HHS Public Access

Author manuscript

Inj Prev. Author manuscript; available in PMC 2023 June 01.

Published in final edited form as:

Inj Prev. 2022 June ; 28(3): 262–268. doi:10.1136/injuryprev-2021-044397.

Non-fatal injury data: characteristics to consider for surveillance and research

Andrea E. Carmichael[1,2], **Michael F. Ballesteros**[2], **Judith R. Qualters**[2], **Karin A. Mack**[2]

[1]Oak Ridge Associated Universities (ORAU), Division of Injury Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

[2]Division of Injury Prevention, National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

Abstract

Background—All data systems used for non-fatal injury surveillance and research have strengths and limitations that influence their utility in understanding non-fatal injury burden. The objective of this paper was to compare characteristics of major data systems that capture non-fatal injuries in the USA.

Methods—By applying specific inclusion criteria (eg, non-fatal and non-occupational) to well-referenced injury data systems, we created a list of commonly used non-fatal injury data systems for this study. Data system characteristics were compiled for 2018: institutional support, years of data available, access, format, sample, sampling method, injury definition/coding, geographical representation, demographic variables, timeliness (lag) and further considerations for analysis.

Results—Eighteen data systems ultimately fit the inclusion criteria. Most data systems were supported by a federal institution, produced national estimates and were available starting in 1999 or earlier. Data source and injury case coding varied between the data systems. Redesigns of sampling frameworks and the use of International Classification of Diseases, 9th Revision, Clinical Modification/International Classification of Diseases, 10th Revision, Clinical Modification coding for some data systems can make longitudinal analyses complicated for injury surveillance and research. Few data systems could produce state-level estimates.

Correspondence to Andrea E. Carmichael, Centers for Disease Control and Prevention, Atlanta, USA; acarmichael@cdc.gov.

Contributors AEC was involved in all aspects of the study, including planning, writing and reviewing the paper, and analysing and reporting the results. MFB, JRQ and KAM were involved in planning, writing and reviewing the article. AEC is the guarantor for this work.

Competing interests None declared.

Ethics approval This study did not use human subjects, as the data systems are publicly available. Therefore, no institutional review board approval was necessary for this study.

Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/injuryprev-2021-044397).

Conclusion—Thoughtful consideration of strengths and limitations should be exercised when selecting a data system to answer injury-related research questions. Comparisons between estimates of various data systems should be interpreted with caution, given fundamental system differences in purpose and population capture. This research provides the scientific community with an updated starting point to assist in matching the data system to surveillance and research questions and can improve the efficiency and quality of injury analyses.

BACKGROUND

Fatal injury data are often used to assess burden of injury due to strong data quality and easy availability through query systems such as the Centers for Disease Control and Prevention's (CDC) Wide-Ranging Online Data for Epidemiologic Research (WONDER)[1] and Web-based Injury Statistics Query and Reporting System (WISQARS).[2] However, fatal injury data focus on only the most severe injuries, capturing only a portion of the total injury burden. Surveillance and research on non-fatal injuries, which can cause substantial morbidity and disability and incur hefty economic costs, are essential to inform policy and practices that protect individuals from non-fatal injury.[3] Additionally, when comparing non-fatal injury with fatal injury, leading causes for each often differ by age group, indicating decisions on priority setting and prevention strategies are better informed if non-fatal injury burden is also considered.[4 5]

In order to conduct surveillance and research activities to assess non-fatal injury burden and to understand non-fatal injury risk and protective factors, characteristics of databases and other data resources (hereafter referred to as data systems) for monitoring non-fatal injury should be understood and considered. Since data system characteristics can vary widely, there are many elements to consider when selecting the appropriate data system for non-fatal injury surveillance and research. Existing non-fatal injury data systems have strengths and limitations which influence their utility for surveillance and answering specific non-fatal injury research questions. The objective of this paper was to expand on previous literature[6 7] by comparing characteristics of major data systems that capture information on non-fatal injuries in the USA. The focus is on those characteristics that influence the accessibility and other high-level attributes that may impact data utility. However, it is not intended to provide an in-depth evaluation of each data system and does not address aspects of data management, storage and specific quality metrics such as completeness of variables.[8] The information provided in this paper can assist with selecting appropriate data systems for specific non-fatal injury surveillance and research activities and can be used to compare the qualities of data systems within and outside the USA.

METHODS

Appendix 3 of the *Handbook of Injury and Violence Prevention*[9] and the CDC 'Inventory of National Injury Data Systems'[10] were reviewed to identify common data systems for injury analyses. Other data systems were also considered based on prior knowledge of data systems used in injury research. Data systems were included based on the following criteria:

- Collects data on the US population or a specific population therein (eg, children, subnational).

- Does not solely capture violence-specific, occupational-specific or fatal injuries.
- Contains at least one variable that can be used to define a non-fatal injury case (eg, International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM)/International Classification of Diseases 10th Revision, Clinical Modification (ICD-10-CM) codes; predefined injury variables based on ICD-9-CM/ICD-10-CM or self-reported injury).
- May be used for surveillance (eg, not a one-time survey, not discontinued).
- Collects data on more than one unintentional injury mechanism and more than one body region affected.
- Is publicly available.

Routinely collected survey data and data systems focused on injury risk factors were included if all other inclusion criteria were met. Documentation and codebooks for the selected data systems were acquired, and characteristics of the systems were consolidated. The following data system characteristics were gathered using 2018 (as of August 2021, a majority of data systems had most recent data starting in 2018 or later; therefore, 2018 was selected) documentation, unless otherwise noted:

- Institutional support (primary organisation(s) overseeing data management and providing funding).
- Years of data available (years of publicly available data across the entire span of data releases).
- Access/format (how to obtain the data and which formats are publicly available).
- Description (brief overview of data system).
- Data source (where data are obtained).
- Sampling method (sample and sampling procedures).
- Injury definition/coding (how injury variables and/or injury case is defined).
- Available demographic variables (ie, age, sex, race, ethnicity, income and urban/rural).
- Geographical representation (national/regional/state incidence estimates).
- Timeliness/lag (most recent data year available).
- Strengths/limitations (notable strengths/limitations that may affect planning of analyses, across the entire span of data releases).

It was not appropriate or possible to involve patients or the public in the design, conduct, reporting or dissemination plans.

RESULTS

Data systems identified

Of the 71 data systems considered, 53 data systems were outside the study scope based on exhibiting one or more of the following characteristics: mechanism-specific (24 data systems); violence-only (10); not routinely released, have been discontinued or are one-time surveys (8); fatal-only (8); no injury-related variables (5); occupational-specific (3); body region-specific (2); not publicly available (2), and/or no injury mechanism variables (1). Ultimately, 18 data systems total were included (online supplemental appendix 1):

- Healthcare Cost and Utilization Project (HCUP) data systems.[11]
  - Nationwide Emergency Department Sample (NEDS).
  - National (Nationwide) Inpatient Sample (NIS).
  - Nationwide Readmissions Database (NRD).
  - Nationwide Ambulatory Surgery Sample (NASS).
  - Kids' Inpatient Database (KID).
  - State Emergency Department Databases (SEDD).
  - State Inpatient Databases (SID).
  - State Ambulatory Surgery and Services Databases (SASD).
- Outcome Assessment Information Set (OASIS).[12]
- MarketScan Commercial Claims and Encounters Database (referred to as MarketScan from hereout).[13]
- Medicare Provider Analysis and Review (MedPAR) Limited Data Sets.[12]
- National Ambulatory Medical Care Survey (NAMCS).[14]
- National Electronic Injury Surveillance System (NEISS).[15]
- National Electronic Injury Surveillance System-All Injury Program (NEISS-AIP).[16]
- National Emergency Medical Services Information System (NEMSIS).[17]
- National Health Interview Survey (NHIS).[18]
- National Hospital Ambulatory Medical Care Survey (NHAMCS).[14]
- National Trauma Data Bank (NTDB).[19]

Aggregate characteristics

Incidence estimates and data source—Characteristics of data systems varied (table 1, figure 1 and online supplemental appendix 1). Twelve of the 18 data systems could produce nationally representative incidence estimates; 4 could produce state estimates; and 2 could produce regional estimates (table 1). Many of the data systems sampled administrative data for hospitalisations (5) and/or emergency department (ED) visits (5). Other data systems

sampled medical claims (3), ambulatory surgeries (2), hospital readmissions (1), office-based physician visits (1), household interviews (1), trauma admissions (1), and emergency medical services (EMS) activations (1).

Demographic variables—All 18 data systems contained demographic variables for age and sex, while race (14), ethnicity (11), urban/rural (11) and income (9) were less consistently available.

Injury definition/coding—Fifteen data systems contained diagnosis variables based on the ICD-9-CM/ICD-10-CM coding systems, which can be used to create injury variables; external cause-of-injury variables were available for 14 data systems. In most instances, the diagnosis variable arrays also contained the external cause-of-injury codes. NEISS and NEISS-AIP, however, used trained data coders to abstract information directly from medical records to code injury mechanism. Eight data systems provided predefined injury variables based on ICD-9-CM/ICD-10-CM or self-reported injury. Of the data systems, only NEISS, NEISS-AIP and NTDB were data systems specifically designed for injury surveillance (consumer product-related injury, all injury-related ED visits, and injuries treated in a trauma centre, respectively). All data systems contained variables helpful for analysing discharge status.

Access/format—Data files were available to researchers for all data systems through institutional support websites for free or for purchase (table 1 and online supplemental appendix 1); 10 data systems also provided a free online query tool.

Data availability years—As of August 2021, data systems varied in the most recently released data year: two had year 2020; nine had year 2019; six had year 2018; and one had year 2016 as the most recent data year available (figure 1). The earliest data year available was 1963 (NHIS). Ten data systems had less than 20 years of data, while 8 had greater than 20 years of data available. KID released data every 3 years starting in 1997, with no data available for 2015 due to the ICD-9-CM to ICD-10-CM transition; therefore, 2016 was the most recently released year.[20]

Considerations when planning analyses

Accessibility—All data systems included in this study were available through a query system or through other online access (free public use files or for-purchase data files, table 1). NEISS, NEISS-AIP, the HCUP data systems (excluding KID, NASS and SASD), NHIS and NEMSIS all provided online query tools for analyses with limited data.[2 15 18 21 22] The availability of a query tool helps provide the researcher with quick, aggregated results, which can be used for surveillance or help guide more in-depth analyses.

Data lag—Average lag in data availability ranged from a few months to approximately 2 years, with NEISS and NEMSIS data being released the quickest by full calendar year of data released (approximately 4–5 months after close of previous calendar year), while the HCUP national data systems (NEDS, NIS, NRD, NASS and KID) and NAMCS observed some of the longest average lag times to data release (approximately 18–24 months).

Injury coding systems—The HCUP data systems, MarketScan, MedPAR, NAMCS, NHAMCS, NEMSIS, NHIS, NTDB and OASIS all rely on the ICD-10-CM coding system. Prior to 1 October 2015, these data systems used the ICD-9-CM coding system. The ICD-9-CM/ICD-10-CM coding systems offer information to analyse specific injuries within a mechanism, especially after the transition (eg, in the ICD-10-CM system: W17.0XXA, 'Fall into well, initial encounter').[23 24] However, data systems that include narratives such as NEISS or NEISS-AIP can help the researcher better understand the context in which the injury occurred. Unlike other data systems, NEISS and NEISS-AIP, which are specifically designed for injury surveillance, each contain precipitating cause of injury variables coded by hospital coders who review medical records comprehensively. Understanding the context of injury through narratives is important for guiding public health interventions.

Geographical analyses—The SID, SEDD, SASD and NHIS all contain state data files, which can be used to perform state-level analyses, although availability by state may vary over time. While the NHIS sample is too small for state-level analyses by year, combining data years may provide a big enough sample size for certain analyses.[25] The other data systems should not be used to make state-level estimates due to sampling methods.

Demographic variables—Variables for analysing injury by sex and age were available across all data systems. Income variables were available for NHIS and the HCUP data systems. A race and/or ethnicity variable was available in all data systems except for SEDD, NEDS, NRD and NASS. However, race/ethnicity data were commonly missing or unreliable and should be interpreted understanding measurement limitations.[26] The large sample size of MarketScan and the HCUP data systems lent themselves well to stratification by demographic variables.

DISCUSSION

While much injury surveillance and research has been conducted using the data systems described in this paper, to our knowledge, there has not been a recent review of specific characteristics of data systems commonly used in non-fatal injury surveillance and research. This project adds to prior research conducted by the Agency for Healthcare Research and Quality (AHRQ) in 2009[6] by taking inventory of 14 additional data systems not captured by the AHRQ study. Further, the current study captures not only US ED visits but also hospitalisations, medical claims, ambulatory surgeries, office-based physician visits, household interviews, trauma admissions, hospital readmissions, and EMS activations. This research provides the scientific community with a starting point to select the appropriate data system for non-fatal injury surveillance and research and can be used as a tool to improve the quality of non-fatal injury analyses.

Understanding data

When selecting a data system to use for analyses, it is key to know data system characteristics in detail. A few specific examples of data system characteristics and changes over time are highlighted further; however, this list is not exhaustive. Data system documentation should be consulted prior to performing analyses.

Sample, injury case definitions and coding differences

A data system's sample and injury case definition both impact the types of analyses that may be appropriate for a given data system and may make comparisons across different data systems challenging. Each data system collects data through a variety of methods (eg, probability sample, convenience sample and census) and are abstracted from diverse sources (eg, ED visits, hospitalisations, trauma admissions, hospital readmissions, office-based physician visits, household interviews, EMS activations, ambulatory surgery and medical claims). Similarly, since injury case definitions rely on the coding system from which they are derived, injury case definitions may also vary across data systems, often causing incomparability between injury estimates across data systems.

Additionally, injury coding can change over time both between and within data systems. For example, data systems that rely on the ICD-9-CM/ICD-10-CM coding systems for injury coding have been affected by the transition from the 9th revision to the 10th revision on 1 October 2015.[27 28] There are also slight coding changes within these coding systems each October when codes may be added, modified or deleted.[23] The deletion or modification of codes may render trend analyses challenging over certain time frames for specific codes within a data system. Analysing trends over the 2015 change for data systems that use ICD-9-CM/ICD-10-CM coding should be done with caution.

Some data systems do not rely on the ICD coding system to define injury mechanism, and thus injury mechanism in those systems may contribute different or additional information. For example, NHIS contains injury risk factor variables, such as a variable indicating whether or not an individual was wearing a helmet or seatbelt at the time of a crash,[29] and this information does not have a comparable ICD-9-CM/ICD-10-CM code. NEISS and NEISS-AIP include narratives describing the context of the injury. The context of an injury, whether assessing fatal or non-fatal injury, is often lacking in data systems, especially claims data, but is nonetheless a deeply important aspect of understanding how to prevent injuries.[30]

Sampling framework changes and modifications to data elements

Sampling framework changes and redesigns to data elements over time may also make longitudinal analyses over certain periods complicated. For example, prior to 2012, NIS captured *all* discharges from a sample of hospitals, used the American Hospital Association hospital and discharge definition and included data elements that were not necessarily supplied by all states. In 2012, NIS's sampling framework was redesigned to capture a *sample* of discharges, to use the state organisation's hospital and discharge definition, and to use a standard set of data elements. This redesign was carried out to minimise sampling error, improve confidentiality and produce more precise national estimates. The data user should be cautious about performing NIS trend analyses with data before and after 2012 and be certain to use the proper trend weights as described in the NIS documentation.[31] Other data systems have also experienced redesigns or have undergone revisions to data elements. NHIS redesigns its sampling process following each decennial census.[25] OASIS redesigned their data most recently in 2019.[12] Modifications to data elements and redesigns of sampling frameworks should be identified prior to initiating any longitudinal analyses.

Examples of appropriate surveillance and research uses for non-fatal injury data systems

Non-fatal injury research questions, study plans and analyses can vary widely; goals for injury surveillance can also vary. As mentioned previously, NEISS and NEISS-AIP provide important details about the context in which the injury occurred, which can be helpful for surveillance and research activities. Narratives in NEISS-AIP were used in a surveillance study to categorise traumatic brain injuries (TBIs) based on the sports-related consumer product involved and to understand the context surrounding the TBI occurrence.[32] Haarbauer-Krupa *et al*[33] used NEISS and NEISS-AIP to categorise unintentional TBIs among children, indicating any product and the activity associated with the fall. Using the narratives, certain action words were selected (eg, 'jump' and 'climb') that showed child-initiated action that precipitated a fall and subsequent brain injury, and these cases were compared with cases where a caregiver caused the precipitation of the event or series of events leading to the TBI injury. NEISS-AIP and NEISS are appropriate systems for this research, considering the amount of detail needed from narratives to categorise the injuries, which other data systems, such as the ICD-9-CM/ICD-10-CM-based data systems, do not supply.

Using a combination of data systems with different samples can help provide a more comprehensive picture of injury burden. Peterson *et al*[34] investigated the average medical cost of fatal and non-fatal injuries by type using three data systems that used data from three different sources (medical claims (MarketScan), ED visits (NEDS) and hospitalisations (NIS)). This resulted in capturing a larger range of medical costs due to injury. The study period avoided the ICD-9-CM/ICD-10-CM transition date (1 October 2015). This avoided potential inconsistency in injury definition/coding that could arise by analysing a study population containing cases coded using both the ICD-9-CM and ICD-10-CM coding systems.

Study limitations

This study provides a high-level overview of data systems commonly used for non-fatal injury analysis and can be a resource when seeking to address surveillance or research questions. However, it represents a starting point and does not include all details necessary to determine the most appropriate data system to use. Prior to any analyses, documentation for each data system should be carefully investigated in order to understand other possible limitations, any changes which may have occurred within the time frame of interest, and the data system's suitability for the question being asked. Data quality, not directly addressed in this study, is also important to account for; the reliability of results from a well-planned analysis relies heavily on a good quality data source.

Additionally, many of the data system characteristics were captured for a single data year, and these characteristics could be different for other data years. Further, the data systems included in this paper are not an exhaustive list of commonly used data systems for injury analysis. Fatal-only data systems; violence-specific, topic-specific, risk factor-specific or occupational-specific data systems; and one-time surveys were not included (online supplemental appendix 2). Data systems in online supplemental appendix 2 may be a better source of data for specific injury mechanism surveillance and research questions and

can provide additional context to help inform decision-making. For example, the Fatality Analysis Reporting System covers only motor vehicle traffic crash fatalities but does so with a wider range of variables than a claims-based data system.

CONCLUSION

Thoughtful consideration of strengths and limitations should be exercised when selecting a data system to answer injury-related surveillance and research questions. Comparisons between estimates of different data systems should be interpreted with caution, given fundamental system differences in purpose, data source, sampling methods, and injury definition/coding. Proper justification for selecting a data system is essential to conducting quality surveillance and scientific research. Data systems inherently exhibit characteristics that make each more or less suitable for certain surveillance and research endeavours. The researcher must be fully aware of these differences and select a data system accordingly. This study adds to the previous literature by providing an updated starting place for categorising characteristics of data systems useful for non-fatal injury research and surveillance. Additionally, this study provides a framework for injury researchers to better understand which data system may be most appropriate for a research question or surveillance activity and can be the basis for new investigations or categorisation of other non-fatal data systems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

REFERENCES

1. Centers for Disease Control and Prevention, 2020.CDC WONDER. Available: https://wonder.cdc.gov/
2. National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, 2020.Nonfatal Injury Reports, 2000 - 2019. Available: https://www.cdc.gov/injury/wisqars/nonfatal.html
3. Peterson C, Xu L, Florence C. Average medical cost of fatal and non-fatal injuries by type in the USA. Inj Prev 2021;27:24–33. [PubMed: 31888976]
4. Ballesteros MF, Schieber RA, Gilchrist J, et al. Differential ranking of causes of fatal versus non-fatal injuries among US children. Inj Prev 2003;9:173–6. [PubMed: 12810747]
5. Drew JAR, Xu D. Trends in fatal and nonfatal injuries among older Americans, 2004-2017. Am J Prev Med 2020;59:3–11. [PubMed: 32201184]
6. Owens PL, Barrett ML, Gibson TB, et al. Emergency department care in the United States: a profile of national data sources. Ann Emerg Med 2010;56:150–65. [PubMed: 20074834]

7. Mitchell RJ, Williamson AM, O'Connor R.The development of an evaluation framework for injury surveillance systems. BMC Public Health 2009;9:1–14. [PubMed: 19121216]
8. German R, Lee L, Horan J. Updated guidelines for evaluating public health surveillance systems. MMWR Recomm Rep 2001;50:1–35.
9. Doll L, Bonzo S, Mercy J. Handbook of injury and violence prevention. New York, NY: Springer, 2007.
10. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. Inventory of national injury data systems, 2018. Available: https://www.cdc.gov/injury/wisqars/inventoryinjurydatasys.html [Accessed January 21, 2021].
11. Agency for Healthcare Research and Quality (AHRQ). Healthcare Cost and Utilization Project (HCUP), 2019. Available: https://www.hcup-us.ahrq.gov/ [Accessed 21 Jan 2021].
12. Centers for Medicare and Medicaid Services. Advance equity, 2019. Available: https://www.cms.gov/ [Accessed 21 Jan 2021].
13. IBM. IBM MarketScan research databases. Available: https://www.ibm.com/us-en/marketplace/marketscan-research-databases/details [Accessed 21 Jan 2021].
14. Centers for Disease Control and Prevention, National Center for Health Statistics. Ambulatory health care data, 2021. Available: https://www.cdc.gov/nchs/ahcd/index.htm [Accessed 21 Jan 2021].
15. United States Consumer Product Safety Commission. National Electronic Injury Surveillance System (NEISS). Available: https://www.cpsc.gov/Research-Statistics/NEISS-Injury-Data [Accessed 21 Jan 2021].
16. United States Consumer Product Safety Commission. NEISS coding manual: all trauma, 2018. Available: https://www.cpsc.gov/Research--Statistics/NEISS-Injury-Data [Accessed 21 Jan 2021].
17. NEMSIS. NHTSA office of EMS. Available: https://nemsis.org/ [Accessed 21 Jan 2021].
18. Centers for Disease Control and Prevention, National Center for Health Statistics. National Health Interview Survey, 2020. Available: https://www.cdc.gov/nchs/nhis/index.htm [Accessed 21 Jan 2021].
19. American College of Surgeons. Annual call for data: National Trauma Data Bank (NTDB). Available: https://www.facs.org/quality-programs/trauma/tqp/center-programs/ntdb [Accessed 21 Jan 2021].
20. Agency for Healthcare Research and Quality. Introduction to the HCUP Kids′ Inpatient Database (KID) 2016, 2018. Available: https://www.hcup-us.ahrq.gov/db/nation/kid/kid_2016_introduction.jsp [Accessed January 21, 2021].
21. Agency for Healthcare Research and Quality (AHRQ). HCUPnet. Available: https://hcupnet.ahrq.gov/#setup [Accessed 21 Jan 2021].
22. NEMSIS. EMS Data Cube. Available: https://nemsis.org/view-reports/public-reports/ems-data-cube/ [Accessed 21 Jan 2021].
23. Hedegaard H, Johnson R, Garnett M. The International Classification of Diseases, 10th Revision, Clinical Modification (ICD–10–CM) external cause-of-injury framework for categorizing mechanism and intent of injury. Natl Health Stat Report 2019;136:1–21.
24. Johnson RL, Hedegaard H, Pasalic ES, et al. Use of ICD-10-CM coded hospitalisation and emergency department data for injury surveillance. Inj Prev 2021;27:i1–2. [PubMed: 33674325]
25. CDC National Center for Health Statistics. About the National Health Interview Survey, 2019. Available: https://www.cdc.gov/nchs/nhis/about_nhis.htm [Accessed 21 Jan 2021].
26. James CA, Bourgeois FT, Shannon MW. Association of race/ethnicity with emergency department wait times. Pediatrics 2005;115:e310–5. [PubMed: 15741357]
27. Slavova S, Costich JF, Luu H, et al. Interrupted time series design to evaluate the effect of the ICD-9-CM to ICD-10-CM coding transition on injury hospitalization trends. Inj Epidemiol 2018;5:36. [PubMed: 30270412]
28. Peterson A, Gabella BA, Johnson J, et al. Multisite medical record review of emergency department visits for unspecified injury of head following the ICD-10-CM coding transition. Inj Prev 2021;27:i13–18. [PubMed: 33674328]

29. National Center for Health Statistics (NCHS). 2016 National Health Interview Survey (NHIS) episode public use file, 2017. Available: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2016/Injpoiep_freq.pdf [Accessed 21 Jan 2021].

30. Hemenway D. Let's make it a priority to improve injury data. Inj Prev 2020;26:395–6. [PubMed: 32694193]

31. Agency for Healthcare Research and Quality (AHrQ). Introduction to the HCUP National Inpatient Sample (NIS), 2012. Available: https://www.hcup-us.ahrq.gov/db/nation/nis/NIS_Introduction_2012.jsp [Accessed 21 Jan 2021].

32. Waltzman D, Womack LS, Thomas KE, et al. Trends in Emergency Department Visits for Contact Sports-Related Traumatic Brain Injuries Among Children - United States, 2001-2018. MMWR Morb Mortal Wkly Rep 2020;69:870–4. [PubMed: 32644984]

33. Haarbauer-Krupa J, Haileyesus T, Gilchrist J, et al. Fall-related traumatic brain injury in children ages 0-4 years. J Safety Res 2019;70:127–33. [PubMed: 31847987]

34. Peterson C, Xu L, Florence C. Average medical cost of fatal and non-fatal injuries by type in the USA. Inj Prev 2021;27:24–33. [PubMed: 31888976]

What is already known on the subject

- Data systems used for non-fatal injury surveillance and research have strengths and limitations.
- Surveillance and research are essential to inform policies and practices that protect individuals from non-fatal injury.

What this study adds

- This study compares characteristics of major data systems that capture non-fatal injuries in the USA and expands on previous literature.
- The information in this paper can assist with selecting appropriate data systems for non-fatal injury surveillance and research activities and can be used to compare the qualities of data systems within and outside the USA.

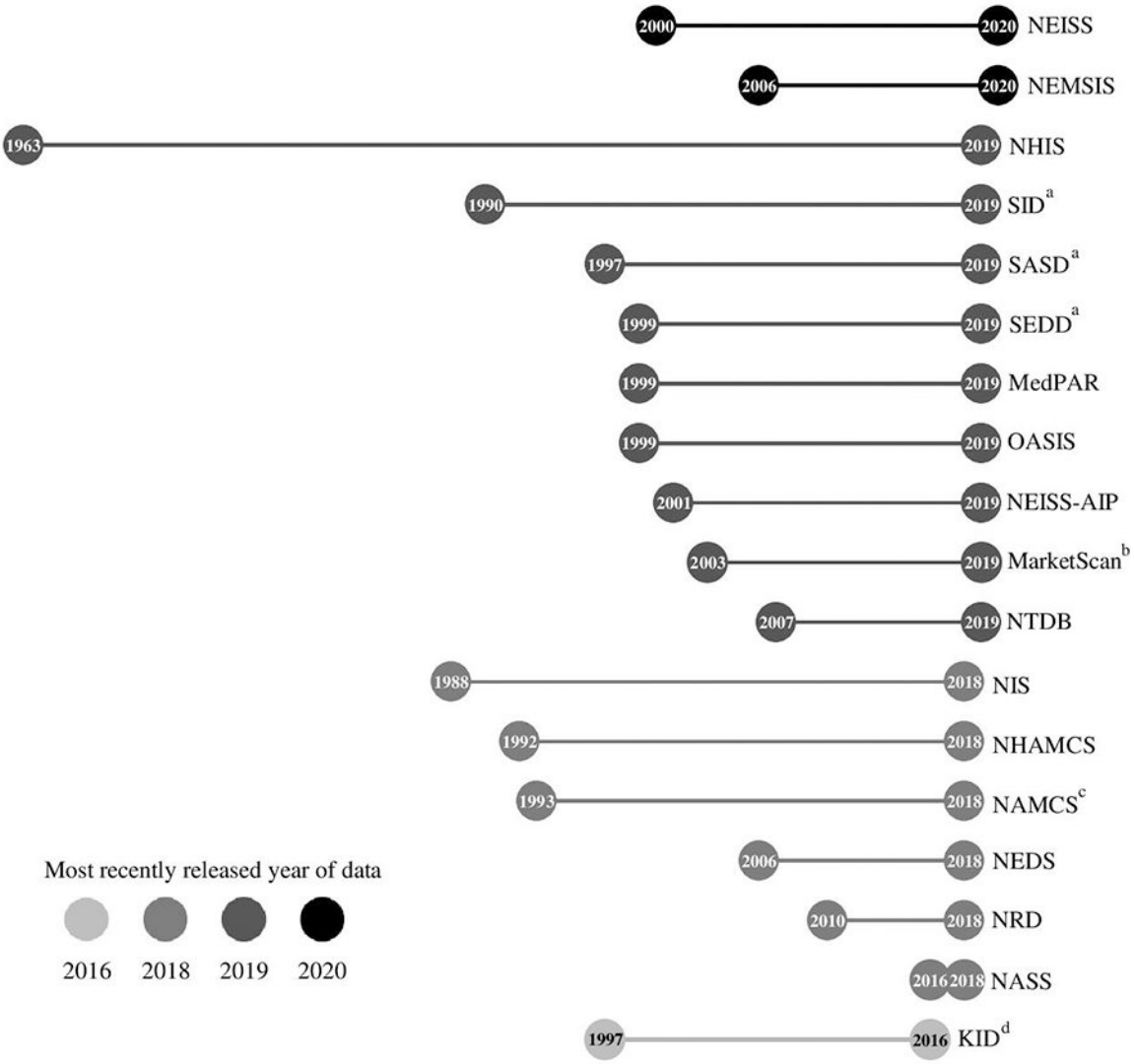

Figure 1.

Data availability and years for non-fatal injury data systems. Note: Availability is current as of August 2021. Data systems are as follows: KID, Kids' Inpatient Database; MedPAR, Medicare Provider Analysis Review Limited Data Sets; NAMCS, National Ambulatory Medical Care Survey; NEDS, Nationwide Emergency Department Sample; NEISS, National Electronic Injury Surveillance System; NEMSIS, National Emergency Medical Services Information System; NEISS-AIP, National Electronic Injury Surveillance System-All Injury Program; NHIS, National Health Interview Survey; NIS, National (Nationwide) Inpatient Sample; NTDB, National Trauma Data Bank; MarketScan; NHAMCS, National Hospital Ambulatory Medical Care Survey; NRD, Nationwide Readmissions Database; NASS, Nationwide Ambulatory Surgery Sample; OASIS, Outcome and Assessment Information Set; SID, State Inpatient Databases; SASD, State Ambulatory Surgery and Services Databases; and SEDD, State Emergency Department Databases. (a) State availability varies

over time. (b) Refers to the MarketScan Commercial Claims and Encounters Database. (c) Data not available for 2017. (d) Available 1997, 2000, 2003, 2006, 2009, 2012, 2016.

Inj Prev. Author manuscript; available in PMC 2023 June 01.

Table 1

Inventory of institutional support, estimates, demographic variables, injury coding, access/format and data source for non-fatal injury data systems, USA—2018 *

Institutional support	**CPSC**	**CDC NCIPC/ CPSC**	**CDC NCHS**			**AHRQ (HCUP)**								**ACS**	**NHTSA**	**IBM**	**CMS**	
Data system	**NEISS**	**NEISS-AIP**	**NHIS**	**NHAMCS**	**NAMCS**	**SID**	**SEDD**	**SASD**	**KID** *	**NEDS**	**NIS**	**NRD**	**NASS**	**NTDB**	**NEMSIS**	**MarketScan** †	**MedPAR**	**OASIS**
Incidence estimates ‡ (n)																		
National (12)	X	X	X	X	X				X	X	X	X	X	X §		X		
Regional (2)			X	X														
State (4)			X			X	X	X										
Demographic variables (n)																		
Sex (18)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Age (18)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Race ¶ (14)	X	X	X	X	X	X		X	X		X			X	X	X	X	X
Ethnicity ¶ (11)	X	X	X	X	X	X					X			X	X		X	X
Urban/rural (11)				X	X	X	X	X	X	X	X	X	X					X
Income (9)			X			X	X	X	X	X	X	X	X					
Injury coding (n)																		
Discharge status (18)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Diagnosis codes (15)				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
External cause codes (14)				X	X	X	X	X	X	X	X	X	X	X	X		X	X
Predefined injury variables (8)	X	X	X	X	X					X				X	X			
Access/format (n)																		

Institutional support	CPSC	CDC NCIPC/ CPSC	CDC NCHS			AHRQ (HCUP)								ACS	NHTSA	IBM	CMS	
Data system	**NEISS**	**NEISS-AIP**	**NHIS**	**NHAMCS**	**NAMCS**	**SID**	**SEDD**	**SASD**	**KID***	**NEDS**	**NIS**	**NRD**	**NASS**	**NTDB**	**NEMSIS**	**MarketScan†**	**MedPAR**	**OASIS**
Restricted access files (16)			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Fee (15)			X	X	X	X	X	X	X	X	X	X	X	X		X	X	X
Data use agreement required (18)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Free online query tool (10)	X	X	X			X	X		X	X	X	X			X			
Data source (n)																		
ED visits (5)	X	X		X			X			X								
Hospitalisations (5)						X			X		X	X					X	
Medical claims (3)																X	X	X
Ambulatory surgeries (2)								X					X					
Office-based physician visits** (1)					X													
Hospital readmissions (1)												X						
Household interviews (1)			X															
EMS activations (1)															X			
Trauma admissions (1)														X				

Note: Availability of characteristics and definitions of characteristics may vary by data system, by state and over time. Additionally, data for all states are not necessarily captured in data systems that can be used to produce state estimates. Variable availability does not imply completeness. The documentation for 2018 was consulted to construct this table, unless otherwise noted. Characteristics for previous or later years may vary.

The institutional support organisations are as follows: ACS, American College of Surgeons; AHRQ, Agency for Healthcare Research and Quality; CDC, Centers for Disease Control and Prevention; CPSC, Consumer Product Safety Commission; CMS, Centers for Medicare & Medicaid Services; HCUP, Healthcare Cost and Utilisation Project; IBM, International Business Machines; NCIPC, National Center for Injury Prevention and Control; NCHS, National Center for Health Statistics; NHTSA, National Highway Traffic Safety Administration.

Data systems are as follows: KID, Kids' Inpatient Database; MarketScan[†]; MedPAR, Medicare Provider Analysis and Review Limited Data Sets; NAMCS, National Ambulatory Medical Care Survey; NASS, National Ambulatory Surgery Sample; NEDS, National Emergency Department Sample; NEISS, National Electronic Injury Surveillance System; NEISS-AIP, National Electronic Injury Surveillance System-All Injury Program; NEMSIS, National Emergency Medical Services Information System; NHIS, National Health Interview Survey; NHAMCS, National Hospital Ambulatory Medical Care Survey; NIS, National (Nationwide) Inpatient Sample; NRD, Nationwide Readmissions Database; NTDB, National Trauma Data Bank; OASIS, Outcome and Assessment Information Set; SASD, State Ambulatory Surgery and Services Databases; SID, State Inpatient Databases; and SEDD, State Emergency Department Databases.

[*] 2016 documentation was used to compile characteristics for KID since this was the most recent data available at the time of the study.

[†] Refers to the MarketScan Commercial Claims and Encounters Database.

[‡] Coding for the ability to produce estimates at the national, regional or state level was based on data system documentation. It should be noted that some data systems never had the intent to produce estimates at those levels. Further, some estimates (in particular regional or state estimates) may be unreliable due to small numbers, especially if stratified into smaller study populations.

[§] National estimates are possible through an extension of NTDB called the National Sample Program.

[¶] Exhibits a high degree of missing or unreliability across most data systems.

[**] Includes office-based physician visits and community health centre physician visits.

ED, emergency department; EMS, emergency medical services.

Inj Prev. Author manuscript; available in PMC 2023 June 01.