International Journal of Medical and Health Research
ISSN: 2454-9142
Impact Factor: RJIF 5.54
www.medicalsciencejournal.com
Volume 3; Issue 12; December 2017; Page No. 128-131

Predictors of hospital admission in exercise-related injuries: Use of decision tree analysis

Peter D Hart
Associate Prof., Health Promotion, College of Education, Arts & Sciences and Nursing, Montana State University, Northern, USA

Abstract

Background: The purpose of this study was to use a novel data mining technique to identify predictors of hospital admission in adults injured during an exercise-related (ER) activity.

Methods: Data for this research came from the 2015 National Electronic Injury Surveillance System (NEISS) which collects data annually from a representative sample of U.S. emergency departments (EDs). Product codes were used to identify adults 18+ years of age who were injured in an ER activity. Variables utilized in the analysis were hospital admission (yes/no), body part (upper/lower), location (recreation or sport facility/other), age group (18-24/25-49/50-64/65+ years), race (white/black/other), and sex (male/female). SAS survey procedures and SPSS CHAID were used for the analyses.

Results: An estimated 16,958 (5.4%) out of 311,563 adults were admitted in 2015 after presenting to an ED with upper or lower body ER injury. Multiple logistic regression showed body part, age group, and race as independent predictors of admission. CHAID analysis with 95.3% accuracy showed that the first best predictor of admission was age group. Among the 65+ age group, race and then body part were significant predictors. The three younger age groups showed a similar pattern with body part then age group and location significantly predicting admission.

Conclusion: Results from this study support the use of a novel data mining tool to find specialized subgroups of ER injuries predictive of hospital admission.

Keywords: CHAID, epidemiology, exercise, physical activity, injury

Introduction

Physical activity (PA) is a recommended behavior because of its protective benefits against chronic disease and premature mortality [1-3]. Current U.S. guidelines for PA recommend all adults accumulate at least 150 minutes of moderate-intensity PA each week [4]. Furthermore, recent surveillance efforts have shown that majority of adults participate in either recommended amounts or amounts considered insufficient or below guideline thresholds or muscle strengthening activity [5, 6]. Given a large number of active adults in the U.S., health promotion concerns shift to understanding the negative side-effects associated with activity. More specifically, PA-related or exercise-related (ER) injury may be of increased concern as more individuals adopt the behavior. One recent study has shown that ER injuries presenting to U.S. emergency departments (EDs) has linearly increased (R^2=.939) from the years 2006 to 2015 [7].

Another concern associated with ER injury is the extent of the harm suffered by those who are injured. There are many gradations of injury, some requiring over-the-counter medication and self-help [8]. While other ER injuries may require a visit to an ED, and if severe enough, admittance to the hospital itself. Thus, a better understanding of the predictors of hospital admission due to ER injury is warranted for proper health promotion. Chi-squared automatic interaction detection (CHAID) is a classification tree method that has the ability to find the most predictive factors associated with certain outcomes, such as hospital admission [9]. Therefore, the purpose of this study was to use a novel data mining technique to identify predictors of hospital admission in adults injured during an ER activity.

Methods

Participants and design

The National Electronic Injury Surveillance System (NEISS) was used for this study. The NEISS collects data annually from a representative sample of U.S. EDs [10]. The 2015 dataset was downloaded from the NEISS and included a small set of variables: date of treatment, case number, date of birth, age of patient, sex of patient, diagnosis, body part affected, case disposition, product code (s), injury intention, location of incident, fire-related, work-related, race of patient, and comments. The dataset also included a sampling weight, a stratum, and a PSU for each record to include in the complex survey analysis. This study was limited to adults 18+ years of age who presented to one of the sampled EDs with an ER injury.

Measures

Product codes were used to identify ER injuries, which included events coded as non-equipment exercise injuries (product code 3299) as well as strength training activity injuries (product code 3265) [11]. Two different product code variables were included in the NEISS dataset. Therefore, if a record included the above codes in either product code variable, the record was considered presenting to an ED with an ER injury. The research at hand used the following variables in the study: hospital admission status (yes/no), body part (upper/lower), location (recreation or sport facility/other), age group (18-24/25-49/50-64/65+ years), race (white/black/other), and sex (male/female).

Statistical Analysis

Descriptive statistics consisted of unweighted sample frequencies and weighted estimates (%s) with 95% confidence intervals (CIs) of hospitals admission overall and across related characteristics. The Rao-Scott chi-square test of independence was used to determine significant differences in estimates. Logistic regression was used to model the bivariate relationship between each predictor variable and admittance status. Multiple logistic regression was used to simultaneous model the independent relationship of each predictor and admittance status. Odds ratios and 95% CIs were presented for all models. The CHAID decision tree procedure was used to determine the best set of categorical predictors associated with hospital admission. SPSS version 24 was used for CHAID [12] and SAS version 9.4 was used for descriptive statistics and modeling [13, 14]. A probability less than. 05 was used define statistical significance in all hypothesis tests.

Results

Table 1 displays distributions of ER injuries by admittance and other related characteristics. Overall, approximately 5.4% (95% CI: 3.1-9.3) of patients presenting to EDs with an ER injury were admitted to the hospital. Older patients were admitted as compared to younger ones. More other races were admitted as compared to their counterparts. And more patients presenting with upper-body injuries were admitted as compared to patients with lower-body injuries. Table 2 displays results of the bivariate and multiple logistic regression models. Results were fairly similar between the adjusted and unadjusted models. Most notably, odds of hospital admission increased linearly with age (p for trend <.001). Additionally, odds of hospital admission were greater for white and other patients, as compared to black patients with ER injuries. Lastly, odds of admission were greater for those with upper-body injuries, as compared to those with lower-body injuries.

Figure 3 displays the results of the CHAID analysis. The first distribution (far left box) represents the overall frequency of hospital admissions among those presenting to an ED with an ER injury. The first branch is the result of the CHAID analysis finding that the chi-square statistic was largest (χ^2=328.2, p<.001) when a two-by-three table is constructed with hospital admission status and age group, with the two younger age groups collapsed. Inspecting the three age group distributions (Nodes 1 thru 3), shows that hospital admissions was greatest (17.3%) among the older (65+ years) ED patients, as compared to the middle-aged (7.9%) and younger-aged (2.5%) groups. Continuing along the tree branch of the older age group, the next largest chi-square statistic (χ^2=22.8, p<.001) is when a two-by-two table was constructed with hospital admissions and race/ethnicity, with white and black groups collapsed. Examining the two race/ethnicity nodes shows that the race/ethnicity patients categorized as “other” had a greater rate (23.4%) of hospital admission, as compared to white and black patients combined (9.5%). No other factors were significantly related to hospital admissions among the older patients who were white or black. However, among older patients categorized as “other”, body part was a significant predictor of hospital admissions, yielding a two-by-two table (χ^2=18.1, p<.001). Among the final nodes on this branch, patients with an upper-body ER injury were more likely (31.9%) to be admitted to the hospital, as compared to those with lower-body injuries (13.5%).

The CHAID results for the other two newly formed age groups were similar in that body part was the next best predictor. However, significant factors were different from that point forward. Most notable, middle-aged patients with upper-body ER injuries who were categorized as “other” race/ethnicity were more likely to be admitted to the hospital (17.0%), as compared to the white and black combined group (4.7%).

Table 1: Distribution of ER Injuries by Admittance and Other Characteristics, U.S. EDs 2015.

Admitted for Hospitalization							
	Yes			No			
Characteristic	*N*	*%*	*95% CI*	*N*	*%*	*95% CI*	*p*
Overall	362	5.4	3.1-9.3	7,294	94.6	90.7-96.9	<.001
Sex							.482
Male	222	5.6	3.1-9.9	4,373	94.4	90.1-96.9	
Female	140	5.2	3.1-8.7	2,921	94.8	91.3-96.9	
Age (yr)							<.001
18-24	34	2.1	1.2-3.7	1,709	97.9	96.3-98.8	
25-49	114	3.1	1.7-5.5	3,915	96.9	94.5-98.3	
50-64	95	9.4	5.7-15.2	1,100	90.6	84.8-94.3	
65+	119	15.9	8.9-26.7	570	84.1	73.3-91.1	
Race/Ethnicity							.011
White	92	4.3	2.9-6.4	2,305	95.7	93.6-97.1	
Black	8	1.4	0.6-2.9	745	98.6	97.1-99.4	
Other	262	7.0	3.7-13.0	4,244	93.0	87.0-96.3	
Body part							.002
Upper-body	262	9.0	4.5-17.0	3,208	91.0	83.0-95.5	
Lower-body	100	2.4	1.4-4.1	4,086	97.6	95.9-98.6	
Location							.590
Recreation center	239	5.7	3.0-10.6	4,972	94.3	89.4-97.0	
Other	123	5.0	3.3-7.6	2,322	95.0	92.4-96.7	

Note. N is unweighted sample frequency. Estimates (%) are weighted using sample weights. p value is for the Rao-Scott chi-square statistic.

Table 2: Odds Associated with Admittance Due to an ER Injury, U.S. EDs 2015.

Admitted for Hospitalization						
	Unadjusted			Adjusted		
Characteristic	*OR*	*95% CI*	*p*	*OR*	*95% CI*	*p*
Sex						
Male	1.08	0.87-1.34	.482	1.16	0.95-1.41	.149
Female	1.00	-		1.00	-	
Age (yr)			<.001			<.001

18-24	1.00	-		1.00	-	
25-49	1.50	0.70-3.23		1.60	0.72-3.58	
50-64	4.89	2.16-11.06		4.96	2.23-11.03	
65+	8.85	3.56-22.03		8.43	3.36-21.13	
Race/Ethnicity			<.001			<.001
White	3.24	1.55-6.77		2.95	1.34-6.50	
Black	1.00	-		1.00	-	
Other	5.47	2.52-11.87		5.13	2.41-10.93	
Body part						
Upper-body	3.99	1.58-10.11	.004	3.60	1.39-9.30	.009
Lower-body	1.00	-		1.00	-	
Location						
Recreation center	1.00	-		1.00	-	
Other	1.13	0.72-1.79	.591	1.08	0.71-1.63	.729

Note. OR is odds ratio. CI is confidence interval. P value is for the Wald F statistic. Unadjusted column refers to simple bivariate models. Adjusted column refers to a single multivariate model.

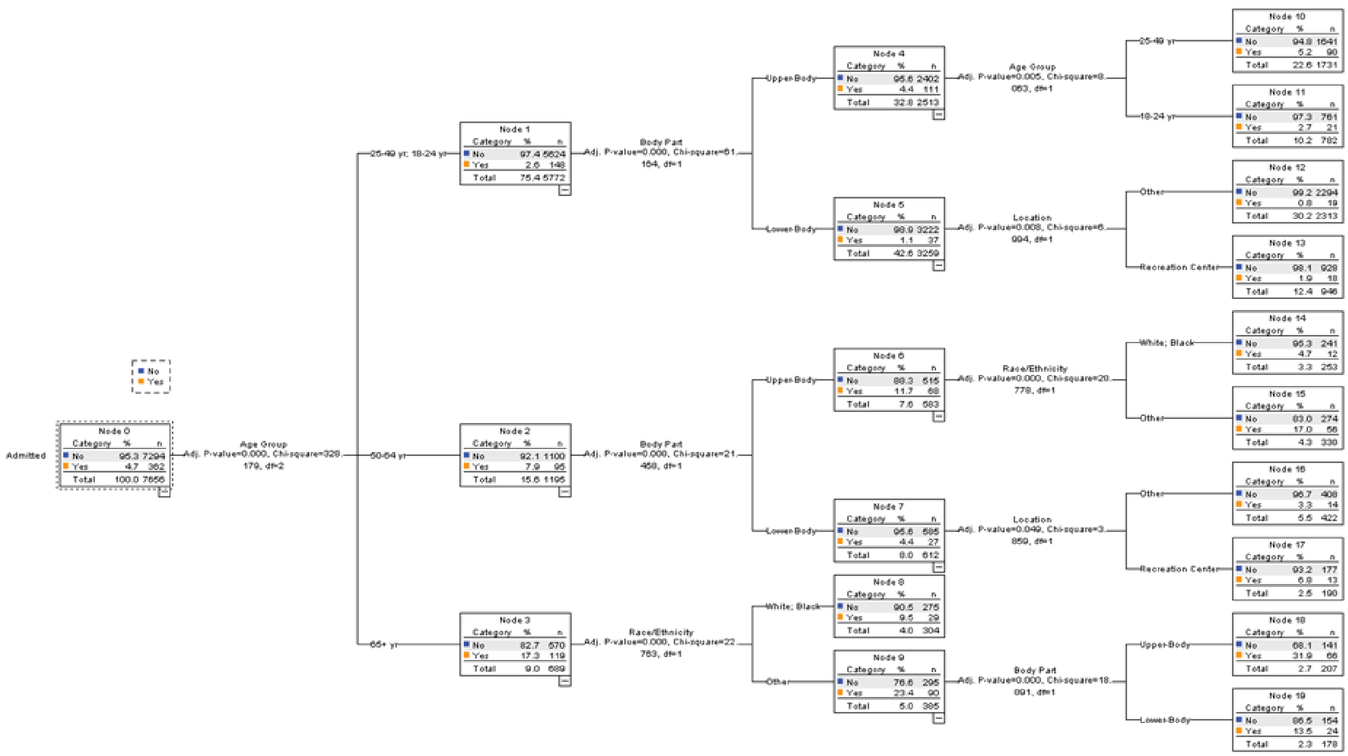

Fig 1: CHAID Tree Diagram of ER Injuries by Admittance and Other Characteristics, U.S. EDs 2015.

Discussion

The purpose of this study was to use a data mining decision tree tool (CHAID) to examine the most useful predictors of hospital admissions among adults presenting to an ED with an ER injury. Results showed that the most useful predictors of hospital admissions depended first on the age of the injured patient. Additionally, the highest rate of hospital admissions were seen in the older patients who were categorized as "other" race/ethnicity, and who suffered an upper-body injury. The novel aspect of this study, however, was the ability of CHAID to create all possible frequency tables relating to the root node of hospital admissions status [15]. Furthermore, the CHAID analysis was able to subdivide the sample while searching for the best predictors of hospital admissions [16]. This latter aspect of CHAID is innovative because it allows for protection against Simpson's Paradox [17, 18]. That is, the relationship between two variables becomes different in the presence of other variables. The current research specifically shows evidence of where CHAID provides such protection. That is, in the bivariate and multiple regression analyses, the location factor (recreation or sport facility/other) consistently showed to be unrelated to hospital admissions in ER injuries. However, in the CHAID analysis, location was a significant predictor of hospital admissions among the three younger-aged groups who had lower-body injuries. More specifically, among the three younger-aged groups who had lower-body ER injuries, those with injuries occurring at recreational/sport facilities were more likely to be admitted to the hospital, as compared to their counterparts. This specific finding would have been difficult to find without a procedure such as CHAID.

This study has limitations worth discussing. One limitation is the fact that these findings can only generalize to ER injuries presenting to U.S. EDs. That is, some adults in 2015 likely suffered an ER injury and sought medical care from providers other than an ED (e.g., urgent care, primary physician, athletic trainer, etc.). This fact should be considered when interpreting the findings of this study. Another limitation is the fact that the CHAID could not incorporate complex sampling variables into its analysis, which is normally required to adjust standard errors for statistical tests. Given this fact, follow-up analyses were performed using the SAS PROC SUREVYFREQ procedure [19, 20]. In these post-hoc analyses, all cross-tabulation chi-square statistics were run in their respective subpopulations. For example, among the 50-64 year old age group, a weighted two-by-two chi-square test was run on hospital admissions status and body part. All such weighted chi-square tests were run and statistical significance was confirmed in all tests. Although not a perfect solution, this procedure does support the validity of the CHAID results.

Conclusions

Results from this study support the use of a novel data mining tool to find specialized subgroups of ER injuries predictive of hospital admission. Moreover, older adults of race/ethnicity other than white or black, who suffer an upper-body ER injury are at greatest risk for hospital admission. Injury epidemiologists should consider decision tree methods when identifying useful predictors of important outcomes. Furthermore, health promotion efforts should focus on older adults and preventing upper-body ER injury.

References

1. De Rezende LF, de Sá TH, Markozannes G, Rey-López JP, Lee IM, Tsilidis KK, *et al*. Physical activity and cancer: an umbrella review of the literature including 22 major anatomical sites and 770 000 cancer cases. Br J Sports Med, 2017.
2. Smith AD, Crippa A, Woodcock J, Brage S. Physical activity and incident type 2 diabetes mellitus: a systematic review and dose–response meta-analysis of prospective cohort studies.
3. Kyu HH, Bachman VF, Alexander LT, Mumford JE, Afshin A, Estep K, *et al*. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study. BMJ. 2013-2016; 354:i3857.
4. Lee IM, Shiroma EJ, Evenson KR, Kamada M, LaCroix AZ, Buring JE. Accelerometer-Measured Physical Activity and Sedentary Behavior in Relation to All-Cause Mortality: The Women's Health Study. Circulation. Circulationaha, 2017.
5. Quick Stats. Percentage of U.S. Adults Who Met the 2008 Federal Physical Activity Guidelines for Aerobic and Strengthening Activity, by Sex—National Health Interview Survey, 2000–2014. MMWR Morb Mortal Wkly Rep. 2016; 65:485. DOI: http://dx.doi.org/10.15585/mmwr.mm6518a9
6. Keadle SK, McKinnon R, Graubard BI, Troiano RP. Prevalence and trends in physical activity among older adults in the United States: A comparison across three national surveys. Preventive medicine. 2016; 89:37-43.
7. Hart PD. Epidemiology of Exercise-Related Injuries Presenting to U.S. Emergency Departments: 10-year Trends. J Phy Fit Treatment & Sports. 2017; 1(3):555562.
8. American College of Sports Medicine. ACSM's guidelines for exercise testing and prescription. Lippincott Williams & Wilkins, 2017.
9. Liu X, Liu YY, Liu SH, Zhang XR, Du L, Huang WX. Classification tree analysis of the factors influencing injury-related disability caused by the Wenchuan earthquake. Journal of international medical research. 2014; 42(2):487-93.
10. US Consumer Product Safety Commission. NEISS Sample Design 1997-Present. US Consumer Product Safety Commission, 2001.
11. US Consumer Product Safety Commission. NEISS coding manual. US Consumer Product Safety Commission, 2017.
12. Song YY, Ying LU. Decision tree methods: applications for classification and prediction. Shanghai archives of psychiatry. 2015; 27(2):130.
13. Stokes ME, Davis CS, Koch GG. Categorical data analysis using SAS. SAS institute, 2012, 31.
14. Cody RP, Smith JK. Applied statistics and the SAS programming language. North-Holland, 1985.
15. Rao PH. BUSINESS ANALYTICS: An application focus. PHI Learning Pvt. Ltd, 2013, 29.
16. Ratner B. Statistical modeling and analysis for database marketing: effective techniques for mining big data. CRC Press, 2004, 2.
17. Everitt B, Skrondal A. The Cambridge dictionary of statistics. Cambridge: Cambridge University Press, 2002.
18. Kievit RA, Frankenhuis WE, Waldorp LJ, Borsboom D. Simpson's paradox in psychological science: a practical guide. Frontiers in psychology, 2013, 4.
19. Chen X, Gorrell P. An introduction to the SAS® survey analysis PROCs. NESUG, 2008.
20. Lewis TH. Complex Survey Data Analysis with SAS. CRC Press, 2016, 19.