Sex-specific analysis of traumatic brain injury events: applying computational and data visualization techniques to inform prevention and management

Tatyana Mollayeva1,2,3,4,5,6*, Andrew Tran4,5, Vincy Chan1,3,4,7, Angela Colantonio1,2,3,4,5 and Michael D. Escobar5

Abstract

Background: The interplay of host, agent, and environment implicated in traumatic brain injury (TBI) events is difficult to account for in hypothesis-driven research. Data-driven analysis of injury data can enable insight into injury events in novel ways. This research dissected complex and multidimensional data at the time of the TBI event by exploiting data mining and information visualization methods.

Methods: We drew upon population-based decade-long health administrative data collected through the routine operation of the publicly funded health system in Ontario, Canada. We applied a computational approach to categorize health records of 235,003 patients with TBI versus the same number of reference patients without TBI, individually matched based on sex, age, place of residence, and neighbourhood income quantile. We adopted the basic concepts of the Haddon Matrix (host, agent, environment) to organize emerging factors significantly related to TBI versus non-TBI events. To explore sex differences, the data of male and female patients with TBI were plotted on heatmaps and clustered using hierarchical clustering algorithms.

Results: Based on detected similarities, the computational technique yielded 34 factors on which individual TBI-event codes were loaded, allowing observation of a set of definable patterns within the host, the agent, and the environment. Differences in the patterns of host, agent and environment were found between male and female patients with TBI, which are currently not identified based on data from injury surveillance databases. The results were internally validated.

Conclusions: The study outlines novel areas for research relevant to TBI and offers insight into how computational and visual techniques can be applied to advance the understanding of TBI event. Results highlight unique aspects of sex differences of the host and agent at the injury event, as well as differences in exposure to adverse social and environmental circumstances, which can be a function of gender, aiding in future studies of injury prevention and gender-transformative care.

Keywords: Data-driven research, Environment, Health analytics, Human factors, Physical force agent, The Haddon matrix, Sex
health problem globally [1, 2]. It is estimated that 50–60 million people worldwide are affected by TBI annually, and it is predicted that close to 50% of the global population will sustain a TBI in their lifetime, costing the international economy approximately US$400 billion each year [1, 2]. Given an estimated standardized gross world product of US $73.7 trillion, the economic cost of TBI represents 0.5% of the annual global output [1, 3].

A majority of the TBIs sustained are mild in severity (mild TBI), occurring at an annual incidence of 224 (95% CI: 120–418) per 100,000 person-years, almost ten times higher than the incidence of moderate TBI and 17 times higher than that of severe TBI [1, 4]. Historically, TBI was viewed as predominantly a male injury sustained in war conflicts, occupational accidents, and sports; as a gender-based division of labour and leisure activities dissipate conflicts, occupational accidents, and sports; as a gender-based division of labour and leisure activities dissipate and with more attention on concussion, the prevalence of TBI in female persons increases, matching that observed in males in later life [5]. Considering the increasing burden of TBI in both male and female persons [4, 5], the World Health Organization and the federal agencies of many developed countries have identified TBI prevention as a priority, while also setting requirements for the inclusion of biological sex and sociocultural process of gender into research hypotheses [5–7]. As a result, the field of TBI has witnessed tremendous transformation and growth in both scope and depth over the past decade [5]. Among the most profound changes that have come from the research is the growing acceptance of TBI as an event that is most likely not random and is unequally distributed in the population, based on health status and risk factors preceding the injury [5, 8]. For example, the presence of substance-use disorders, epilepsy, and medication effects can potentiate or modify the risks associated with adverse behaviors (e.g., assault and domestic abuse) or falls [3, 7–9], which may affect male and female patients differently and have different implications for an injury event with a TBI outcome. Adding to these risks, many known adverse social determinants of health (e.g., socioeconomic disparity, racial and gender inequality) have been implicated in vulnerability to injury and decreased help-seeking after injury due to depleted psychosocial reserves [10–13]. A recent study highlighted that disorders of the elderly and medical issues, mental health disorders, and metabolic disorders preceding a TBI event were more frequently occurring in female patients than males; these disorders are implicated in more frequent falls due to increased frailty, especially in older persons [8]. In male patients, relatively higher occurrence of injuries from sharp instruments and moving machinery and emergencies and adversities due to substance use within five years preceding TBI [8] highlight the links to a gender-based division of labour and high risk behaviours, which might be reflected in injuries as a result of being struck by or against an object and interpersonal violence; all of which have implications for clinical and prevention contexts, and TBI surveillance system overall [4, 5].

The interplay of many clinical, social, and environmental factors implicated in injury events is difficult to account for in hypothesis-driven research. For a long time, scientists have been debating the role of hypothesis testing in injury surveillance. Such concerns reflect the centrality of hypothesis-driven research around humans’ level of expertise and knowledge, and the linear process in testing the hypothesis-driven associations. Many would argue that focusing primarily on a hypothesis-driven account diminishes the scientific curiosity and also stops scientific progress by encouraging scientists to focus on narrowly defined questions that can be posed as testable hypotheses. This issue is particularly relevant in complex injuries such as TBI, where the hypothesis-driven approach might stop new directions of analysis (e.g., the understanding the interrelatedness between personal, injury-related and environmental factors in injury or the nature-nurture interaction) and preclude identification of important relationships among many interacting variables in a non-linear fashion. Most recently, data-driven approaches to the analysis of injury data have emerged to enable understanding of TBI events in novel ways, providing evidence that TBI is caused by the interplay of three categories of factors: the host (i.e., variables characterizing the persons involved in injury with TBI outcome), the agent (i.e., an injury-producing agent that reaches the host and allows the transfer of energy in injurious amounts to cause TBI), and the environment (i.e., the context, circumstance, and conditions that may influence, directly or indirectly, the occurrence of injury, consisted of physical and social, including emergency medical services systems) [14–20], known as the epidemiologic triad in the Haddon Matrix [21–23]. The Haddon Matrix consists of three columns (host, agent, and environment) and three rows (pre-event, event, post-event phases) [21, 23]. This study focused on the event phase.

To date, most of data-driven studies that focused on the TBI event phase characterized the person (the host) who sustained the injury or how the injury-producing agent/its energy reached the host in injurious amounts (the agent of injury) as determinants of TBI outcomes (Supplementary Table 1). For example, Hernandes Rocha and colleagues (2017) studied the hospital registry data of 3138 patients utilizing a machine learning technique and reported that higher Glasgow Coma Scale (GCS) score, pulse oximetry values, and having a surgery were predictive of good recovery, while advanced patient age and injuries sustained in a motor vehicle collision (MVC)
were predictive of poor recovery [16]. Recently, Folweiler and colleagues (2020) proposed a hybrid machine learning technique to identify phenotypes of patients with TBI during the acute injury phase and reported that when patients were grouped by baseline GCS score, no differences were observed in their clinical profiles or outcomes [17], suggesting that the GCS score may not be foolproof for all TBI severities and that other factors within the human ecosystem, the agent and the environment [24], are essential to consider in TBI prevention strategies [17]. This notion is further supported by recent population-based studies which have revealed that multiple chronic disorders, alcohol and prescribed drug toxicity, and exposures to environmental adversities preceding a TBI event have implications for TBI outcomes and cost of publicly funded healthcare, affecting male and female patients differently [8, 25–27]. These differences motivated the present study’s design, which aimed to utilize the key concepts of host-agent-environment of the Hadon Matrix [21, 22] to analyze the data of patients with a diagnosis of TBI through a sex lens in the event phase.

Here, we utilized computational approaches to sequence host, agent, and environment data at the injury event from International Statistical Classification of Diseases and Related Health Problems, Tenth Revision Canadian Enhancement (ICD-10-CA) codes [28]. Given recent reports concerning sex and gender differences in TBI [8], we anticipated that data captured during the injury event will differ between the sexes, rooted in biological and social differences between male and female patients [30–32, Supplementary Table 2].

Methods
The research ethics boards approved the study protocol at the clinical (University Health Network) and academic institutions (University of Toronto) with which the authors are affiliated. We followed The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) and Sex and Gender Equity in Research (SAGER) guidelines to conduct and report our research.

Data
Data utilized in this study were acquired from ICES [29], which collects and stores health administrative data on publicly funded services provided to residents of Ontario, Canada, including information on acute care hospitalizations and emergency department (ED) visits. Ontario is Canada’s most populous province, comprising approximately 40% of Canada’s population [30]. Universal healthcare in Ontario covers all medically necessary healthcare services at the point of care [31]. The data were linked deterministically at the individual patient level through a unique, encoded identifier to other administrative data based on name, sex, date of birth, and postal code. Information in hospital and ED databases is recorded using the Canadian coding standards for the ICD-10-CA [28]. Each ICD-10-CA code consists of a combination of alphanumeric characters that characterize broad diagnosis categories arranged hierarchically, with the code length ranging from three to six characters. The first three characters designate the diagnosis category, which is the same as the World Health Organization’s ICD-10 international standard for reporting diseases and health conditions [28].

Materials
We used a previously established cohort of patients discharged from acute care (identified in the Discharge Abstract Database, DAD) and from the ED (identified in the National Ambulatory Care Reporting System, NACRS) between fiscal years 2007/2008 and 2015/2016 with a diagnostic code for TBI (ICD-10-CA codes S02.0, S02.1, S02.3, S02.7, S02.8, S02.9, S04.0, S07.1, and S06) [8, 26, 27]; these patients comprised the TBI cohort in the present study. Data on patient demographics and main and secondary diagnoses, conditions, problems, and injury circumstances were collected for each patient. We followed previously published severity classifications to assign TBI injury severity to every patient [32, 33]. External causes of injury were established using Centers for Disease Control and Prevention’s (CDC; United States) injury matrix, divided into falls, struck by/against an object, MVC, and assaults [34]. We additionally identified sports-related injuries based on the Association of Public Health Epidemiologists in Ontario (APHEO) criteria [35]. We selected a 10% random sample of patients discharged from acute care hospitals or the ED during the same study period for any reason other than TBI, individually matched to patients with TBI on sex, age, place of residence (urban vs. rural), and income quintile; these patients comprised the reference patients’ cohort.

Statistical analyses
To define the TBI event window, a histogram for the days from the index date was constructed. The index date for patients with TBI was their first TBI-related hospital and/or ED visit during the study period (fiscal years between 2007/2008 and 2015/2016). A peak was observed around the index date with the frequency dropping to a stationary point 30 days preceding, and following, the index date. This 61-day window, therefore, was determined as a TBI event phase and was the focus of this study. A similar event-phase window was defined for patients in the reference cohort, with the exception that the midpoint of each patient’s hospital visits was selected as an index date.
To analyze the health status of the host, agent, and the environment at the TBI event phase and gain insight into their observed relationships, we analyzed all ICD-10-CA codes recorded across the 10 and 25 diagnosis fields of the NACRS and the DAD respectively, at the first three-character level, for patients with TBI and their matched reference patients. We converted all ICD-10-CA codes (i.e., 2600) into binary variables, except for provisional codes for research and temporary assignments, U98 and U99. These 2600 binary ICD-10-CA code variables were tested for significant associations with TBI diagnosis codes using a matched McNemar test with correction for multiple testing [36]. The Benjamini-Yekutieli (BY) method was applied to acquire a set of codes controlled at a False Discovery Rate (FDR) of 5% [37]. For a set of ordered observed p-values, the BY method rejects k null hypotheses 
\[ H_k^0 \] where k is defined as

\[ k = \max \left\{ i : p(i) \leq \frac{i}{m \sum_{i=1}^{m} q} \right\} \] (1)

Where \( i \) is the p-value ranking, \( m \) is the total number of tests conducted, and \( q \) is the level of significance (set at 0.05 or 5% for this study).

Any of the three-alphanumeric digit ICD-10-CA codes that fulfilled the following criteria were retained for further analysis: a) identified as significant using McNemar tests following BY correction, and b) had a calculated odds ratio (OR) greater than one, compared with the reference cohort. The procedure was first performed using the training dataset and then repeated using the validation dataset. Only codes that were significant in both the training and validation sets were retained for further factor analysis using the principal components method [8, 38]. The optimal number of factors was determined by the breakpoint on the scree plot, eigenvalue, the highest cumulative proportion of variance accounted for, and via a conditional logistic regression looped across all factors covering the greatest area under the receiver operating characteristic curves [39]. Word clouds were generated to visualize the sex-specific frequencies, proportion, and ORs of the factors, where the size of the words indicated the magnitudes of these values [40].

Each of the factors at the injury event was defined as host, agent, or environment-related by studying its frequencies, proportion, and ORs of the sex-specific associations, U98 and U99. These factor-based scores have been described elsewhere [44, 45].

The conditional logistic regression models were then built using designated binary factor-based scores [43]. Male and female patients were assigned a score of one if they possessed any of the ICD-10-CA codes contained in the factor definition (See Supplementary Table 2); otherwise, they were assigned a zero. These factor-based scores were used to calculate ORs and 95% confidence intervals from a conditional logistic regression for the association between TBI and each factor, adjusting for age, rurality and income quantile in the training dataset first, and then repeated in the testing and validation datasets. Pearson’s correlation matrix was generated for all significant factors in all three datasets to identify groups of male and female patients with similar associative factors, and hierarchical clustering was performed on similar group factors using correlation-based distance. To aid in the visualization of clusters of factors at the event phase, we used heatmaps, where clustering was performed using Ward (minimum variance) linkages. The dendrogram/clustering used to group the factors was based on our previous work [8]. The algorithms for these agglomerative clustering methods have been described elsewhere [44, 45].

To visualize sex-specific differences in the host, agent, and environment factor correlations, raw correlation values were plotted onto heatmaps. To remove noise from weaker correlations, hypothesis tests were conducted using Fisher-transformed correlations to identify a set of significant correlations [46, 47]. All correlations (r) were transformed using the following formula:

\[ r' = \frac{1}{2} \ln \left( \frac{1 + r}{1 - r} \right) \] (2)

The resulting transformed value \( r' \) has variance \( \frac{1}{n-3} \) and quickly converges to normal, a property which is useful for conducting z-tests. Three sets of hypotheses tests were conducted to generate heatmaps for males \( \left( H_M \right) \)
females $\left( H^0_F \right)$, and the differences between both sexes
$\left( H^0_{F-M} \right)$:

$$H^0_M : r' = 0$$  \hspace{1cm} (3)

$$H^0_F : r' = 0$$  \hspace{1cm} (4)

$$H^0_{F-M} : r'_F - r'_M = 0$$  \hspace{1cm} (5)

For the first two hypotheses (3) and (4), the test statistic was calculated as

$$z = \frac{r'}{\sqrt{\frac{1}{n-3}}}$$  \hspace{1cm} (6)

For hypothesis (5), the test statistic was calculated as

$$z = \frac{r'_F - r'_M}{\sqrt{\frac{1}{n^F-3} + \frac{1}{n^M-3}}}$$  \hspace{1cm} (7)

where the resulting statistic comes from a normal distribution $N(0,1)$ which can be used to calculate $p$-values. Multiple testing correction (i.e., the BY method) was then used as previously described to obtain a set of $p$-values with the FDR controlled at 5%.

Finally, the host, agent, environment-related variables were placed on the association matrix for TBI severity, and the external cause of injury as designated by the CDC [7]. Using the same procedure outlined above, sex-specific heatmaps were generated using Fisher transformation, z-tests, and multiple hypothesis testing correction to control the FDR at 5%. SAS software (SAS Inc., Cary, NC version 9.410) and R (R Foundation for Statistical Computing version 3.4.1.11.) were used for data analyses. Figures were created using R (ComplexHeatmap and Wordcloud, R Foundation for Statistical Computing; www.r-project.org).

Results

Among the total Ontario population of 12.9 and 14 million in 2008 and 2016, respectively [30], 239,103 unique patients had a TBI-related visit to either an ED or acute care hospital between the years 2007/08 and 2015/16. Every patient with TBI was matched to a patient from the 10% random sample of patients entering an ED or acute care hospital for any reason but TBI; 4100 (1.7%) patients were left unmatched and were removed from further analysis. The final dataset consisted of 235,003 patients; it was split randomly into testing (25%; $n = 58,516$), validation (25%; $n = 58,798$), and training (50%; $n = 117,689$) datasets.

External causes of TBI

Of the 58,516 patients in the testing dataset, 57% were male and 43% were female patients. The most common TBI mechanisms were falls ($n = 26,480$ [45%]; male: $n = 13,400$ [40%]; female: $n = 13,080$ [52%]) and being struck by/against an object ($n = 20,845$ [36%]; male: $n = 13,370$ [40%]; female: $n = 7475$ [30%]). Ten percent were sustained in MVC, equally distributed among the sexes. Of all injuries, 25% were sports-related, affecting 29% of the male and 19% of the female patients, and assaults accounted for 7% of TBIs, affecting 10% of the male and 4% of the female patients. Injury severity was not established in 25,036 [43%] patients; among these, most cases were coded as concussions without a specified length of unconsciousness (ICD-10-CA code S06.0) (Table 1).

Sex-specific health status of the host, agent, and environment at the TBI event phase

From all possible 2600 codes classifying the TBI and reference patients’ health conditions at the ED or acute care settings at the event phase on which matched McNemar tests were performed, 273 significant associations (i.e., had an OR greater than one) were found after BY correction, of which 226 (83%) were internally validated. Of the 226 codes included in the factor analysis, 164 (73%) unique codes met the factor loading cut-off of 0.2 (Supplementary Table 2). Thirty-four factors were selected based on their interpretability and the breakpoints on the scree plots from further analyses (Table 2 and Fig. 1). For details on frequencies, ORs, and factor loadings of codes that met the factor analysis cut-off and rationale for Host, Agent, and environment designations, please see Supplementary Table 3.

For female patients, the host factors with the highest representation within severe TBIs were Heart and Metabolic Disorders (Factor 2), Metabolic and Elderly Emergencies (Factor 4), and Parkinson’s and Dementia (Factor 14); while the lowest share occurred in Headaches, Blurred Vision and Objects Strikes (Factor 34). The agent (i.e., elements that are necessary for the initiation of the pathological process of a TBI by the external force to the head) was most often Falls and Syncope (Factor 16), Car (Factor 18), and Heavy Transport Accidents (Factors 31), in which less severe injuries were prevalent (Fig. 1). Table 2 presents the frequencies, ORs, and ICD-10 codes composing each of the 34 factors. Supplementary Table 3 presents factor loading and a detailed description of each factor.
Sex-specific distribution of the host, agent, and environment at the TBI event phase

When factors were sorted by their proportion values in the TBI event, Factor 16 (Falls and Syncope), Factor 34 (Headache, Blurred Vision), and Factor 6 (Elderly Disorders & Neoplasms), were the three factors with the highest proportion in both sexes (Fig. 2). When these factors were sorted by their effect size (OR), Abuse and Sexual Assault (Factor 15) and Assault and Alcohol disorders (Factor 26) held the strongest associations with TBI in male patients. In females, the top two factors associated with TBI were injuries induced by transport (e.g., Pedestrian Injuries and Heavy Transport Injuries, Factor 31 and Factor 32, respectively) (Fig. 3).

Heatmaps connecting the host, agent, and environment

Sex-specific heatmaps of the clusters connecting the host, agent, and environment factors at the TBI event phase, FDR-adjusted, are presented in Supplementary Figs. 1 and 2. These heatmaps were generated and compared between the training, validation, and testing datasets to confirm observed patterns. In both male and female patients, three distinct clusters of associations were observed: Cluster A: Advanced Age-related Brain Pathology; Cluster B: Multiple Body System Pathology; and Cluster C: Young Age-related Concerns.

Each of these clusters composed the host, agent, and environment factors. Of the nine factors composing Cluster A (Advanced Age-related Brain Pathology),

Table 1 Characteristics of patients with a first traumatic brain injury-related visit in the emergency department or acute care and matched reference patients

| Variables | Patients with TBI (n = 58,516) | Reference patients (n = 58,516) |
|-----------|-------------------------------|---------------------------------|
|           | Female (n = 25,137; 43%)      | Male (n = 33,379; 57%)          |
|           |                               |                                 |
| Sociodemographic characteristics |                               |                                 |
| Age at injury (years) | 39.41 (26.3) | 33.84 (24.3) | 39.41 (26.3) | 33.85 (24.3) |
| Income quintile |                               |                                 |
| Q1 (lowest) | 4868 (19) | 6597 (20) | 4868 (19) | 6597 (20) |
| Q2 | 4895 (19) | 6645 (20) | 4895 (19) | 6645 (20) |
| Q3 | 4956 (20) | 6538 (20) | 4956 (20) | 6538 (20) |
| Q4 | 5269 (21) | 6913 (21) | 5269 (21) | 6913 (21) |
| Q5 (highest) | 5149 (20) | 6686 (20) | 5149 (20) | 6686 (20) |
| Rurality (yes) | 3842 (15) | 5242 (16) | 3842 (15) | 5242 (16) |
| TBI-related characteristics |                               |                                 |
| TBI main diagnosis | 22,163 (88) | 29,542 (89) | NA | NA |
| Injury severity |                               |                                 |
| Unspecified | 12,075 (48) | 12,961 (39) | NA | NA |
| Mild | 8260 (33) | 12,201 (37) | NA | NA |
| Moderate | 682 (3) | 1484 (4) | NA | NA |
| Severe | 4120 (16) | 6733 (20) | NA | NA |
| Type of first healthcare entry |                               |                                 |
| Emergency Department | 21,284 (85) | 26,858 (80) | NA | NA |
| Acute Care | 1606 (6) | 2541 (8) | NA | NA |
| Emergency & Acutea | 2247 (9) | 3980 (12) | NA | NA |
| External causes of injury, by CDCb |                               |                                 |
| Sports injury | 4668 (19) | 9804 (29) | NA | NA |
| Assault | 928 (4) | 3431 (10) | NA | NA |
| Falls | 13,080 (52) | 13,400 (40) | NA | NA |
| Motor vehicle collisions | 2587 (10) | 3221 (10) | NA | NA |
| Struck by/against | 7475 (30) | 13,370 (40) | NA | NA |
| Other | 2519 (10) | 4272 (13) | NA | NA |
| Missing | 44 (0) | 122 (0) | NA | NA |

Abbreviations: Q Quantile, NA Not applicable, TBI Traumatic brain injury, SD Standard deviation. Data presented as n (%) or mean (standard deviation). aA patient had a transfer to either location on the same day. bA patient may have several designations (i.e., sports injury and struck by/against an object)
### Table 2  Characteristics of male and female patients with TBI by injury severity

| VARIABLES                              | **FEMALES, by injury severity (25,137, %)** | **MALES, by injury severity (33,379, %)** |
|----------------------------------------|--------------------------------------------|-------------------------------------------|
|                                        | Unspecified (12,075, 48.0)                 | Unspecified (12,961, 38.8)                |
|                                        | Mild (8260, 32.9)                          | Mild (12,201, 36.6)                       |
|                                        | Moderate (682, 2.7)                        | Moderate (1484, 4.5)                      |
|                                        | Severe (4120, 16.4)                        | Severe (6733, 20.2)                       |
| **Sociodemographic characteristics**   |                                            |                                            |
| **Age at first TBI, years old**        |                                            |                                            |
| Mean (SD)                              | 33.51 (22.3)                              | 33.1 (22.9)                               |
| Median (Q1-Q3)                         | 27 (16–49)                                | 25 (15–49)                               |
| Income quantile Q1 (lowest)            |                                            |                                            |
| Q2                                     | 2334 (19.3)                               | 1568 (19.0)                              |
| Q3                                     | 2399 (19.9)                               | 1616 (19.6)                              |
| Q4                                     | 2645 (21.9)                               | 1725 (20.9)                              |
| Q5 (highest)                           | 2442 (20.2)                               | 1784 (21.6)                              |
| Rurality (yes)                         | 2000 (16.6)                               | 1269 (15.4)                              |
| **Factors**                            |                                            |                                            |
| 1. Multitrauma                          | 178 (1.5)                                 | 134 (1.6)                                |
| 2. Heart & Metabolic Disorders         | 577 (4.8)                                 | 374 (4.5)                                |
| 3. Alzheimer’s & Dementia              | 17 (0.1)                                  | 14 (0.2)                                 |
| 4. Endocrine, Metabolic & Elderly Emergencies | 497 (4.1) | 324 (3.9)    | 135 (19.8)                              |
| 5. Complications & Resp Emergencies    | 28 (0.2)                                  | 15 (0.2)                                 |
| 6. Elderly Disorders, Neoplasms & Falls | 1580 (13.1) | 985 (11.9) | 200 (29.3)                              |
| 7. Stroke & Brain Emerg Sequelae       | 62 (0.5)                                  | 25 (0.3)                                 |
| 8. Renal Dysfunction                   | 46 (0.4)                                  | 31 (0.4)                                 |
| 9. Resp Emerg, Septicemia              | 48 (0.4)                                  | 27 (0.3)                                 |
| 10. Coagulopathies                     | 50 (0.4)                                  | 39 (0.5)                                 |
| 11. Liver Disorders                    | 8 (0.1)                                   | NA (NA)                                  |
| 12. Resp Infections & ABX Resistance   | 64 (0.5)                                  | 63 (0.8)                                 |
| 13. Airway Obstruction                 | 113 (0.9)                                 | 59 (0.7)                                 |
| 14. Parkinson’s & Dementia             | 14 (0.1)                                  | 9 (0.1)                                  |
| 15. Abuse & Sexual Assault             | 54 (0.5)                                  | 45 (0.5)                                 |
| 16. Falls & Syncope                    | 3260 (27.0)                               | 2180 (26.4)                              |
| 17. Diabetic Emergencies               | 435 (3.6)                                 | 296 (3.6)                                |

---

**Notes:**
- [H] = host; [E] = Environment
- [A/ vector] = agent & vector
- [H/A] = host & vector

---
seven (78%) were depicting the host. In contrast, of 13 factors composed Cluster C (Young Age-related Concerns), nine (69%) were depicting agents. Cluster B (Multiple Body System Pathology) was uniting the host and environment factors equally (Fig. 4). The most discrete differences between male and female patients were observed within Cluster C (Young Age-related Concerns), in the Assault & Alcohol Disorders (Factor 26) and Abuse & Sexual Assault (Factor 15), followed by Cluster A (Advanced Age-related Brain Pathology) in the Brain and Other Hemorrhages (Factor 19) and Airway Obstruction (Factor 13).

### Table 2 (continued)

| VARIABLES | FEMALES, by injury severity (25,137, %) | MALES, by injury severity (33,379, %) |
|-----------|----------------------------------------|---------------------------------------|
|           | Unspecified (12,075, 48.0) | Mild (8260, 32.9) | Moderate (682, 2.7) | Severe (4120, 16.4) | Unspecified (12,961, 38.8) | Mild (12,201, 36.6) | Moderate (1484, 4.5) | Severe (6733, 20.2) |
| 18. Car Collision | 1180 (9.8) | 706 (8.6) | 76 (11.1) | 344 (8.4) | 932 (7.2) | 760 (6.2) | 95 (6.4) | 630 (9.4) |
| 19. Brain & Other Hemorrhages | 148 (1.2) | 96 (1.2) | 50 (7.3) | 864 (21.0) | 159 (1.2) | 129 (1.1) | 74 (5.0) | 1367 (20.3) |
| 20. Seizures & Drug Adversities | 114 (0.9) | 55 (0.7) | 13 (1.9) | 269 (6.5) | 142 (1.1) | 115 (0.9) | 31 (2.1) | 545 (8.1) |
| 21. Assault | NA (NA) | 8 (0.1) | NA (NA) | 16 (0.4) | 25 (0.2) | 36 (0.3) | 11 (0.7) | 75 (1.1) |
| 22. Motorcycle Collision | 218 (1.8) | 152 (1.8) | 40 (5.9) | 234 (5.7) | 414 (3.2) | 437 (3.6) | 90 (6.1) | 548 (8.1) |
| 23. Alcohol & Drugs Misuse | 152 (1.3) | 106 (1.3) | 18 (2.6) | 192 (4.7) | 290 (2.2) | 325 (2.7) | 126 (8.5) | 751 (11.2) |
| 24. Multiple Systems Collapse | 26 (0.2) | 8 (0.1) | 10 (1.5) | 127 (3.1) | 42 (0.3) | 19 (0.2) | 28 (1.9) | 369 (5.5) |
| 25. Intracranial Pathology, Convalescence | 90 (0.8) | 66 (0.8) | 34 (5.0) | 561 (13.6) | 103 (0.8) | 101 (0.8) | 56 (3.8) | 914 (13.6) |
| 26. Assault & Alcohol Disorders | 586 (4.9) | 456 (5.5) | 88 (12.9) | 271 (6.6) | 1256 (9.7) | 1703 (14.0) | 628 (42.3) | 1163 (17.3) |
| 27. Aplastic Anemias & Hemorrhages | 45 (0.4) | 33 (0.4) | 10 (1.5) | 188 (4.6) | 37 (0.3) | 43 (0.4) | 27 (1.8) | 276 (4.1) |
| 28. Risky Behaviors, Drug Disorders & Social Disparities | 199 (1.7) | 128 (1.6) | 26 (3.8) | 229 (5.6) | 385 (3.0) | 421 (3.5) | 150 (10.1) | 879 (13.1) |
| 29. Superficial Injuries | 432 (3.6) | 277 (3.4) | 16 (2.4) | 123 (3.0) | 425 (3.3) | 319 (2.6) | 40 (2.7) | 181 (2.7) |
| 30. Pedal cycle Injuries | 305 (2.5) | 255 (3.1) | 27 (4.0) | 76 (1.8) | 569 (4.4) | 656 (5.4) | 76 (5.1) | 328 (4.9) |
| 31. Heavy Transport Injuries | 693 (5.7) | 419 (5.1) | 46 (6.7) | 189 (4.6) | 546 (4.2) | 432 (3.5) | 53 (3.6) | 283 (4.2) |
| 32. Pedestrian Injuries, Car Collision | 117 (1.0) | 102 (1.2) | 26 (3.8) | 207 (5.0) | 128 (1.0) | 108 (0.9) | 42 (2.8) | 252 (3.7) |
| 33. Falls from Elevation | 297 (2.5) | 186 (2.3) | 22 (3.2) | 125 (3.0) | 471 (3.6) | 467 (3.8) | 66 (4.5) | 513 (7.6) |
| 34. Headache, Blurred Vision & Object Strikes | 3020 (25.0) | 1787 (21.6) | 57 (8.4) | 314 (7.6) | 3208 (24.8) | 2459 (20.2) | 175 (11.8) | 584 (8.7) |

*Abbreviations: ABX Antibiotic, A Agent, E Environment, H Host. NA Not available; frequencies of less than 6 hidden to protect confidentiality*
Fig. 1 Host (H), agent (A), environment (E) and intertwined (H/A and H/E) factors in male and female patients with TBI in Ontario, Canada 2005/06–2015/16. Data are shown by injury severity.

Fig. 2 Wordcloud of factors, by proportion, in male and female patients. The word size is relational to the proportion that the factor appears during the current injury event window in female (N=25,137) or male (N=33,379) patients with traumatic brain injury. The colours represent classifications of the Haddon Matrix as follows: host = “blue”; agent = “brown”; environment = “grey”; host/agent = “deep pink”; host/environment = “sea green”
Fig. 3 Wordcloud of factors by the magnitude of effect size. The word size is proportional to the magnitude of the Odds Ratio (OR), measuring the association between possessing the factor during the current injury event window. The ORs were calculated using logistic regression in the testing dataset, controlled for rurality, age, place of residence, and income quintile. The colors represent classifications of the Haddon Matrix as follows: host = “blue”; agent = “brown”; environment = “grey”; host/agent = “deep pink”; host/environment = “sea green”

Fig. 4 The Host, agent, and environment factors correlations (a) and differences between the sexes (female-male) (b). In (a), a positive association is coded in red, and a negative association is coded in blue. In the heatmap reporting sex differences (b), the red color indicates that if the correlation in a sex-specific heatmap was positive (i.e., red color), females have a stronger positive correlation between those two variables than males, and if it was negative (blue color), females have a weaker correlation than males. The blue color in (b) indicates that if the original correlation was positive, males have a stronger positive correlation between those two factors, and if negative, males have a weaker correlation than females.
Host, agent, and environment in external cause of injury and TBI severity

Many of the factors depicting the host and host/environment showed a significant association with external cause of injury and TBI severity. In both male and female patients, falls had a strong positive correlation with chronic disorders within the host, including Heart and Metabolic Disorders (Factor 2), Diabetic, Endocrine, Metabolic, and Elderly Emergencies (Factors 4 and 17). These disorders were also positively associated with severe TBI and increasing age and negatively associated with Abuse and Sexual Assault (Factor 15) and rural residence (Fig. 5). When comparing the sexes, the strength of the association with increasing age was more profoundly seen in female as compared to male patients. Struck by or against an object showed negative associations with age and chronic disorders of the host; the strength was weaker for male patients as compared to female patients.

Motor vehicle collisions were positively correlated with Multisystem Collapse (Factor 24); Intracranial Pathology and Convalescence (Factor 25); and Complications and Respiratory emergencies (Factor 5) in both sexes with an equal strength of association and shared links to severe TBI. In both sexes, sports related TBIs showed strong positive associations with higher income level and negative associations with chronic disorders of the host and adverse social environment. The associations were stronger for female patients as compared to male patients in Alcohol and Drug Misuse (Factor 23) and Risky Behaviours and Social Disparities (Factor 28). Notably, assault as an external cause of injury highlighted strong positive associations with Alcohol and Drug Misuse (Factor 23), and Risky Behaviours and Social Disparities (Factor 28), as well as Abuse and Sexual Assault (Factor 15), with much stronger effect seen in male patients as compared to female patients.

When we analysed the agent factors (Fig. 5), sex-specific patterns in the host were observed. Car Collisions and Falls from Elevation (Factors 18 and 33) were positively associated with Multiple System Collapse and Intracranial Pathology and Convalescence within the Host (Factors 24 and Factor 25) in male but not female patients. The agent Assault (Factor 21) was positively associated with Abuse and Sexual Assault (Factor 15) in female but not male patients. When sex differences were analysed, it appeared that in male patients, the association between the agent factor Assault (Factor 21) and the environment factor Abuse and Sexual Assault (Factor 15) was stronger than in female patients. Other noteworthy observation included the direction of associations between increasing age and Falls from Elevation (Factor 33), positive for male patients and negative for female patients.

All reported results were confirmed in the training and validation datasets, and they were confirmed to be robust (Supplementary Figs. 3 and 4).
Discussion

Our research is the first, to the best of our knowledge, to use the computational and visual techniques and the Haddon Matrix to analyze data from the injury event of patients entering ED or acute care facilities with a TBI-related diagnosis code in Ontario, Canada, over a period of 10 years. We present previously known and well-established associations, which support the theoretical validity of the approach, as well as sex-specific associations that are not initially intuitive, without any predetermined human bias. The results highlight differences between male and female patients with TBI at the injury event in the characteristics of the host, the agent, and the environment. These differences require further investigation for their implications to be established more precisely, as they undoubtedly reflect biological (sex) vulnerability (e.g., probability of sustaining a TBI in falls due to biological differences) and sociocultural (gender) vulnerability (e.g., probability of TBI due to assault), both of which have implications in clinical and preventive contexts, and for TBI surveillance. We also corroborated our prior hypotheses on the dynamics of the TBI-comorbidity relationship at the injury event and its implications for external causes of injury and TBI severity. We issued an exhaustive effort to describe unique aspects of sex differences of the host at the injury event, as well as differences in exposure to adverse social and environmental circumstances, which can be a function of gender, rooted in socially determined expectations and behaviors ascribed on the basis of biological sex [48].

The host: differences between male and female patients

The host’s biological, psychosocial, and behavioral characteristics have been the primary targets of TBI research, identifying high risk groups for primary, secondary, and tertiary preventive initiatives [49–51]. Several prior studies reporting on in-hospital TBI outcomes found more severe brain pathology in males than in females, irrespective of age [4, 5, 52]. Our results add to the findings from previous research, highlighting the greater frequency of severe TBI in male patients as compared to female patients. Our results also provide insight into the potential underpinnings of this difference between the sexes. For instance, the strength of the association between Multiple System Collapse, in-hospital Diabetic and Respiratory Emergencies, Septicemia, and severe TBI was stronger in male patients as compared to female patients, which may point to their more pronounced physiological dysregulation in response to TBI and biological vulnerability to early adverse outcomes [5].

Marked differences between male and female patients at the TBI event have been identified previously in diseases affecting the cardiovascular, endocrine, and metabolic systems [53], and were likewise captured in this study in Factors 2, 4, 17, and 19. These factors are implicated in the risk of Syncope and Falls (Factor 16) and severe TBI and were more profoundly expressed in female patients than male patients in this study. Since the severity of TBI is assigned according to clinical signs and symptoms [54], increased awareness among knowledge users of the influence of these differences when assessing TBI severity is important [55, 56]. Sex differences in alcohol disorders at the injury event, highlighted in this study, likely contribute to injury severity designations. Substances that cross the blood-brain barrier affect brain function and cognition, impact depth and duration of coma following injury and alter indices of TBI severity. The most widely used measure of consciousness and thereby injury severity post-TBI is the GCS, where coma is assessed by ability to open the eyes, obey commands, and/or utter understandable words, each of which can be affected by alcohol, confounding GCS assessment in trauma patients and possibly impacting assessment of acute care performance in male patients to a greater extent than in female patients [57].

Another finding that warrants discussion is the relationship between abuse, sexual assault, and TBI. It was observed that Factor 15 (Abuse and Sexual Assault) is overrepresented in TBI patients as compared to the reference population. This overrepresentation is particularly pronounced in females, which is consistent with results from previous research [58, 59]. However, the stronger association between Factor 15 (Abuse and Sexual Assault) and TBI observed in male patients as compared to female patients may suggest that abuse and sexual assault are more likely to go undetected in male patients, as suggested by research from other settings [60, 61]. This is not to say that sexual assault in female patients is necessarily less challenging to capture but that the collection of abuse and sexual assault data in male patients with TBI should be given appropriate consideration given the likely impact of gender norms, expectations, and expressions on reporting in this group [62–64]. Furthermore, the codes that loaded on Factor 15 (T74, Y07, Y05, Y06) point to the complexity of the social environment of a TBI patient and highlighting the importance of providing customized assessment and care to a patient in TBI management.

The agent: differences between male and female patients

The fundamental tasks in injury control are to prevent the agent from reaching people in a magnitude that exceeds injury thresholds and to minimize the consequences of the injury [21, 22]. The ‘agent’ with respect to the injury event includes everything that determines whether a TBI
results from the injury. We studied implications of agent in falls, strikes by/against an object, MVCs, sports injuries, and assaults. A note-worthy potential contribution to the evolution of TBI surveillance is our data-driven confirmation that the host could also be the agent, as in the case of a person with syncope who falls (Factors 6 and 16) or a person under the influence of alcohol who is assaulted (Factor 23). The role of medication management requires ongoing attention in light of the sex differences that we observed in abnormalities of balance and gait and spontaneous postural sway due to changes in neuromuscular activity, a key physiological system that facilitates and controls movement and stability of the person [65–67].

The environment: differences between male and female patients

In the present work, we characterized the environment of male and female patients at the TBI event. We found that sustaining TBI in a motorcycle collision (Factor 22) was positively and strongly correlated with living in rural areas in male but not female patients; these results are consistent with those of other studies [68–70]. The sex differences found in socioeconomic status associated with rural community suggests that TBI is subject to a social gradient and impacted by geographical, cultural, and personal values as well as residents’ economic and health status [5]. With more refined data on social standing, future data-driven studies may be able to pinpoint the social equity parameters that underlie the geographic variation in TBI [71].

External cause and TBI severity

External causes of morbidity and mortality codes classify many dimensions of the cause of injury in a single ICD code [72, 73]. The US CDC external causes of injury codes include the dimensions of intent (i.e., assault and unintentional injuries), and cause (falls, struck by/against an object) [7, 34]. The APHEO, when categorizing injuries, adds circumstance of injury (i.e., sport-related injuries) [35]. Our findings, therefore, provide more detailed injury classification schemes as compared to external course codes of injury alone, providing more precision with respect to sex differences in trauma [74, 75], and the mechanism of injury, the circumstances, and the agent involved.

Our data-driven research contributes to the discussion on the clinical definition of TBI severity for the purposes of surveillance, specifically in monitoring sex-specific subgroups by intent of injury (e.g., assault), mechanism of injury (e.g., falls), and circumstance of injury (e.g., struck by/against and object) during a specified activity or place (e.g., in a car when driving machinery, in a field when playing sports). While previous research on TBI severity focused on injury events assumed to be independent of the agent and environment, our results show that the Haddon Matrix designations are not independently contributing to TBI severity but are uniformly present at the injury event and correlate with one another in a sex-specific manner. It follows that if the host, agent, and environment present in a sex-specific manner in a TBI severity, then the workup for designating TBI severity should incorporate the aforementioned triad.

There are limitations to consider when discussing the results of our research. The first limitation concerns the reliability of administrative healthcare data used, and the second concerns complexity of the clustering technique. Despite the population-wide universal health coverage in Ontario, which strives to provide equal access, and the reliability/validity of ICES data on ED and acute care visits [76–78], there may remain uncovered variation between the sexes due to differences in help-seeking behaviors. This may be especially true for certain external causes of TBI, for example, assault related TBI. Sex differences in reporting TBI are well documented in sports and other contexts [79–81]. In an effort to mitigate this issue, we internally validated the results that emerged in the testing dataset using the training and validation datasets and detailed these results in the supplementary material of the manuscript.

The data analyzed in this work concerned the TBI event window and was used to measure the contributions of the host, agent, and environment at the injury event captured in the ED and acute care hospitals. The clustering technique we utilized offers a sound basis for extending the body of knowledge for the fields of medicine, public health, and injury surveillance, as it offered internally validated hierarchical structures into which new ideas concerning the TBI event window and injury severity in male and female patients can be input. Nevertheless, the data-driven process of discovering interesting and meaningful patterns in data using computational analytics methods is still evolving, and how the new methodology may intersect with existing hypothesis-driven research practices in epidemiological research remains uncertain. It is key that our results are not used to make conclusions about causality, but as new sex-specific hypotheses about complex and non-linear interactions of the host, the agent, and the environment in the TBI event window, which are difficult to disentangle via traditional research methodologies. Follow-up to this work should involve a sex-specific prediction of TBI events and injury severity, establishing projection of resource needs and priorities based on the characteristics of the host within their environment.
Implications for prevention
The goal of prevention is to protect, promote, and maintain health and well-being and to prevent disease, disability, and death [82–84]. Primary prevention aims to circumvent pathology before it occurs. Secondary prevention entails recognition of pathology at its earliest stages, to stop its progression. Tertiary and quaternary prevention involve treatment directed to prevent complications and minimize disability [84]. Stemming from this research, central prevention ideas encompass the need for balance training [85] and addressing cardiovascular syncope [86] in the elderly to prevent TBI sustained in falls (i.e., Elderly disorders & Neoplasms (Factor 6), associated with Alzheimer’s and Dementia (Factor 1) and Falls and Syncope (Factor 16), and clustering on Advanced Age-related Brain Pathology. Education about alcohol intoxication [87] to prevent assault related TBI is similarly important, associated with Factor 26 Assault and Alcohol disorders and Factor 15 Abuse and Sexual Assault in the Young Age-related Concerns cluster; the latter was expressed to a greater extent in male than in female persons with TBI as compared to a reference cohort. Equally important is regulating the environment through government action on road planning, traffic laws, protective equipment, airbags and seatbelts [88–90] to prevent the agent’s energy from reaching the host in an amount that exceeds the injury threshold in traffic accidents associated with Multitrauma (Factor 1) and severe TBI. The observed sex differences across the agent and external cause of injury highlight risk behaviors that differ between the sexes, potentially steeped in socially-promoted gendered behaviours [91]. While it is extremely challenging to modify behaviors [92], it might be possible to educate the public about the effects of gender at the microgeographic level, comprehensively depicting the injury incident and its consequences for groups at risk [93]. Ideas for secondary, tertiary, and quaternary prevention strategies emerged from the results of this research, including attention to interventions directed on the risk associated with the loss of patient autonomy in severe TBI [94], strongly linked to Elderly Disorders & Neoplasms (Factor 6) in both sexes. There exist experimental therapies that inhibit the release of excitotoxins that play an important role in secondary injury, to attenuate cellular oxidative and metabolic stress [95]. Likewise, deep brain stimulation targeting the ascending reticular activating system and areas of the thalamus has been proposed for promoting the autonomy of vulnerable elderly people with TBI [95, 96]. With the widening temporal agent-host-environment window to pre- and post-injury phases through sex lens, the lines of this study may produce substantial advances in the surveillance and prevention of TBI.

Conclusions
The study findings support the rationale, feasibility, and benefit of computational and data visualization techniques in examining sex-specific patterns within TBI events shaped by social circumstances, providing an enhanced understanding of complex injury data and room to advance TBI surveillance and care. We uncovered meaningful differences between the sexes in the host, agent, and environment and thus contributed to the existing literature in three significant ways. First, our results support the value of ongoing consideration of the differences between male and female patients across the Haddon matrix dimensions. Second, we found that the host’s health status, the agent and the environment are of great relevance to the designation of injury severity and external cause of injury, providing room to advance TBI surveillance. Finally, this study expands the existing knowledge on the healthcare services needed based on the intent of injury (e.g., assault), nature of injury (e.g., falls), or activity when injured (e.g., sports). Future development and application of data-driven methods will facilitate a better understanding of health and harmful environmental exposures of male and female persons across the lifespan and advance the fields of both precision TBI medicine and injury surveillance.

Abbreviations
BY: Benjamini-Yekutieli; CDC: Centers for Disease Control and Prevention; CI: Confidence interval; DAD: Discharge Abstract Database; ED: Emergency department; GCS: Glasgow Coma Scale; FDR: False Discovery Rate; ICD-10: International Statistical Classification of Diseases and Related Health Problems, Tenth Revision; NACRS: National Ambulatory Care Reporting System; OR: Odds ratios; TBI: Traumatic brain injury; US: United States; SD: Standard deviation; SAGER: Sex and Gender Equity in Research; STROBE: The Strengthening the Reporting of Observational Studies in Epidemiology.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12874-021-01493-6.

Additional file 1: Supplementary Table 1. Data-driven studies examining TBI event, with results presented by the Haddon Matrix designations.
Supplementary Table 2. Frequencies, ORs, and factor loadings of codes that met the factor analysis cut-off. Supplementary Table 3. Frequencies, ORs, and factor loadings of codes that met the factor analysis cut-off and rationale for Host, Agent, and environment designations. Supplementary Figure 1. Heatmap of correlation between factors during the injury event-phase period for patients with TBI in the training dataset. Supplementary Figure 2. Heatmap of correlation between factors during the injury event-phase period for patients with TBI in the validation dataset. Supplementary Figure 3. Heatmap of correlation between Haddon Matrix-classified factors and external cause as well as TBI severity for patients with TBI in the training set. Supplementary Figure 4. Heatmap of correlation between Haddon Matrix-classified factors and external cause as well as TBI severity for patients with TBI in the validation set.
Acknowledgements
This study utilized ICES Data & Analytic Services (DAS) and used de-identified data from the ICES Data Repository, which is managed by ICES with support from its funders and partners: Canada’s Strategy for Patient-Oriented Research (SPOR), the Ontario SPOR Support Unit, the Canadian Institutes of Health Research and the Government of Ontario. The opinions, results, and conclusions reported are those of the authors. No endorsement by ICES or any of its funders or partners is intended or should be inferred. Parts of this material are based on data and information compiled and provided by CIHI. However, the analyses, conclusions, opinions, and statements expressed herein are those of the author and not necessarily those of CIHI.

Authors’ contributions
AC, TM, VC and MDE conceived the original concept and initiated the work. MDE designed and optimized statistical analyses for this work. AT carried out the analyses with the support of MDE. TM optimized reporting. All authors discussed the results briefly and further steps for analyses and interpretation. TM wrote the manuscript. All authors read the paper and commented on the text. The author(s) read and approved the final manuscript.

Funding
The research reported in this presentation was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number R21HD089106 and National Institute Of Neurological Disorders And Stroke of the National Institutes of Health under Award Number R01NS117921, and in part, by the Canada Research Chairs Program. The content is solely the authors’ responsibility and does not necessarily represent the official views of the National Institutes of Health or Canada Research Chairs Program.

Availability of data and materials
ICES is an independent, non-profit research institute funded by an annual grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC). As a prescribed entity under Ontario’s privacy legislation, ICES is authorized to collect and use health care data for the purposes of health system analysis, evaluation, and decision support. Secure access to these data is governed by policies and procedures that are approved by the Information and Privacy Commissioner of Ontario. While data sharing agreements prohibit ICES from making the dataset publicly available, access may be granted to those who meet prespecified criteria for confidential access, available at www.ices.on.ca/DAS under accession DAS 2016–257(2018 0970 084 00). The complete dataset creation plan and analytic code remain accessible through authors upon request, understanding that the computer programs could rely upon coding templates or macros that are unique to ICES and are therefore either inaccessible or may require adjustment. All of the computer programs and scripts are available upon request from authors.

Declarations

Ethics approval and consent to participate
The study protocol was approved by the ethics committees at the (Toronto Rehabilitation Institute-University Health Network) and academic (University of Toronto) institutions. This research utilised encrypted administrative health data with no access to personal information. No humans were directly involved in this study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests or personal relationships that could have appeared to influence the results reported in this paper.

Author details
1. KITE-Toronto Rehabilitation Institute, University Health Network, Toronto, Ontario, Canada. 2. Department of Occupational Science & Occupational Therapy, University of Toronto, Toronto, Ontario, Canada. 3. Rehabilitation Sciences Institute, Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada. 4. Acquired Brain Injury Research Lab, University of Toronto, Toronto, Ontario, Canada. 5. Dalla Lana School of Public Health, University of Toronto, Toronto, Canada, Ontario. 6. Trinity College Institute of Neuroscience, Global Brain Health Institute, Dublin, Ireland. 7. Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada.

Received: 7 June 2021 Accepted: 29 November 2021

Published online: 30 January 2022

References
1. GBD 2016 Traumatic Brain Injury and Spinal Cord Injury Collaborators. Global, regional, and national burden of traumatic brain injury and spinal cord injury, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 2019;18(1):56–87.
2. Menon DK, Schwab K, Wright DW, Maas AI, Demographics and Clinical Assessment Working Group of the International and Interagency Initiative toward Common Data Elements for Research on Traumatic Brain Injury and Psychological Health. Position statement: definition of traumatic brain injury. Arch Phys Med Rehabil. 2010;91(11):1637–40.
3. Maas A, Menon DK, Adelson PD, Andelic N, Bell MJ, Belli A, et al. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. Lancet Neurol. 2017;16(12):987–1048.
4. Nguyen R, Frest KM, McChesney J, Kwon CS, Jette N, Frolikis AD, et al. The international incidence of traumatic brain injury: a systematic review and Meta-analysis. The Canadian journal of neurological sciences. J Can Sci Neurol. 2016;43(6):774–85.
5. Mollayeva T, Mollayeva S, Colantonio A. Traumatic brain injury: sex, gender and intersecting vulnerabilities. Nature reviews. Neurology. 2018;14(12):711–22.
6. World Health Organization. Violence and Injury Prevention and Disability. Neurotrauma. Retrieved May 20, 2021 from: https://www.who.int/violence_injury_prevention/road_traffic/activities/neurotrauma/en/.
7. Centers for Disease Control and Prevention. Traumatic brain injury and concussion. Retrieved May 20, 2021 from: https://www.cdc.gov/traumaticbraininjury/index.html
8. Mollayeva T, Sutton M, Chan V, Colantonio A, Jana S, Escobar M. Data mining to understand health status preceding traumatic brain injury. Sci Rep. 2019;9(1):5574.
9. Daugherty J, Waltzman D, Sarmiento K, Xu L. Traumatic brain injury-related deaths by race/ethnicity, sex, intent, and mechanism of injury - United States, 2000-2017. MMWR Morb Mortal Wkly Rep. 2019;68(46):1050–6.
10. Gao S, Kumar RG, Wisniewski SR, Fabio A. Disparities in health care utilization of adults with traumatic brain injuries are related to insurance, race, and ethnicity: a systematic review. J Head Trauma Rehabil. 2018;33(3):E40–50.
11. Peterson AB, Sarmiento K, Xu L, Haileyesus T. Traumatic brain injury-related hospitalizations and deaths among American Indians and Alaska natives - United States, 2008-2014. J Saf Res. 2019;71:315–8.
12. Haines KL, Nguyen BP, Vatsaas C, Alger A, Brooks K, Agarwal SK. Socioeconomic status affects outcomes after severity-stratified traumatic brain injury. J Surg Res. 2019;235:131–40.
13. Yamagami K, Kurogi R, Kurogi A, Nishimura K, Onozuka D, Ren N, et al. The influence of age on the outcomes of traumatic brain injury: findings from a Japanese Nationwide survey (I-ASPECT study-traumatic brain injury). World Neurosurg. 2019;130:e26–46.
14. Agari S, Adams H, Kasprzowicz M, Cosynska M, Smieleswski P, Eccle A. Feasibility of hidden Markov models for the description of time-varying physiologic state after severe traumatic brain injury. Crit Care Med. 2019;47(11):e880–5.
15. Matsuo K, Aihata H, Nakai T, Morishita A, Tohma Y, Kohmura E. Machine learning to predict in-hospital morbidity and mortality after traumatic brain injury. J Neurotrauma. 2020;37(1):202–10.
16. Hernandez Rocha TA, Elahi C, Cristina da Silva N, Sakita FM, Fuller A, Mmbaga BT, et al. A traumatic brain injury prognostic model to support in-hospital triage in a low-income country: a machine learning-based approach. J Neurosurg. 2019;12(6):1961–9.
17. Follweiler KA, Sandsmark DK, Diaz-Anastasia R, Cohen AS, Maisano AJ. Unsupervised machine learning reveals novel traumatic brain injury patient phenotypes with distinct acute injury profiles and Long-term outcomes. J Neurotrauma. 2020;37(12):1431–44.
18. Serviá L, Montserrat N, Badia M, Llopart-Pou JA, Barea-Mendoza JA, Chico-Fernández M, et al. Machine learning techniques for mortality prediction in critical traumatic patients: anatomic and physiologic variables from the RETRAUCI study. BMC Med Res Methodol. 2020;20(1):262.

19. Rau CS, Kuo PJ, Chien PC, Huang CY, Hsieh HY, Hsieh CH. Mortality prediction in patients with isolated moderate and severe traumatic brain injury using machine learning models. Plos One. 2018;13(11):e0207192.

20. Gravesteijn BY,Nieboer D, Ecote A, Lingsma HF, Nelson D, van Calster B, et al. Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. J Clin Epidemiol. 2020;122:95–106.

21. Haddon W Jr. The changing approach to the epidemiology, prevention, and amelioration of trauma: the transition to approaches etiologically rather than descriptively based. Am J Public Health Nations Health. 1968;58(8):1431–8.

22. Haddon W Jr. Advances in the epidemiology of injuries as a basis for public-policy. Public Health Rep (Washington, D.C. 1974). 1980;95(5):411–21.

23. Coronado VG, Mcguire L, Faul M, Sugerman D, Pearson W. Brain injury medicine. In: Zasler ND, Katz DI, Zafonte RD, editors. Chapter: traumatic brain injury epidemiology and public health. Publisher, Principles and Practice. 2nd ed. Springer Publishing Company; 2012. p. 72–83. https://cop.credoreference.com/component/booktracker/edition/10613.html.

24. Parnatayak SK, Kramer RA, Vincent JR. Ecosystem change and human health: implementation economics and policy. Philos Trans R Soc Lond B Biol Sci. 2017;372(1722):20160130.

25. Freeman J, Maxwell JC, Davey J. Unraveling the complexity of driving while intoxicated: a study into the prevalence of psychiatric and substance abuse comorbidity. Accid Anal Prev. 2011;43(1):34–9.

26. Chan V, Hurst M, Petersen T, Liu J, Molloyeva T, Colantonio A, et al. A population-based sex-stratified study to understand how health status preceding traumatic brain injury affects direct medical cost. Plos One. 2020;15(10):e0240208.

27. Chan V, Sutton M, Molloyeva T, Escobar MD, Hurst M, Colantonio A. Data mining to understand how health status preceding traumatic brain injury affects functional outcome: a population-based sex-stratified study. Arch Phys Med Rehabil. 2020;101(10):1523–31.

28. World Health Organization. International Statistical Classification of Diseases and Related Health Problems 10th Revision. Retrieved May 20, 2021 from: https://icd.who.int/browse10/2019/en/

29. ICES data. Assessed May 20, 2021 at: https://www.ices.on.ca/Data-and-Policy/ICES-data

30. Statistics Canada. 2017. Ontario [province] and Canada [country] (table). Census profile. 2016 census. Statistics Canada catalogue no. 98-316-X2016001. Ottawa. Released November 29, 2017. Retrieved May 20, 2021 from: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E

31. Health Canada. Canada’s Health Care System. Retrieved May 20, 2021 from: https://www.canada.ca/en/health-canada/services/health-care-system/reports-publications/health-care-system/canada.html

32. Corrigan JD, Kreider S, Cuthbert J, Whyte J, Dams-O’Connor K, Faul M, et al. Components of traumatic brain injury severity indices. J Neurotrauma. 2014;31(11):1000–7.

33. Lesko MM, Woodford M, White L, O’Brien SJ, Childs C, Lecky FE. Using abbreviated injury scale (AIS) codes to classify computed tomography trauma. 2014;31(11):1000–7.

34. Coronado VG, Huerta M, Petersen T, Liu J, Molloyeva T, Colantonio A, et al. A population-based sex-stratified study to understand how health status preceding traumatic brain injury affects direct medical cost. Plos One. 2020;15(10):e0240208.

35. Vriend J, Gouttebarge V, Finch CE, van Mechelen W, Verhagen A. Intervention strategies used in sport injury prevention studies: a systematic review identifying studies applying the Haddon matrix. Sports Med. (Auckland, N.Z.). 2017;47(10):2027–43.

36. Zetterqvist J, Vermeulen K, Vansteelandt S, Sjölander A. Doubly robust conditional logistic regression. Stat Med. 2019;38(23):4749–60.

37. Chishtie JA, Marchand JS, Tuncote A, Libioka I, Babineau J, Cepoiu-Martin M, et al. Visual analytic tools and techniques in population health and health services research: scoping review. J Med Internet Res. 2020;22(12):e17892.

38. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for R. Bioinformatics (Oxford, England). 2008;24(5):719–20.

39. Ramseyer G. Testing the difference between dependent correlations using the Fisher Z. J Exp Educ. 1979;47(4):307–10 Retrieved May 20, 2021 from: http://www.jstor.org/stable/1951129.

40. Asuero AG, Sayago A, González AG. The correlation coefficient: an overview. Crit Rev Anal Chem. 2006;36(1):41–59.

41. Canadian Institutes of Health Research. Sex, Gender and Health Research. 2020. Retrieved May 20, 2021 from: https://cihr-irsc.gc.ca/e/50833.html

42. Chell J, Brandani CE, Frascetti S, Chakraverty J, Camomilla V. Limitations of the European barrier crash testing regulation relating to occupant safety. Accid Anal Prev. 2019;133:105239.

43. Fajkowiak RJ, Yogananand N, Zhang J, Pintar FA. A finite element model of region-specific response for mild diffuse brain injury. Stapp Car Crash J. 2009;53:193–213.

44. Wu YY, Hwang HF, Hu MH, Chen CY, Lin MR. Effects of fall injury type and discharge placement on mortality, hospitalization, falls, and ADL changes among older people in Taiwan. Accid Anal Prev. 2013;50:887–94.

45. Brazinova A, Rehorcikova V, Taylor MS, Buckova V, Majdan P, Posta M, et al. Epidemiology of traumatic brain injury in Europe: a living systematic review. J Neurotrauma. 2018. http://doi.org/10.1089/neu.2015.4126 Advance online publication.

46. Gaddam SS, Buell T, Robertson CS. Systemic manifestations of traumatic brain injury. Handb Clin Neurol. 2015;127:205–18.

47. Hawryluk GW, Manley GT. Classification of traumatic brain injury: past, present, and future. Handb Clin Neurol. 2015;127:15–21.

48. Girona M, Tousi A, Wilson M, Shahzad K, Martin R. The increasing age of TBI patients at a single level 1 trauma Center and the discordance between GCS and CT Rotterdam scores in the elderly. Front Neurol. 2020;11:112.

49. Navas JF, Martín-Pérez C, Petrova D, Verdejo-García A, Cano M, Sagranti-Mauzin O, et al. Sex differences in the association between impulsivity and driving under the influence of alcohol in young adults: the specific role of sensation seeking. Accid Anal Prev. 2019;124:174–81.

50. DiGioia AM, Wittenberg BA, Cutcher CJ, 2nd, Kennamer B, Greene CS, Velander AJ, et al. The impact of drug and alcohol intoxication on Glasgow coma scale assessment in patients with traumatic brain injury. World Neurosurg. 2020;135:e6664–70.

51. Connell M. Expert testimony in sexual assault cases: alcohol intoxication and memory. Int J Law Psychiatry. 2015;42:43–98. 105.

52. Loder RT, Robinson TP. The demographics of patients presenting for sexual assault at US emergency departments. J Forensic Legal Med. 2020;69:101887.

53. Mclean IA. The male victim of sexual assault. Best practice & research. Clin Obstet Gynaecol. 2013;27(1):39–46.
