Unintentional injuries are a significant cause of morbidity in infants (children ≤12 months of age) in Australia with a rate of hospitalisation of 839/100,000 male infants and 757/100,000 female infants in 2011/12. Falls represent the main cause of hospitalised injuries in infants accounting for around half of the injury hospitalisations in this age group in 2011/12, and about 10.6% of infants and toddlers (under 4 years of age) experienced a fall in the previous four weeks, according to a 2001 Australian Health Survey. Many falls in this age group are associated with consumer products, and the main products implicated in falls in infants are products that are designed for use by/with infants. The most common products associated with infant falls are prams, cots, highchairs, baby walkers, change tables and bouncers.

In Australia, there are limited injury data that allow analysis of the frequency, severity, injury mechanisms, product association and risk profiles for non-fatal injuries in infants. National hospitalisation data are available in a standardised format that enables serious injury trend reporting on broad mechanisms of injury. However, reporting of infant product association is limited by the narrow range of infant product codes available: prams (W02.7), cots (W06.2), high chairs (W07.4), baby walkers (W02.8) and change tables (W08.0). Hospitalisations represent the tip of the iceberg of injury burden and are not representative of the range of mechanisms/products involved in lower severity injuries, therefore injury prevention policy and programs need to be informed by more than just hospitalisation data.

Routine collected emergency department data that captures presentations at public emergency departments throughout Australia are available in an un-standardised format across jurisdictions. Data are recorded in a mixture of coded (usually a subset of ICD-10-AM diagnosis codes) and text data fields. Injury causation information (i.e., external cause) is not standardised, being captured as free-text and therefore to identify causes of injuries and product involvement, interrogation of the text field is required. Special emergency department injury surveillance data are available for only two of the eight states/territories: Queensland Injury Surveillance Unit (QISU) and Victorian Injury Surveillance Unit (VISU). While there are a few hospital-based injury surveillance systems operating in other states, these data are not routinely aggregated or compared. Given the limitations of coded data, one of the key data fields of use in Australian injury surveillance systems is the injury description text field, a short non-standardised text field

Objective: To explore the potential for using a basic text search of routine emergency department data to identify product-related injury in infants and to compare the patterns from routine ED data and specialised injury surveillance data.

Methods: Data was sourced from the Emergency Department Information System (EDIS) and the Queensland Injury Surveillance Unit (QISU) for all injured infants between 2009 and 2011. A basic text search was developed to identify the top five infant products in QISU. Sensitivity, specificity, and positive predictive value were calculated and a refined search was used with EDIS. Results were manually reviewed to assess validity. Descriptive analysis was conducted to examine patterns between datasets.

Results: The basic text search for all products showed high sensitivity and specificity, and most searches showed high positive predictive value. EDIS patterns were similar to QISU patterns with strikingly similar month-of-age injury peaks, admission proportions and types of injuries.

Conclusions: This study demonstrated a capacity to identify a sample of valid cases of product-related injuries for specified products using simple text searching of routine ED data.

Implications: As the capacity for large datasets grows and the capability to reliably mine text improves, opportunities for expanded sources of injury surveillance data increase. This will ultimately assist stakeholders such as consumer product safety regulators and child safety advocates to appropriately target prevention initiatives.

Key words: infants, injury surveillance, product safety, emergency department, text searching

Abstract

© 2015 Public Health Association of Australia

Aust NZ J Public Health. 2015; Online; doi: 10.1111/1753-6405.12466

1. Queensland University of Technology, Centre for Accident Research and Road Safety – Queensland
2. Queensland Injury Surveillance Unit, Mater Health Services, Queensland

Correspondence to: Dr Kirsten Vallmuur, Centre for Accident Research and Road Safety – Queensland, School of Psychology and Counselling, Faculty of Health, Queensland University of Technology, Kelvin Grove, Queensland 4059; e-mail: k.mckenzie@qut.edu.au

Submitted: March 2015; Revision requested: May 2015; Accepted: June 2015

The authors have stated they have no conflict of interest.
that describes how the injury occurred. This injury description can be used to identify and code products, validate coded data, and provide context to how an injury occurred. In Australia, routine emergency data also contains a free text description of the presenting problem. The presenting problem field has a standardised national definition: “The clinical interpretation of the problem or concern that is identified by the triage clinician as the main reason for the person’s non-admitted patient emergency department service episode.” This presenting problem data auto-populates the injury description field for injured patients presenting to sites that collect VISU/QISU injury surveillance data. Therefore, the presenting problem field in the broader ED data is another avenue for exploring injury mechanism and product association. A recent data scoping exercise of Australian emergency departments by the author of this paper found that at least six of the eight states/territories routinely record the presenting problem text data for most of their EDs, although there are different levels of availability of text data for research and variability around whether information is stored locally at hospitals or centrally with health departments.

Emergency department-based injury text descriptions are regularly used in many jurisdictions throughout the world and are increasingly being cited in research papers as a valuable tool for injury surveillance.10 There has been limited research conducted in Australia assessing the utility of the presenting problem text alone (that is, where the text field is not part of a broader specialised injury surveillance data collection such as VISU/QISU) for specific injury surveillance purposes, with the few studies that have been conducted focusing on sports injury11 and alcohol identification.12 Given that these text data are routinely recorded relatively widely throughout Australia and are not onerous in terms of data collection demands on clinical staff, evaluating methods for searching these data for specific injury surveillance purposes is valuable to inform the future development and use of ED text for injury surveillance in Australia.

While there are a range of natural language processing and machine learning approaches to processing text-based data,16 many users of routine emergency department data (such as clinicians, government agencies (such as product safety regulators), community injury prevention advocacy groups, etc) may not have the technological skills or resources to utilise complex computer algorithms to identify cases of interest. Often, the aim in these cases is simply to identify a cohort of patients for follow-up or to gather broad descriptive data about age groups, frequency of presentations and severity of injuries. As such, it is important to assess the potential to use simple, easily applied text search techniques to identify cases from these unstructured data sources for the average user. To identify high risk infant products, prioritise child injury prevention policies and response and monitor injury trends requires injury data that accurately identifies cases and circumstances, and which is widely available. This study aimed to explore the potential for using a basic text search of routinely collected emergency department data to identify infants injured while using particular products, and to compare the patterns of injury from routine ED data and specialised injury surveillance data.

Methods

Population
The population were all infants aged 12 months and younger who presented for treatment of an injury at an emergency department in Queensland.

Data sources
Data was sourced from the Queensland Emergency Department Information System (EDIS) and the Queensland Injury Surveillance Unit (QISU). EDIS data is collected for all patients in Queensland who attend EDs who use the EDIS software system and the data is estimated to cover 75% of ED presentations in Queensland.16 QISU data contains injury surveillance information collected from persons presenting with an injury or – in the case of children – from the accompanying adult, for a sample of EDs in Queensland and the data is estimated to cover about 25% of ED presentations in Queensland.17

Variables
For EDIS, infants were included in the data collection if the discharge diagnosis was an injury-related diagnosis code (ICD-10 code range S00-T79, excluding complications of medical/surgical care) or one of the limited range of unintentional external cause codes used in the discharge diagnosis field of the EDIS system (V00-X59). The EDIS is designed to capture information about the patient, and their diagnosis, treatment and movement through the system, and is not designed to collect detailed injury surveillance data. Relevant injury data collected in EDIS includes the date and time of presentation, patient demographics, the presenting problem, nurse assessment, triage category, disposition and diagnostic code.

QISU is designed to capture level 2 National Data Set for injury surveillance (NDSIS) data in addition to the routinely collected emergency data outlined above. QISU data is collected by the ED triage nurse and includes: cause of injury, mechanism of injury, place and part of place (where the injury occurred), activity of the injured person, the main object or substance involved in the injury and human intent. QISU sites use a combination of text-based and coded fields to enter data, depending on the patient management system used. The Injury Description text field within the QISU data is a short free text field (maximum 255 characters) that (at EDIS sites) autopopulates from the EDIS ‘presenting problem’ field. The triage nurse then has the option of adding additional information if they deem it to be necessary. QISU data is subsequently validated by trained coders to ensure the text and coded fields are consistent.

Study design
QISU data was analysed to identify the top five infant products recorded in the major injury factor codes. The top five infant products identified were consistent with international trends and were, in order: prams, change tables, highchairs, cots and baby walkers. A list of potential reliable text search terms was developed by reviewing text terms used by triage staff for each product group. This process involved identification of common stems, misspellings and alternative names for coded products. The final list of search terms after refinement and review ranged from single terms for some products to multiple terms for other products as follows: Cot (searched for cot), Pram (searched for pram, strol, stoller, pusher, pushchair), Change table (searched for change table, changetable, change mat, changing mat, change drawer), Highchair (searched for high chair, highchair, hgh chair), Walker (searched for walker). The refined list was then applied to EDIS data using text search syntax in SPSS to
flag relevant cases. All cases were manually reviewed by the author to evaluate whether the case was a true or false positive, and cases deemed to be a false positive were removed from further analysis.

Statistical analysis

From the first phase QISU analysis, sensitivity, specificity and positive predictive value were calculated for the five products identified in QISU data using the text field search as the test results and the coded product data field (NDSIS Major Injury Factor) as the gold standard results.

The specificity and positive predictive value were calculated for the text search strategy developed for each of the five products identified in EDIS data using the text field search as the test results and the manually reviewed revised data as the 'surrogate' gold standard results for a 'true case' where the text search was deemed to accurately flag a case as having the product of interest. Sensitivity of the EDIS search could not be calculated, as there was no way of knowing the exact number of true positives in the data that weren't identified by the text search.

Descriptive analysis of the top five products using both QISU and EDIS data was conducted to examine the frequency of presentations, age (in months)/sex differences, and outcomes between the two datasets.

Ethics statement

This research study was approved by the University Human Research Ethics Committee and access to data was approved by Queensland Health's Human and Medical Research Unit.

Results

Sample characteristics

Between 2009 and 2011 there were 10,250 injury presentation for infants aged 12 months and younger recorded in Queensland EDIS data and 3,647 injury presentation recorded in QISU data, with about 54% male and 46% female in both datasets. The number of presentations increased uniformly across months of age with the most common age of presentation around 11–12 months across both datasets representing a quarter of infants in this group (Table 1).

Validity of QISU search algorithm

The sensitivity, specificity, and positive predictive value were calculated for the five products identified in QISU data (Table 2). All of the searches showed high sensitivity (albeit acknowledging the limitations noted in the methodology section for sensitivity estimates) and high specificity. All searches except for Cots showed high positive predictive value. Just over 25% of cases identified using the search for 'cot' were false positives with the word stem 'cot' a common stem of other words (e.g. 'cotton'), however including spaces in the search (e.g. 'cot ') compromised the sensitivity of the search missing cases where the word appeared at the start or end of the phrase or where it was included as part of 'portacot'. Hence no changes were made to the approach to identify cot-related injuries for the next phase of analysis given a manual review of all identified EDIS cases was to be conducted.

EDIS search results

The specificity and positive predictive value for the search were calculated for the five products identified in EDIS data (Table 3). All of the searches showed high specificity. Similar to the QISU analysis, all searches except for Cots showed high positive predictive value. Around 27% of cases identified using the search for 'cot' were false positives with the word stem 'cot' a common stem of other words.

Comparison of patterns using QISU and EDIS data for products

False positives were removed and infants presenting with an injury related to one of the five products were compared between QISU and EDIS data in regards to infant age distributions, nature of injury and admission outcomes. EDIS patterns were similar to QISU patterns for all of these aspects as shown in Figures 1 to 8. Figures 1 to 6 show

---

Table 1: Number of injury presentations in infants by month of age recorded in QISU and EDIS.

| Infant age | QISU 2009 | EDIS 2009 | QISU 2010 | EDIS 2010 | QISU 2011 | EDIS 2011 | QISU TOTAL | EDIS TOTAL |
|------------|-----------|-----------|-----------|-----------|-----------|-----------|------------|------------|
| < 1 month  | 12        | 104       | 11        | 95        | 15        | 116       | 38         | 315        |
| 1 month    | 63        | 146       | 54        | 145       | 48        | 166       | 165        | 457        |
| 2 months   | 45        | 128       | 59        | 158       | 56        | 143       | 160        | 429        |
| 3 months   | 46        | 155       | 64        | 128       | 51        | 147       | 161        | 430        |
| 4 months   | 50        | 185       | 60        | 160       | 58        | 188       | 168        | 533        |
| 5 months   | 83        | 231       | 67        | 195       | 69        | 220       | 219        | 646        |
| 6 months   | 98        | 254       | 100       | 271       | 82        | 259       | 280        | 784        |
| 7 months   | 97        | 326       | 102       | 307       | 98        | 317       | 297        | 811        |
| 8 months   | 121       | 357       | 124       | 320       | 116       | 337       | 361        | 1014       |
| 9 months   | 151       | 394       | 142       | 354       | 124       | 372       | 417        | 1120       |
| 10 months  | 147       | 395       | 144       | 357       | 123       | 369       | 414        | 1121       |
| 11 months  | 141       | 394       | 192       | 417       | 149       | 428       | 482        | 1239       |
| 12 months  | 160       | 379       | 149       | 418       | 176       | 415       | 485        | 1212       |
| TOTAL      | 1,214     | 3,448     | 1,268     | 3,325     | 1,165     | 3,477     | 3,647      | 10,250     |

Table 2: Sensitivity, specificity, and positive predictive value for five products in QISU data.

| Product | Number of true cases identified | Number of false positives | Sensitivity | Specificity | PPV | Total number of true cases |
|---------|--------------------------------|----------------------------|-------------|------------|-----|---------------------------|
| Pram    | 121                            | 1                          | 0.98        | 0.99       | 0.99| 123                       |
| Change table | 104                        | 0                          | 0.98        | 1.00       | 1.00| 106                       |
| Highchair | 76                         | 0                          | 0.95        | 1.00       | 1.00| 80                        |
| Cot     | 40                             | 14                         | 0.93        | 0.99       | 0.74| 43                        |
| Walker  | 24                             | 0                          | 1.00        | 1.00       | 1.00| 24                        |

Table 3: Specificity, and positive predictive value for five products in EDIS data.

| Product | Number of true cases identified | Number of false positives | Specificity | PPV |
|---------|--------------------------------|----------------------------|-------------|-----|
| Pram    | 252                            | 21                         | 0.99        | 0.92|
| Change table | 218                        | 0                          | 1.00        | 1.00|
| Highchair | 174                        | 2                          | 0.99        | 0.99|
| Cot     | 84                             | 31                         | 0.99        | 0.73|
| Walker  | 68                             | 1                          | 0.99        | 0.98|
strikingly similar month-of-age patterns for presentations for the selected products, with pram-related injuries peaking in the 2–3 month-olds, change table-related injuries peaking around 8 months of age, highchair-related injuries peaking around 10 months, cot-related injuries peaking around 8–10 months, and baby walker-related injuries peaking around 8–9 months. Figure 7 shows that change table-related injuries resulted in the highest proportion of cases admitted in both QISU and EDIS data, with about 28% of infants admitted when presenting with change table-related injuries. The head was the most commonly affected body region in both QISU and EDIS data, with about 28% in both QISU and EDIS data, accounting for around 70% of cot-related injuries and up to 95% of highchair-related injuries.

Conclusions

This study explored the capacity of a simple text mining strategy to be used to explore a large routinely collected emergency department dataset in order to identify specific product-related injuries in infants. This strategy was used on a coded injury surveillance dataset (where product-related injuries were coded in real time) and on the larger routine emergency department dataset and the results compared. Firstly, interrogation of QISU data identified prams, change tables, highchairs, cots and baby walkers as the most common infant products involved in injury-related presentations of infants. Furthermore, these products could be identified using fairly simple text searches that showed high sensitivity and high specificity for all of these products, and high positive predictive value (for all products except Cots) using coded QISU data as a gold standard.

While the EDIS data was considerably larger in magnitude, with the sample-based QISU system capturing around one-third of the number of cases as was captured in EDIS, the patterns by age (in months), outcomes, and head injury proportions were remarkably similar. In younger infants (2–3 months), prams were the most common product involved, while for infants aged 8–10 months, change tables, highchairs, cots and baby walkers were more commonly involved. Furthermore, in terms of severity of outcomes, change tables produced the highest number of admitted patients (28%), as well as the second highest proportion of head injuries (90% of all change table-related injuries), second only to highchairs (95% of all highchair-related injuries).

While QISU data is used as a gold standard, QISU data is a smaller (and variable) sample of ED cases than EDIS and there may be some errors of omission where a triage nurse didn’t document or code the products involved in the incident. However, it is not possible to know to what extent this occurred without a larger data quality audit. QISU does undertake training and quality assurance of their data on a regular basis, and data that is collected is likely to be comparable to other systems that collect injury data in busy emergency departments; hence, findings
in relation to the products in this study are likely to be generalisable to other comparable systems.

**Implications**

This research has implications for organisations involved in identifying and investigating consumer product-related injuries and risks such as regulators and standards developers, as well as those groups involved in the prevention of child injury, such as health departments, child care agencies and child safety advocates.

One aspect of interest could be the potential to use existing data in Australia (both retrospectively and prospectively) to explore product-related injury patterns for other products of concern, as well as conducting more detailed investigations of the injury patterns for these five products in other states and territories. Using simple text search approaches to flag relevant cases could lead to follow-up studies with patients injured using products of interest to gather more detail regarding the product itself to assist regulators to target safety initiatives.

Secondly, it is important to note that only three of the five products investigated have mandatory standards (prams, cots and baby walkers) regulating the safe design of products sold in Australia. Pram-related injuries showed the highest number of presentations of all of these products, followed by the two unregulated products (change tables and highchairs). Investigation of the common patterns of injury over time (particularly those that relate to changes in product design or manufacturing or consumer use of products) could contribute to safer design solutions.

Thirdly, the main period of concern within the infant age groups was between 8 and 10 months of age. Further investigation of parents’ understanding of development and risks at different age groups and approaches to supervision across infancy and toddlerhood is needed to identify opportunities for prevention campaigns.

Specialised injury surveillance systems only operate in two states and a small number of individual hospitals throughout Australia. This study demonstrated a capacity to identify a sample of valid cases of product-related injuries for specified products using simple text searching of emergency presentation data that is routinely available for most states and territories. Furthermore, the patterns identified using data obtained via this simple method was extraordinarily consistent with data collected within the specialised injury surveillance system. Nevertheless, the need for specialised injury surveillance systems are still apparent to enable the identification of an inclusive range of search terms (including brand names for products to improve search algorithm sophistication), early detection of new brand names/terms, and refinement of terms with high sensitivity but low specificity. Furthermore, specialised injury surveillance collections (even if only sample-based collections) remain a vital part of the system to ensure continued advocacy for the important role of emergency clinical staff in injury prevention through maintaining a focus on the need for clear documentation and accurate coding of injury circumstances during emergency presentations.

As the capacity for storage and interrogation of large datasets grows, and the capability
to reliably mine unstructured text data improves, opportunities for expanded sources of injury surveillance data increase. Resources for collection of ‘non-essential’ prevention-focused injury surveillance data in the clinical setting continue to shrink; hence, it is critical that we look for complementary methods for gathering injury surveillance data into the future.

Acknowledgements

This work was supported by the Australian Research Council Future Fellowship under Grant FT120100202.

References

1. Pointer S. Hospitalised Injury in Children and Young People 2011–12. Injury Research and Statistics Series No: 91. Canberra (AUST): Australian Institute of Health and Welfare; 2014.
2. Australian Bureau of Statistics. National Health Survey: Summary of Results 2001. Canberra (AUST): ABS; 2002.
3. Sengölge M, Vincenten J. Child Safety Product Guide: Potentially Dangerous Products. Birmingham (UK): European Child Safety Alliance, EuroSafe; 2013.
4. Watson W, Routley V, Oztanne-Smith J. Nursery Furniture Injuries. Vic Inj Surveillance Unit Hazard Bull. 1998:37.
5. Krahin D, Barker R, Spinks D, Pitt R. Injuries under 12 months. Qld Inj Surveillance Unit Inj Bull. 2007:98.
6. McKenzie K, Scott D, Limbong J, Li E. Feasibility of Using Health Data Sources to inform Product Safety Surveillance in Queensland. A Report for the Queensland Injury Prevention Council.Brisbane (AUST): Queensland University of Technology National Centre for Health Information Research and Training; 2011.
7. Mack KA, Gilchrist J, Ballesteros MF. Injuries Among Infants Treated in Emergency Departments in the United States, 2001-2004. Pediatrics. 2008;121(5):930-7.
8. National Casemix and Classification Centre. The International Statistical Classification of Diseases and Related Health Problems, 10th Revision. Australian Modification (ICD-10-AM). 8th ed. Sydney (AUST): University of Wollongong; 200.
9. Australian Institute of Health and Welfare. Non-admitted Patient Emergency Department Service Episode–Presenting Problem. Canberra (AUST): AIHW; 2007.
10. McKenzie K, Scott D, Campbell M, McClure R. The use of narrative text for injury surveillance research: A systematic review. Accid Anal Prev. 2010;42(2):354-63.
11. Mitchell R, Finch C, Boufous S, Browne G. Examination of triage nurse text narratives to identify sports injury cases in emergency department presentations. Int J Inj Contr Saf Promot. 2009;16(3):153-7.
12. Indig D, Copeland J, Conigrave K, Rotenko I. Why are alcohol-related emergency department presentations under-detected? An exploratory study using nursing triage text. Drug Alcohol Rev. 2006;25(6):584-90.
13. Indig D, Copeland J, Conigrave KM. Comparing methods of detecting alcohol-related emergency department presentations. Emerg Med J. 2009;26(8):596-600.
14. Indig D, Copeland J, Conigrave KM, Arcuri A. Characteristics and comorbidity of drug and alcohol-related emergency department presentations detected by nursing triage text. Addiction. 2010;105(5):897-906.
15. Vallmuur K, Limbong J, Barker R, Hides L. A comparison of methods to identify alcohol involvement in youth injury-related emergency department presentation data. Drug Alcohol Rev. 2013;32:19-26.
16. Australian Institute of Health and Welfare. Australian Hospital Statistics 2010/11. Canberra (AUST): AIHW; 2012.
17. Inoue N, Baker R, Scott D. Pram and Stroller Related Injuries in Queensland Children Under 5 years of age. Qld Inj Surveillance Unit Inj Bull. 2009;108.