Accident Analysis and Prevention 52 (2013) 133–143

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Development of a robust mapping between AIS 2+ and ICD-9 injury codes

Ryan T. Barnard [c], Kathryn L. Loftis [a,b], R. Shayn Martin [b], Joel D. Stitzel [a,b,*]

[a] Virginia Tech-Wake Forest University, Center for Injury Biomechanics, Medical Center Boulevard, Winston-Salem, NC 27157, USA
[b] Wake Forest University, School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, USA
[c] Wake Forest University, Health Sciences, Medical Center Boulevard, Winston-Salem, NC 27157, USA

ARTICLE INFO

Article history:
Received 21 August 2012
Received in revised form 7 November 2012
Accepted 30 November 2012

Keywords:
Abbreviated Injury Scale
Injury code
International Classification of Diseases
Motor vehicle crash

ABSTRACT

Motor vehicle crashes result in millions of injuries and thousands of deaths each year in the United States. While most crash research datasets use Abbreviated Injury Scale (AIS) codes to identify injuries, most hospital datasets use the International Classification of Diseases, version 9 (ICD-9) codes. The objective of this research was to establish a one-to-one mapping between AIS and ICD-9 codes for use with motor vehicle crash injury research. This paper presents results from investigating different mapping approaches using the most common AIS 2+ injuries from the National Automotive Sampling System-Crashworthiness Data System (NASS-CDS). The mapping approaches were generated from the National Trauma Data Bank (NTDB) (428,637 code pairs), ICDMAP (2500 code pairs), and the Crash Injury Research and Engineering Network (CIREN) (4125 code pairs). Each approach may pair given AIS code with more than one ICD-9 code (mean number of pairs per AIS code: NTDB = 211, ICDMAP = 7, CIREN = 5), and some of the potential pairs are unrelated. The mappings were evaluated using two comparative metrics coupled with qualitative inspection by an expert physician. Based on the number of false mappings and correct pairs, the best mapping was derived from CIREN. AIS and ICD-9 codes in CIREN are both manually coded, leading to more proper mappings between the two. Using the mapping presented herein, data from crash and hospital datasets can be used together to better understand and prevent motor vehicle crash injuries in the future.

1. Introduction

In 2008, motor vehicle crashes (MVCs) resulted in over 37,000 deaths and over 2.3 million injuries in the US (NHTSA, 2010b). It is important to study injuries that occur in real-world crashes to gain better understanding of how they are being caused and what can be done to prevent them in the future. In the study of these injuries using disparate databases, it has become apparent that a mapping between Abbreviated Injury Scale (AIS) codes and International Classification of Diseases version 9 (ICD-9) codes is needed. The objective of this study is to provide researchers such a link between AIS and ICD-9 codes for the most frequent MVC-induced injuries.

1.1. Injury codes

The primary injury coding lexicon used by hospitals in the US is the ICD-9 coding system. ICD-9 codes are nominal, meaning they are unordered, qualitative categories not ranked by severity. ICD is not trauma-specific, but rather is a general, all-purpose diagnosis taxonomy for all health conditions and it is generally used for hospital billing. It is over 110 years old and is currently in its 10th revision (ICD-10), though in the United States the 9th revision (ICD-9) is most commonly used (National Center for Health Statistics (NCHS) and Centers for Medicare & Medicaid Services (CMS), 2008).

The most advanced trauma-specific, anatomically based coding lexicon is the AIS code, which was first conceived as a system to define the type and severity of injuries arising from MVCs. To calculate AIS scores, medical records of traumatic incidents are transcribed into specific codes that capture individual injuries. AIS is a proprietary classification system, requiring specialized training for coding personnel and it is typically used for research purposes (AAAM, 2001).

The actual AIS code consists of two numerical components. The first component is a six-digit injury descriptor code ("pre-dot"), which is unique to each traumatic injury. Pre-dots classify the injury by region, type of anatomic structure, specific structure, and level. The second component is a severity score ("post-dot"), graded from 1 (minor) to 5 (critical injury), with the caveat that all unsurvivable injuries are scored a 6. These severity scores, or "AIS severity," are consensus-derived assessments assigned by a group of experts and they were last updated in 2008. Of note, AIS is used as both a classification scheme for injury coding (the pre-dots) as

* Corresponding author at: Medical Center Boulevard, Winston-Salem, NC 27157, USA. Tel.: +1 336 716 5597; fax: +1 336 458 9306.
E-mail addresses: rybarnar@wakehealth.edu (R.T. Barnard), kaloftis@wakehealth.edu (K.L. Loftis), romartin@wakehealth.edu (R.S. Martin), jdstitzel@gmail.com, jstitzel@wfubmc.edu (J.D. Stitzel).

0001-4575/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aap.2012.11.030

well as a severity score (the post-dots). AIS scores are useful in the research setting because they allow studies comparing individuals or populations by injury severity. AIS severities can also be used to calculate the Injury Severity Score (ISS) where the highest AIS severity in each of three body regions is squared and then summed to result in an overall injury severity metric (Baker et al., 1974). The most current AIS codes are 2005 version 2008 update, but most research datasets are primarily coded using AIS 1990 update 1998.

1.2. Motivation

While hospital databases utilize ICD-9 codes, many research databases use AIS codes. As an example, the National Trauma Data Bank (NTDB) uses primarily ICD-9 codes while the National Automotive Sampling System–Crashworthiness Data System (NASS-CDS) uses AIS codes (National Highway Traffic Safety Administration, 2011). The plurality of coding systems prevents standardization between databases, but researchers often need to combine and compare injury data gleaned from more than one source. Without a reliable mechanism for translating from one coding system to the other, such combinations are impossible. This need coupled with the dominance of ICD-9 coding—especially contrasted against the comparative paucity of sources depending on the trauma-specific AIS system—motivates an ICD-9-based standardization. This is accomplished via a comprehensive, accurate, and targeted mapping from AIS codes to ICD-9 codes.

Such a mapping has been attempted previously in the ICDMAP tool, but, though it may be useful, its nature poses potential shortcomings that suggest that a superior mapping is possible. In particular, ICDMAP is comprised of two distinct mappings rather than a single cohesive mapping. These mappings are the "AIS High" and "AIS Low" mappings—referred to hereafter simply as AIS-High and AIS-Low, respectively. These provide mappings from ICD-9 to either higher or lower severity AIS codes, and the user must make a decision on the preferred mapping. Furthermore, the mappings provided by ICDMAP are neither exhaustive nor highly specific: many AIS and ICD-9 codes are missing, and among those codes that are present, the suggested ICD-9 code for a given AIS code may be overly general even though a relevant and specific code may exist (Johns Hopkins University and Tri-Analytics Inc., 2007).

1.3. Databases

There are two main databases used in the creation of the mapping between AIS codes and ICD-9 codes—NTDB and CIREN. A third database—NASS-CDS—performs an auxiliary role in identifying the most frequently occurring injuries to help constrain the set of injuries involved. These databases are summarized in Table 1.

1.3.1. NASS-CDS

The NASS-CDS database contains close to 1000 variables on motor vehicle crashes that occur in the US. These NASS-CDS cases are weighted to provide a population sample for studying aspects of a real-world crash. NASS-CDS has detailed data on a representative, random sample of thousands of minor, serious, and fatal crashes, all of which is supplied by 24 highly trained field research teams that study about 5000 crashes a year involving passenger cars, light trucks, vans, and utility vehicles.

The NASS-CDS database is provided in multiple different "tables", and each one represents a different aspect of the crash. Injuries in NASS-CDS are coded using AIS scores. No ICD-9 codes are provided within NASS. Using these tables within the NASS-CDS database, it is then possible to make a connection between the crash, vehicle, and occupant characteristics and the injuries (UMTRI, 2009).

1.3.2. NTDB

The NTDB collects trauma registry data from participating trauma centers on an annual basis and is the most complete trauma database available to researchers. All hospitals with trauma registries are encouraged to participate, and the version 7.2 dataset includes 256 hospitals across the country, both American College of Surgeons (ACS) verified and state designated trauma centers. The NTDB Research Data System (RDS) contains all records sent to NTDB for each admission year. Data regarding patient demographics, injury severity and injury origin is collected as well as descriptive accounts of each traumatic incident. Criteria for participation in the database include the reporting institution's trauma accreditation and hospital or emergency department admission. Patients who are dead upon arrival are excluded from the NTDB. Data submitted to NTDB are rigorously examined using both the NTRACS system institutionally and an additional logical checks system created and enforced by NTDB administrators (American College of Surgeons, 2007).

The NTDB dataset is used in the current study to explore data on trauma centers and patient information. Cases coded in NTDB include both AIS and ICD-9 codes. These codes are supplied independently of one-another; there is no database-supplied mechanism to identify which AIS code(s) relate to which ICD-9 codes. NTDB's strength is instead in its extent (containing over 1.9 million distinct cases), its broad employment, and its rigorous review process.

1.3.3. CIREN

The Crash Injury Research and Engineering Network (CIREN) is a database sponsored by the National Highway Traffic Safety Administration (NHTSA) to investigate US motor vehicle crashes resulting in serious injury. Currently, six trauma centers across the US recruit and collect information from occupants upon admission to the hospital sustaining AIS 2+ injuries due to MVCs. There are now more than 4000 cases within the CIREN database, which began in 1996. Enrolled crashes must be "clean" (simple crash scenarios, no complicated multi-event crashes) and are restricted to vehicles with current model years (within six years of the current year). By studying the crash, vehicle, scene, and medical records, the injury causation (Biotab) is determined for each injury (Schneider et al., 2011). This CIREN data can be used to research mechanisms and occupant outcomes for specific injuries (Gayzik et al., 2009; Danelson et al., 2011; Loftis et al., 2011; Urban et al., 2012; Weaver et al., 2012a, 2012b). In the CIREN database, each case contains injury information coded using both the AIS and ICD-9 systems (NHTSA, 2010a).

2. Methods

Two new mappings (accompanied by several variations of each) have been constructed and evaluated against the code pairings suggested by ICDMAP. To extract all possible ICD-9 to AIS pairings suggested by ICDMAP, we generated the list of all possible ICD-9 codes from 800 to 959.99. This list consists of the 160 integer-part-only codes and, for each, the 100 decimal codes from .00 to .99 and the 10 single-digit decimal codes from .0 to .9, resulting in $160 + (160 \times (100 + 10)) = 17{,}760$ possible ICD-9 codes. Each of these codes was run through ICDMAP twice; once on its AIS-Low setting and again on its AIS-High setting.

To determine the most frequent injuries occurring in MVCs in the US, NASS-CDS 2000–2011 was used. For each injury in the database, we computed two values: the total (weighted) number of times that injury appears, and the percentage of all weighted injuries represented by that injury. We then ordered all of the injuries with an AIS severity of two or greater in decreasing order of