Guide to Statistics and Methods

Practical Guide to Surgical Data Sets: National Trauma Data Bank (NTDB)

Zain G. Hashmi, MBBS; Amy H. Kaji, MD, PhD; Avery B. Nathens, MD, MPH, PhD

Introduction

Trauma remains a leading cause of death and disability and accounts for a substantial portion of health care expenditures.[1] Therefore, research to improve trauma care is a leading public health priority.

Spearheading this effort, the American College of Surgeons Committee on Trauma coordinated a landmark multi-institutional endeavor, the Major Trauma Outcomes Study.[2] On its completion in 1989, the American College of Surgeons Committee on Trauma recognized the importance of national trauma data aggregation to inform quality improvement. In 1997, it formed a subcommittee to develop the National Trauma Data Bank (NTDB), a standardized collection of national trauma data (**Box**). Today, to our knowledge, the NTDB is the world's largest trauma data repository, with more than 7.5 million electronic records from more than 900 trauma centers.

Box. Attributes of the National Trauma Data Bank

1. The National Trauma Data Bank (NTDB) is the world's largest trauma data repository with more than 7.5 million electronic records.
2. The NTDB is considered a convenience sample because of voluntary trauma data submission.
3. The large sample size of NTDB facilitates hypothesis-generating research and the study of rare injuries, procedures, and outcomes.
4. The NTDB contains prehospital and in-hospital injury data, including anatomic and physiologic severity.
5. The NTDB does not include data on costs, laboratory values, readmissions, or long-term outcomes.
6. Careful study design, sample selection, and analytics can help mitigate NTDB limitations resulting from missing data and selection/information bias.

Data Compilation and Structure

Annually, between February and May, the NTDB collects voluntarily submitted data from individual hospitals, concordant with the National Trauma Data Standard, a set of standardized data definitions.[3] Inclusion/exclusion criteria are based on *International Statistical Classification of Diseases, Tenth Revision, Clinical Modification* diagnoses codes and certain admission characteristics.[4] The NTDB is an incident-based record; each incident is recorded independently of repeated injuries. Validation rules mitigate the submission of missing or nonsensical data. A fully deidentified data set is reported annually in compliance with the Health Insurance Portability and Accountability Act, such that no patient can be identified.

Data quality in the NTDB has improved significantly since the adoption of the National Trauma Data Standard in 2007 and the implementation of American College of Surgeons Trauma Quality Improvement Program (ACS TQIP) in 2010, which is based on NTDB data. Over time, ACS TQIP has added several fields that are only required for participating centers. Further, ACS TQIP has focused significantly on registrar education to achieve greater data standardization. The quality of data in the NTDB has improved significantly as the number of centers participating in ACS TQIP has increased. Data from ACS TQIP centers represent a large subset of NTDB data, which differ from the non-TQIP centers by virtue of (1) higher-quality data and (2) expanded fields. Researchers may access the data in the form of the research data set after completing a brief project proposal and payment of the data set fee. This includes data from all centers (TQIP and non-TQIP) but excludes the expanded fields. Investigators from TQIP centers may request the public use file, which includes only TQIP centers along with the expanded fields.

The research data set exists as a set of relational tables that can be imported using most statistical software. The user manual describes each file and its contents.[4] Because of aforementioned changes, we caution against combining 2002 to 2006 data with later years and recommend using data from 2007 to the present.

Methods

Retrospective, cross-sectional and matched case-control study designs have commonly been used to analyze outcomes such as mortality, length of stay, and complications. The NTDB's large size facilitates the study of rare injuries, procedures, and outcomes.

A brief description of NTDB should be included to inform the reader about the characteristics of this data set. The rationale for using NTDB vs other available data should also be provided.

Sample inclusion/exclusion criteria should be explicitly stated. We recommend including a flow diagram of sample selection along with reasons for exclusions, documenting a stepwise derivation of the final sample.[5]

All predictor and outcome variables should be defined a priori. A justification should be provided regarding categorizing continuous variables. When studying mortality, at a minimum age, hypotension, pulse, total Glasgow Coma Scale score, the mechanism of injury, and injury severity score should be considered.[6]

Mortality is an important end point for trauma and therefore requires diligence to ensure that ascertainment is adequate. Currently, data capture ceases at the time of discharge and deaths occurring after administrative discharge are not identified. For example, if a patient with severe traumatic brain injury was transferred to hospice (even in the same institution), he or she would first be discharged and then readmitted to hospice. For this reason, investigators have recommended that transfers-to-hospice

be treated as deaths in this data set.[7] Additionally, patients arriving with no signs of life (dead on arrival) may need to be excluded depending on the nature of the analysis.[4]

Standard statistics should be used to report descriptive, univariate, and multivariate analyses. For risk-adjusted analyses, we recommend specifying the model type (eg, logistic or Poisson regression) with a description of its theoretical underpinnings/assumptions. Variable selection should be based on prior evidence and biological/clinical plausibility.[6] If selection is based on statistical significance criteria, the model should be presented as being hypothesis-generating rather than conclusive. Additionally, model performance statistics and whether multicollinearity and effect modification were assessed should be specified. If data include a facility identifier, hierarchical analyses should be used to account for correlated patient outcomes, as patients are nested within facilities.

Limitations

The NTDB is a convenience sample of voluntarily submitted data and therefore is not nationally representative. However, virtually all level I/II trauma centers now submit data to the NTDB and the data can be considered nationally representative of those facilities. The NTDB does not include data on costs, laboratory values, and long-term outcomes including readmissions and functional outcomes.

Selection bias refers to the differences between groups due to differences in inclusion. Two important sources of selection bias in the NTDB arise because of differences in hospital-level inclusion of isolated hip fractures and transferred patients.[8] Researchers should be cognizant of these and other distinct patient subpopulations that can alter their results. Strategies to mitigate this bias may involve including an indicator variable in the model or using sensitivity analyses.

Information bias refers to differences between groups due to differential data availability. This may be caused by missing data or variations in data abstraction. For example, an injury severity score is manually abstracted and depends on diagnostic aggressiveness. Hospitals with more liberal imaging protocols may detect more incidental, clinically insignificant injuries and subsequently report higher injury severity scores than other centers. Further, patients who die shortly after arrival to the emergency department, before operative or radiologic evaluation, will have inadequate injury ascertainment, and severity scores might be challenging to interpret. Additionally, injury severity scores may be recorded differently depending on which registry software was used. To mitigate this, NTDB now requires the use of the Abbreviated Injury Scale 2005 to standardize injury severity data.[4]

Most of the demographic and injury data included in the NTDB are considered robust. However, certain data, like emergency medical services parameters, comorbidities, and complications, may not be adequately captured, especially at some trauma centers. As previously mentioned, missing data in the NTDB can pose a significant challenge. We advise researchers to undertake thorough exploratory analyses to understand and describe patterns of "missingness" (missing completely at random, missing not at random). Strategies to mitigate these include restricting the analyses to hospitals with known superior data-quality (level I/II trauma centers, centers reporting >100 cases per year and/or at least >1 outcome of interest) and using multiple imputation techniques.[9]

Recommended Reading

The NTDB Data Manual is a comprehensive resource providing detailed insight into several methodological issues highlighted here.[4] Additionally, several high-impact publications, such as those by Galvagno[5] and Haider et al,[6] serve as excellent examples of appropriate data use and presentation.[10]

Conclusions

The NTDB is a powerful repository providing increasingly granular insight into trauma care and is considered a robust hypothesis-generating resource to guide future research. An in-depth understanding of its characteristics is essential to harness its true potential.

ARTICLE INFORMATION

Author Affiliations: Center for Surgery and Public Health, Harvard Medical School, Harvard T. H. Chan School of Public Health, Department of Surgery, Brigham & Women's Hospital, Boston, Massachusetts (Hashmi); Department of Surgery, Sinai Hospital of Baltimore, Baltimore, Maryland (Hashmi); Department of Emergency Medicine, Harbor-University of California–Los Angeles Medical Center, Torrance (Kaji); Statistical Editor, *JAMA Surgery* (Kaji); Department of Surgery, Sunnybrook Health Sciences Center, University of Toronto, Toronto, Ontario, Canada (Nathens); American College of Surgeons, Chicago, Illinois (Nathens).

Corresponding Author: Zain G. Hashmi, MBBS, Center for Surgery and Public Health, Department of Surgery, Brigham & Women's Hospital, 1620 Tremont St One Brigham Cir, 4th Floor, Ste 4-020, Boston, MA 02120 (mzhashmi@bwh.harvard.edu).

Published Online: April 4, 2018. doi:10.1001/jamasurg.2018.0483

Conflict of Interest Disclosures: Dr Nathens works for the American College of Surgeons as the director of the Trauma Quality Improvement Program. No other disclosures were reported.

REFERENCES

1. Rhee P, Joseph B, Pandit V, et al. Increasing trauma deaths in the United States. *Ann Surg*. 2014;260(1):13-21.

2. Champion HR, Copes WS, Sacco WJ, et al. The Major Trauma Outcome Study: establishing national norms for trauma care. *J Trauma*. 1990;30(11):1356-1365.

3. American College of Surgeons. National trauma data standard. https://www.facs.org/~/media/files/quality%20programs/trauma/ntdb/ntds/data%20dictionaries/ntds%20data%20dictionary%202018.ashx. Accessed July 24, 2017.

4. American College of Surgeons. National Trauma Data Bank research data set user manual and variable description list. https://www.facs.org/~/media/files/quality%20programs/trauma/ntdb/ntdb%20rds%20user%20manual%20all%20years.ashx. Accessed July 24, 2017.

5. Galvagno SM Jr, Haut ER, Zafar SN, et al. Association between helicopter vs ground emergency medical services and survival for adults with major trauma. *JAMA*. 2012;307(15):1602-1610.

6. Haider AH, Hashmi ZG, Zafar SN, et al. Developing best practices to study trauma outcomes in large databases: an evidence-based approach to determine the best mortality risk adjustment model. *J Trauma Acute Care Surg*. 2014;76(4):1061-1069.

7. Kozar RA, Holcomb JB, Xiong W, Nathens AB. Are all deaths recorded equally? The impact of hospice care on risk-adjusted mortality. *J Trauma Acute Care Surg*. 2014;76(3):634-639.

8. Gomez D, Haas B, Hemmila M, et al. Hips can lie: impact of excluding isolated hip fractures on external benchmarking of trauma center performance. *J Trauma*. 2010;69(5):1037-1041.

9. Oyetunji TA, Crompton JG, Ehanire ID, et al. Multiple imputation in trauma disparity research. *J Surg Res*. 2011;165(1):e37-e41.

10. Haider AH, Hashmi ZG, Zafar SN, et al. Minority trauma patients tend to cluster at trauma centers with worse-than-expected mortality: can this phenomenon help explain racial disparities in trauma outcomes? *Ann Surg*. 2013;258(4):572-579. doi:10.1097/SLA.0b013e3182a50148