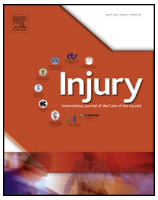
Injury, Int. J. Care Injured 47 (2016) 559–567

Contents lists available at ScienceDirect

Injury

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



Review

Classifying, measuring and improving the quality of data in trauma registries: A review of the literature

Gerard M. O'Reilly[a,b,*], Belinda Gabbe[a,c], Lynne Moore[d], Peter A. Cameron[a,b,e]

[a] Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Commercial Rd, Melbourne, 3004, Australia
[b] Emergency and Trauma Centre, Alfred Health, Commercial Rd, Melbourne, Victoria, 3004, Australia
[c] Swansea University, United Kingdom
[d] University of Laval, Quebec, Canada
[e] Emergency Medicine, Hamad Medical Corporation, Doha, Qatar

ARTICLE INFO

Article history:
Accepted 9 January 2016

Keywords:
Trauma registry
Trauma database
Trauma databank
Injury registry
Injury database
Injury databank
Data quality
Accuracy
Completeness
Capture

ABSTRACT

Introduction: Globally, injury is a major cause of death and disability. Improvements in trauma care have been driven by trauma registries. The capacity of a trauma registry to inform improvements in the quality of trauma care is dependent upon the quality of data. The literature on data quality in disease registries is inconsistent and ambiguous; methods used for classifying, measuring, and improving data quality are not standardised. The aim of this study was to review the literature to determine the methods used to classify, measure and improve data quality in trauma registries.

Methods: A scoping review of the literature was performed. Databases were searched using the term "trauma registry" and its synonyms, combined with multiple terms denoting data quality. There was no restriction on year. Full-length manuscripts were included if the classification, measurement or improvement of data quality in one or more trauma registries was a study objective. Data were abstracted regarding registry demographics, study design, data quality classification, and the reported methods used to measure and improve the pre-defined data quality dimensions of accuracy, completeness and capture.

Results: Sixty-nine publications met the inclusion criteria. Four publications classified data quality. The most frequently described methods for measuring data accuracy ($n = 47$) were checks against other datasets ($n = 18$) and checks of injury coding ($n = 17$). The most frequently described methods for measuring data completeness ($n = 47$) were the percentage of included cases, for a given variable or list of variables, for which there was an observation in the registry ($n = 29$). The most frequently described methods for measuring data capture ($n = 37$) were the percentage of cases in a linked reference dataset that were also captured in the primary dataset being evaluated ($n = 24$). Most publications dealing with the measurement of a dimension of data quality did not specify the methods used; most publications dealing with the improvement of data quality did not specify the dimension being targeted.

Conclusion: The classification, measurement and improvement of data quality in trauma registries is inconsistent. To maintain confidence in the usefulness of trauma registries, the metrics and reporting of data quality need to be standardised.

Contents

---

[*] Corresponding author at: Emergency and Trauma Centre, The Alfred, Commercial Rd, Melbourne, Victoria 3004, Australia. Tel.: +61 438 338 414.
E-mail address: gerard.oreilly@monash.edu (G.M. O'Reilly).

http://dx.doi.org/10.1016/j.injury.2016.01.007

Introduction

Injury is a major cause of disability and death. The burden of injury is of particular concern amongst young and working-age adults and in developing countries [1]. In developed countries, trauma care systems have reduced morbidity and mortality from injury [2–8]. Improvements in the delivery of trauma care have been driven by the information contained in trauma registries, which are dedicated to storing data on the outcomes and processes of trauma care [9]. Trauma registries have played a very important role in developed trauma systems for more than 30 years [10]. Conversely, the use of trauma registries in developing countries, where the burden of injury is most prominent, remains sporadic [11,12].

The capacity of a trauma registry to inform improvements in the quality of trauma care depends upon the quality of its data. Incomplete and erroneous data are a threat to the use of trauma registries for comparing and benchmarking systems of trauma care [10,13]. If the quality of data in a trauma registry is unknown, questionable or poor, it will be rendered less valuable as a tool for improving the quality of trauma care. Systems of trauma care, at the single-hospital or multi-hospital level, need methods to measure and improve the quality of new or established trauma registries; users of observational research based on registry data need to be confident of the quality of the source data [14].

For trauma registries, there is no single definition or classification of data quality. Amongst datasets in general, the taxonomy of data quality varies considerably according to the type and intended use of the dataset being considered [15–21]. For example, one publication defined data quality as "fitness for purpose" and listed its six dimensions as: relevance, accuracy and reliability, timeliness, accessibility, interpretability and coherence [15]. Another publication listed seven dimensions of data quality: valid, complete, consistent, unique, timely, accurate and precise [16]. Wang and Strong, in their iterative approach to developing a framework of data quality grouped a total of 15 dimensions under four categories of data quality: intrinsic, contextual, representational and accessibility [17].

The literature dealing with data quality amongst *health*-related datasets is similarly inconsistent [18,19]. Several publications have provided a relatively simple set of definitions and classification for data quality [20,21]. Sorensen et al. included accuracy and completeness of variables, and completeness of registration, amongst their list of dimensions [20]. Similarly, Arts et al. noted that the two most cited data quality attributes are accuracy (the extent to which registered data are in conformity to the truth) and completeness (the extent to which all necessary data that could have been registered have actually been registered) [21].

The aim of this study was to review the accessible literature to determine the methods used to classify, measure and improve data quality in trauma registries.

Methods

A scoping review of the literature was performed. Abstracts were sourced by searching MEDLINE, Embase and CINAHL databases on 12 February 2015. There were no year restrictions placed on the search. Key words used for the search were: "trauma registry", "trauma registries", "trauma database*", "trauma databank*", "injury registry", "injury registries", "injury database*", and "injury databank*". An additional subject heading search was undertaken in: MEDLINE using "Wounds and Injuries" AND "Registries"; Embase using "Registry" AND "Injury" OR "Trauma"; and CINAHL using "Registries, Trauma". Finally, in each database, the results of the search strategy described above were combined (using "AND") with each of the keywords denoting data quality and its dimensions in previous literature [15–21] (see Table 1).

From the results of the search, the authors included those full-length manuscripts which met the following criteria:

1. The trauma registry was a dedicated database, prospectively collecting data regarding injured patients at one or more health facilities, and
2. The classification, measurement or improvement of data quality in one or more trauma registries was a study objective.

Data on the following were abstracted from each full-text manuscript:

1. Characteristics of the trauma registry (source country, level of country development, jurisdiction, stage of registry existence) [22].
2. Characteristics of the publication (year published, study design).
3. Did the publication provide a classification of data quality?
4. Did the study of trauma registry data quality deal with measuring data quality *or* improving data quality *or* both?
5. What dimensions of trauma registry data quality were named?
6. For those publications dealing with measuring data quality, what dimensions of data quality were measured, how were the measurements defined and what specific metrics were used?
7. For those publications dealing with the improvement of data quality, what dimensions of data quality were targeted and what domains of trauma registry function were described?

For the final question of the trauma registry domains described in publications looking at *improving* data quality (7), a modified version of the Trauma Registry Assessment Tool was used [11,12,23]. The details of this tool and its validation have been published elsewhere [12]. Under the broad headings of physical resources, human resources and processes it comprises a checklist

Table 1
Dimensions of quality used in search terms.

Quality	Correctness	Comparability
Accuracy	Consistency	Uniqueness
Completeness	Usability	Availability
Capture	Relevance	Representativeness
Coverage	Timeliness	Case ascertainment
Missing	Accessibility	Error
Reliability	Interpretability	concordance
Validity	Coherence	precision
	Verification	