



Motor Vehicle Injuries

Scientific

Nonfatal motor vehicle crash injuries: Wisconsin's experience with linked data systems

Trudy A. Karlson, PhD, Chip Quade, and Martha Florey, MS, JD, Madison

Introduction

Between 1989 and 1993, motor vehicle crashes killed 3,860 people in Wisconsin. Many more persons sustained nonfatal injuries. Nationally, the annual cost of motor vehicle-related injuries in 1985 dollars is estimated at \$48.7 billion.¹

Preventing crash-related injuries and reducing their severity requires the application of numerous strategies, including engineering, education and enforcement of existing laws. Choosing an appropriate strategy and evaluating its effectiveness is essential in a time of scarce resources. Fortunately, Wisconsin has been able to take advantage of existing data sources originally designed for other purposes to build a new and powerful tool for health officials, transportation

professionals, researchers and others who seek information about the nature and extent of the state's crash injury problem.²

This study provides a brief overview of Wisconsin's unique five-year nonfatal crash injury data base and encourages its use in the investigation of motor vehicle crash injuries and their prevention. Since this is a relatively new data base creation method, we've also chosen to describe the technology in detail.

Background

Wisconsin's efforts to produce linked data have been the result of collaborative agreements between several state and national agencies, specifically the Office of Health Care Information (OHCI) in the Office of the Commissioner of Insurance, the Wisconsin Department of Transportation (DOT), the University of Wisconsin's Center for Health Systems Research and Analysis and the National Highway Traffic Safety Administration,

Karlson and Quade are associated with the Center for Health Systems Research and Analysis at the University of Wisconsin. Florey is with the Bureau of Transportation Safety, Wisconsin Department of Transportation. Reprint requests to Trudy A. Karlson, PhD, Center for Health Systems Research and Analysis, 11th Floor, WARF Building, 610 Walnut Street, Madison, WI 53705. Copyright 1996 State Medical Society of Wisconsin.

Table 1. Variables used for linking Wisconsin crash and hospital discharge data.

| Variable Type | Crash variable | Hospital discharge variable |
|---------------------|-----------------------|-----------------------------|
| Event information | Crash date | Admission date |
| | County of crash | County of hospitalization |
| | HSA of crash | HSA of hospitalization |
| Patient information | Date of birth | Date of birth |
| | Sex | Sex |
| | Zip code of residence | Zip code of residence |
| | Injury flag | (No direct counterpart) |
| | EMS transport flag | (No direct counterpart) |

*OHCI's public use data do not contain birth or admission dates, both of which are required for the linkage. The use of this information is subject to strict confidentiality requirements. Because of this, we linked the crash data with a custom OHCI data file under an administrative arrangement that assured the OHCI confidentiality requirements wouldn't be violated. We removed birth and admission dates before generating the final data set for analysis.

which funded much of the work. With funds supplied in part by the federal Department of Transportation, similar efforts to link motor vehicle crash data and health data took place in seven states.³⁴

Data sources

Inpatient hospital discharge data. Wisconsin statutes require every hospital in the state to submit outpatient discharge information to OHCI. The information includes patient birth date, sex and zip code, but no other identifiers. Information about the hospitalization includes diagnoses and procedures, total charges; type and source of admission, discharge status, length of stay and expected primary and secondary payer.

The OHCI data base contains information about Wisconsin hospitalizations only. It contains no information about people injured in Wisconsin but transported to other states for treatment.

Crash data. Law enforcement agencies submit crash report forms to DOT's Traffic Accident Section. In Wisconsin, a crash is reportable if it occurs on public property and results in any injury or in property damage of \$500 or more (\$1,000 or more as of January 1, 1996). The crash data base includes vehicle occupants' demographics, their seating location, use of helmets, safety belts or airbags, vehicle type, crash location and circumstances.

DOT's crash data base contains information about occupants of all reported crashes that occur in Wisconsin. It contains no information about occupants of unreported

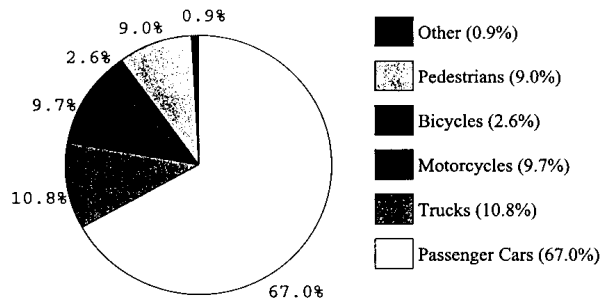


Fig 1 - Vehicle types for hospitalized crash occupants, Wisconsin 1993.

crashes, nor those who are treated in Wisconsin hospitals after crashes that occur in other states.

Methods

Data linkage. In linking records from two data sources, the goal is to identify records in each source representing the same person and event. In the absence of personal identifiers such as names or social security numbers, other information must be used.⁵ We used information about the occupant (or patient) and the date and location of the event (the crash or hospitalization). Occupant information includes the birth date, sex and zip code.

Event information includes the crash date and county and the admission date and county of hospitalization. Crash occupants, especially those with severe injuries, often are transported across county lines for medical treatment. For this reason, we map each county code

into a code representing a region, or Health Service Area (HSA), and use this variable in the linkage as well.

We also take advantage of variables in the crash data indicating whether the occupant was injured, and whether the occupant was transported via EMS to an emergency department. Table 1 summarizes the linkage variables.

Probabilistic linking is an efficient technique for linking data files in the absence of definitive personal identifiers. In this approach, a computer algorithm determines the likelihood that two records, one from each data file, represent the same person and event.⁶ The algorithm compares the fields in the two records and computes a score for the pair that reflects:

- *Similarities and differences:* The score goes up when information in the two records is similar or identical.

Table 2. Vehicle types and rates of hospitalization per crash occupant, Wisconsin 1993.

| | Passenger Vehicles | Trucks | Motorcycles | Bicycles | Pedestrians |
|-------------------------------|--------------------|--------|-------------|----------|-------------|
| Hospitalizations Occupants of | 3,269 | 528 | 474 | 127 | 436 |
| Reported Crashes | 295,030 | 58,564 | 2,886 | 1,921 | 2,630 |
| Hospitalization Rate | 1.1% | 0.9% | 16.4% | 6.6% | 16.6% |

- *The underlying frequency distribution:* The score goes up when the records match on values that occur less frequently in the data sets.
- *The effect of user-assigned weights:* The overall score for the pair is the sum of the scores for comparisons of individual fields. The user assigns a weight to each field that reflects its relative importance and reliability.

The linking program computes scores for all possible matched pairs. A probabilistic linking program can calculate the scores quickly, even for large data files. When linking 250,000 crash records against 90,000 hospital records on a 486-based microcomputer, the matching program computes its scores in approximately three hours.

Data linkage does require trial and error to determine an optimum set of user-assigned weights, and computing the scores is only one step in the overall linkage process. Even so, a relatively inexperienced user can link a year's worth of crash and hospital data in approximately 80 hours, with time for trial and error, data management and other associated tasks.

Definitions

- A passenger vehicle is any automobile, van, sport utility vehicle or pickup truck.
- A truck is any truck except a pickup.
- A motorcyclist is a rider of either a motorcycle or moped.
- Brain injuries include skull fractures with intracranial injury; cerebral, cerebellar or brain stem lacerations or contusions; intracranial hemorrhages; and intracranial injuries without skull fracture but with loss of consciousness for greater than one hour. These were identified by ICD-9-CM diagnosis codes in the discharge record.
- Hospital charges don't include

Table 3. Vehicle types and rates of hospitalization for brain injury, Wisconsin 1993.

| | Passenger Vehicles | Motorcycles | Pedestrians |
|-----------------------------------|--------------------|-------------|-------------|
| Brain injury Hospitalizations | 548 | 110 | 95 |
| Occupants of reported crashes | 295,030 | 2,886 | 2,630 |
| Brain injury hospitalization rate | 0.2% | 3.8% | 3.6% |

physician's fees, nor do they include care rendered outside of the inpatient setting.

Findings

In 1993, the most recent year for which linked data currently are available, 376,220 people were involved in motor vehicle crashes in Wisconsin. Of these, 4,878 (1.3%) received inpatient treatment for crash injuries. Figure 1 shows the distribution of vehicle types for hospitalized crash occupants.

Passenger vehicle occupants accounted for the greatest share of both the overall crash victims (78.4%) and the hospitalized victims (67.0%). However, motorcyclists represented a much greater share of the hospitalized occupants than they did of the total crash victims. The same is true for pedestrians. Motorcyclists, pedestrians and bicyclists are more susceptible to injury in event of a crash because they're not protected by a passenger compartment. Table 2 shows the hospitalization rates for each type of vehicle.

Brain injuries also occurred more frequently among motorcyclists and pedestrians than among passenger vehicle occupants. Table 3 shows the brain injury hospitalization rates for passenger vehicles, motorcycles and pedestrians.

While overall hospitalization rates may be declining, including hospitalization rates for motor vehicle crash injuries, the average hospitalization charge is rising.

Table 4 takes a five-year look at motor vehicle crash occupants, hospitalizations and deaths.

Discussion

Data linkage has proven to be an efficient, cost-effective way to create an injury surveillance system from existing data sources designed for other purposes. As with all data systems, a linked data file has its limitations. Some arise from the linkage methodology, and others are inherent in the primary data sources.

Some crash-related hospitalizations are missing from the linked data set. In our experience, the linked data files have an estimated 80 percent predictive value positive, or about a 20 percent false negative rate. These false negatives result in an underestimate of the motor vehicle crash injury problem. Hospitalizations may be excluded from the data set for a variety of reasons, including:

- People who are in Wisconsin crashes but transported out of state for treatment are excluded.
- People in unreported crashes who have no records in the crash data, and therefore can't be linked. Similarly, persons whose crash records contain insufficient identifying information (missing birth date or sex) can't be linked with the hospital data file, and aren't included in the final linked data file.
- The linkage is based on likelihood, and the scores for some

Table 4. Crash occupants, hospitalizations, hospital charges and deaths, Wisconsin 1989- 1993.

| | 1989 | 1990 | 1991 | 1992 | 1993 | Total |
|------------------------------------|--------------|--------------|--------------|--------------|--------------|---------------|
| Crash Occupants | 371,692 | 374,153 | 363,011 | 362,901 | 376,220 | 1,847,977 |
| Hospitalizations | 6,402 | 5,939 | 4,956 | 5,282 | 4,878 | 27,457 |
| Total Hospital Charges | \$59 million | \$56 million | \$60 million | \$71 million | \$70 million | \$316 million |
| Average Charge Per Hospitalization | \$9,228 | \$9,422 | \$12,009 | \$13,513 | \$14,357 | \$11,508 |
| Deaths | 844 | 795 | 823 | 674 | 724 | 3,860 |

valid matched pairs may fall below a minimum threshold.

Some linked pairs are false positives. Validation of the match results with auxiliary data sources indicates that in approximately 7 percent of the linked pairs, the crash record and hospital discharge record don't represent the same person and event. False positives are more likely to occur among population groups with high crash frequencies, and among occupants who live in populous areas.

Some crash types may be under-represented in the DOT crash data. If a pedestrian or bicyclist is injured but the crash doesn't involve a motor vehicle, the crash may not be reported to the DOT. Uninjured motorcyclists also may be under-represented.

There may be biases in the reporting of some variables in the crash data. Eighty-five percent of crash occupants reportedly use seat belts, even though observational studies have determined the statewide rate of seat belt use to be approximately 55% to 60%.⁷ Other information, such as alcohol use, is likely to be selectively reported.

Despite limitations, a linked data base is an extremely useful tool for documenting the nature and magnitude of crash injuries

and evaluating the effectiveness of specific countermeasures. Nonfatal motor vehicle crash injuries sometimes are very severe, and can result in lifelong disability. This makes them expensive to treat. In an era of spiraling health care costs, public health professionals should explore all possible means for preventing these injuries.⁸ A linked data system contains information on occupant, vehicle and environment factors, which can be used for problem identification at the national, state, regional and local levels.

The Office of Health Care Information has 5 years of linked crash data available, and is linking a sixth year. These data are available to researchers, policy makers and health professionals, and we encourage data acquisition and use to further the development and evaluation of programs to prevent motor vehicle crash injuries.

References

1. Rice DP, MacKenzie EJ et al. *Cost of Injury in the United States: A report to Congress*. San Francisco: Institute for Health and Aging, University of California and Injury Prevention Center, The Johns Hopkins University, 1989.
2. Karlson TA, Quade CA. Head Injuries Associated with Motorcycle Use -Wisconsin, 1994. *MMWR* 43:423, 429-432.
3. Johnson S, Walker J, Utter D. *Crash Outcome Data Evaluation System (CODES) Project Safety Belt and Helmet Analyses*. National Center for Statistics and Analysis Research Note. Washington, DC: National Highway Traffic Safety Administration, U.S. Department of Transportation, 1995.
4. National Highway Traffic Safety Administration. *Report to Congress: Benefits of Safety Belts and Motorcycle Helmets, Based on Data from the Crash Outcome Data Evaluation System (CODES)*. DOT HS 808 347. Washington, DC: US Department of Transportation, 1996.
5. Quade CA. Injury Surveillance Using Linked Data. <http://linear.chsra.wisc.edu/chip/linkinfo>. Madison, WI: Office of Transportation Safety, Wisconsin Department of Transportation, 1995.
6. Jaro M. *Automatch User's Guide, Version 2.0*. Silver Spring, MD: MatchWare, Inc., 1994.
7. Wisconsin Office of Transportation Safety. *Motorcycle Helmet Use in Wisconsin, 1993*. Madison, WI: Office of Transportation Safety, Wisconsin Department of Transportation, 1994.
8. Karlson TA. Injury Control and Public Policy. *Critical Reviews in Environmental Control* 22: 195-241, 1992. ♦