comparative data sources

For comparison with the NIS and SID, we examined the 2001 National Hospital Discharge Survey (NHDS), a probability sample of inpatient hospital records acquired from a national sample of about 500 hospitals.<sup>11</sup> For comparison with the SEDD, we used two sources. The first was the National Hospital Ambulatory Medical Care Survey (NHAMCS): 2001 Emergency Department Summary, a national probability sample of visits to hospital EDs.<sup>12</sup>

### Measures

Injury related hospitalizations were identified by a principal ICD-9-CM diagnosis code in the range of 800–995: 800–909.2, 909.4, 909.9, 910–994.9, 995.5–995.59, and 995.80–995.85. We then calculated the proportion of those records that contained an injury related E-code in any of the secondary diagnosis fields, in accordance with previously published recommendations.<sup>13</sup>

To supplement our understanding of why E-code completeness may vary by state, we considered five state-specific factors that may affect completeness rates, including: state mandates for inclusion of E-codes on injury records; state policies for enforcing those mandates; the presence of additional diagnosis fields on state reporting forms; the recording of E-codes separately from other secondary diagnoses; and verification of the presence of E-codes on injury related records. These data were collected via structured telephone interviews with representatives from 32 of the 33 statewide data organizations that participate in HCUP.

The association of each of these reporting practices to E-code completeness was examined using logistic regression and the estimated odds ratio.

## RESULTS

### E-code reporting was high in the HCUP Inpatient and ED databases

Across the inpatient databases included in this study, injuries accounted for approximately 5% of all US discharges. This was similar to the proportion of cases identified as injuries in the NHDS. Within the NIS, the percent of those injury records with a corresponding E-code was 86%. For the 33 states represented in the SID, we found that completeness

varied from 53% to 99%, with more than half of the states reporting E-codes on at least 90% of injuries. On average, completeness was 87% across states. In comparison, E-code completeness in the NHDS was found to be 68% (table 1).

E-code reporting in the ED setting was somewhat higher. Among the nine states that provided ED data, E-coding ranged from 72% to 99%, with an average completeness of 93%. Seven of these nine states reported E-codes on at least 90% of injuries. In comparison, E-code completeness in the NHAMCS (81%) was slightly lower than in the majority of states in the HCUP (table 1).

### State mandates for E-code collection affect E-code completion rates

We examined several state-specific procedures that might affect completeness rates. The results indicate that all five factors are positively associated with E-code completeness, with two factors—the existence of mandates requiring E-code submission and whether state agencies enforce the mandates—displaying the strongest association (table 2).

Of the 32 states, 22 have mandates or regulations for E-code submission on injury records. States with mandates reported E-codes on at least 94% of their ED injuries. In contrast, states without mandates reported E-codes on 71.9% to 80.1% of their ED injuries.

Of the 22 states with mandates, eight also reported formal or informal mechanisms for enforcing them. In these eight states, the E-code completeness averaged 97.0%. In the 14 states with mandates but no enforcement mechanism, E-code completeness averaged 89.4%.

## DISCUSSION

Our findings support the use of administrative healthcare data for injury research. These data are routinely collected, population based, capable of identifying the mechanism and intent of injury events, and identify injuries serious enough to warrant ED treatment or inpatient hospitalization. Completeness of E-code reporting on inpatient and ED injury records was relatively high in the 2001 HCUP administrative databases.

Our findings on state variations in E-coding practices are congruent with those in the recently released report issued by the Council of State and Territorial Epidemiologists (CSTE).

**Table 2** State mandates associated with E-coding completeness

Factors	Conclusion about association*	p Value	Odds ratio (95% CI)
Number of diagnosis fields available	+	<0.0001	1.06 (1.06–1.06)
Dedicated E-code field available	+	<0.0001	1.29 (1.27–1.30)
State mandates for E-code collection	+	<0.0001	1.71 (1.64–1.77)
Enforceability of state mandates	+	<0.0001	4.98 (4.76–5.22)
E-code edit checks performed	+	0.0032	1.09 (1.03–1.15)

\*+ = positive association between percent of injury records with E-codes reported and factors (statistically significant at  $p < 0.05$ ).

Comparable to our findings, the CSTE survey found that states with such mandates had higher rates of E-code completeness.<sup>7</sup> Both studies suggest that improving E-code reporting will depend on states adopting mandatory reporting requirements. Our study also found that the overall number of diagnosis fields, the presence of a dedicated E-code diagnosis field, the performance of routine edit checks, and the availability of mechanisms to enforce E-code mandates were all associated with more complete E-coding.

This study focused on the essential first step of understanding the completeness of E-coding for injuries reported in administrative healthcare data. Although the focus of this work was on E-code completeness, other investigators in the United States and New Zealand have independently reported on the accuracy of E-codes in hospital discharge records, finding a range of incorrect coding between 13% and 18%.<sup>3-14</sup> These findings suggest that investigators in the United States and other countries using ICD-9 coding of administrative data can have a relatively high degree of confidence in the E-codes contained therein, particularly in jurisdictions that mandate and enforce E-code reporting. Countries considering the implementation of E-coded hospital discharge data systems should also consider the other factors we have found to be associated with E-code completeness as important adjuncts to mandatory reporting. In the United States, efforts should be made to strengthen E-coding in states with current deficiencies and to increase the number of states regularly collecting and reporting ED data.

#### ACKNOWLEDGEMENTS

The authors gratefully acknowledge support for this study from the statewide data organizations participating in HCUP. The following 33 states provided inpatient data for 2001: Arizona, California, Colorado, Connecticut\*, Florida, Georgia, Hawaii, Illinois, Iowa, Kansas, Kentucky, Maine\*, Maryland\*, Massachusetts, Michigan, Minnesota\*, Missouri\*, Nebraska\*, New Jersey, New York, North Carolina, Oregon, Pennsylvania, Rhode Island, South Carolina\*, Tennessee\*, Texas, Utah\*, Vermont, Virginia, Washington, West Virginia, and Wisconsin. The nine states marked with an asterisk also provided emergency department data. This work was funded by the Agency for Healthcare Research and Quality.

#### Authors' affiliations

**J H Coben**, Injury Control Research Center, West Virginia University, Morgantown, WV, USA

**C A Steiner**, Agency for Healthcare Research and Quality, Center for Delivery, Organization, and Markets, Rockville, MD, USA

**M Barrett**, M.L. Barrett, Inc, San Diego, CA, USA

**C T Merrill, D Adamson**, Medstat, Washington, DC, USA

Competing interests: none declared.

Correspondence to: Dr J H Coben, Injury Control Research Center, West Virginia University, PO Box 9151, Morgantown, WV 26506-9151, USA; [jcoben@hsc.wvu.edu](mailto:jcoben@hsc.wvu.edu)

Accepted 9 March 2006

#### REFERENCES

- 1 **Sniezek JE**, Finklea F, Graitcer PL. Injury coding and hospital discharge data. *JAMA* 1989;**262**:2270-2.
- 2 **Finch C**, Cassel E, Stathakis V. The epidemiology of sport and active recreation injury in the Latrobe Valley. Report #151. Victoria, Australia: Monash University Accident Research Centre, 1999.
- 3 **Langley J**, Stephenson S, Thorpe C, et al. Accuracy of injury coding under ICD-9 for New Zealand public hospital discharges. *Inj Prev* 2006;**12**:58-61.
- 4 **Rhodes AE**, Links PS, Streiner DL, et al. Do hospital E-codes consistently capture suicidal behaviour? *Chronic Dis Can* 2002;**23**:139-45.
- 5 **Chan CC**, Cheng JCY, Wong TW, et al. An international comparison of childhood injuries in Hong Kong. *Inj Prev* 2000;**6**:20-3.
- 6 **American Public Health Association**. *How states are collecting and using cause of injury data: a survey on state-based injury surveillance, external cause of injury coding practices, and coding guidelines in the 50 states, the district of Columbia, and Puerto Rico*, A Report of the Data Committee, Injury Control and Emergency Health Services Section. San Francisco, CA: September, 1998.
- 7 **Council of State and Territorial Epidemiologists**. *How states are collecting and using cause of injury data: 2004 update to the 1997 report*, A Report by the Council of State and Territorial Epidemiologists; Data Committee, Injury Control and Emergency Health Services Section. Atlanta, GA: March, 2005. Available at <http://www.cste.org> (accessed March 2006).
- 8 **Agency for Healthcare Research and Quality**. *Healthcare Cost and Utilization Project*. Available at <http://www.hcup-us.ahrq.gov/home.jsp> (accessed March 2005).
- 9 **Agency for Healthcare Research and Quality**. *Healthcare cost and utilization project, 1988-2002*. Available at <http://www.hcup-us.ahrq.gov/overview.jsp> (accessed January 2005).
- 10 **Steiner C**, Elixhauser A, Schnaier J. The healthcare cost and utilization project: an overview. *Effective Clinical Practice* 2002;**5**:143-51.
- 11 **National Center for Health Statistics**. *National hospital discharge and ambulatory surgery data*. Available at <http://www.cdc.gov/nchs/about/major/hdasd/nhds.htm> (accessed March 2005).
- 12 **McCraig LF**, Burt CW. *National Hospital Ambulatory Medical Care Survey: 2001 Emergency Department Summary*, Advance Data from Vital Health Statistics; No 335, June 2003. Available at <http://www.cdc.gov/nchs/data/ad/ad335.pdf> (accessed March 2005).
- 13 **Injury Surveillance Workgroup**. *Consensus recommendations for using hospital discharge data for injury surveillance*. A Report of the State and Territorial Injury Prevention Directors Association. Marietta, GA, 2003.
- 14 **LeMeir M**, Cummings P, West TA. Accuracy of external cause of injury codes reported in Washington State hospital discharge records. *Inj Prev* 2001;**7**:334-8.