Project 1: What influences self reporting of Concussion in undergraduate college students?

Subhrangshu Nandi, UW Madison Department of Statistics

Dr. Traci Snedden (PhD, RN, CPNP, CNE), UW Madison School of Nursing

April 6, 2017

Purpose

To investigate the factors that might influence "self-reporting" of **concussion** and why it may occur, in undergraduate college students.

Data

The data is from a survey send out to a large groups of undergraduate students at a large midwestern university. Overall response rate of the survey was 6.72% (1,950 out of 29,000). The response rate is slightly lower than the average response rate for external surveys which is around 10-15%. Out of 1,950 respondents, 76 of them reported to have had a concussion during since starting college. We analyzed this subset to investigate the factors that might influence "self-reporting" of concussion and the reasons why they may have occurred.

Summary

Of the 76 students who had concussion, the only 31 "self-reported", as shown in table 1. The plot (fig 1)

NA	YES	NO
8	31	37

Table 1: Self-reporting of concussion

below shows the distribution of ages of students in the three sub-groups - "YES", "NO" and Missing information
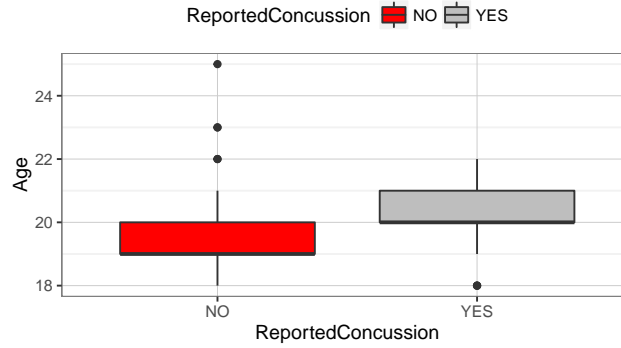
Figure 1: Distribution of Age of students who had concussion, separated by whether they self-reported or not

Figure 1 excludes one student whose age was 37 in the "YES" category. Apart from the three outlier points, students who did not self-report seem younger than the groups of students who did report. Table 2 shows the distribution of self-reporting behavior by gender.

	NA	YES	NO
Female	5	25	25
Male	3	6	12

Table 2: Self-reporting of concussion

A χ^2 test was conducted to test the hypothesis that self-reporting behavior was the same for both gender. Although the number of male students who did not self-report were twice that of female students, the difference in the distributions were statistically not significant. The p-value of the χ^2 test was 0.3465. Hence, we could not reject the null hypothesis. Below is the distrution of self-reporting behavior by "race".

	NA	YES	NO
Asian	3	0	3
Black or African American	0	0	1
Others	2	0	2
White	3	27	31

Table 3: Self-reporting of concussion by Race

There were not enough students in all the cells of table 3 to conduct a statistical test. There were also not enough students in all cells of "ethnicity" in table 4

	NA	YES	NO
Hispanic or Latino or Spanish Origin	0	2	1
Not Hispanic or Latino or Spanish Origin	6	28	36

Table 4: Self-reporting of concussion by Ethnicity

Table 5 below is the distribution of self-reporting behavior by academic status.

2

	NA	YES	NO
NA	1	0	0
Completed 1 year or less	1	5	10
Completed 1-2 years	1	3	9
Completed 2-3 years	4	14	8
Completed 3-4 years	1	8	6
Completed 4 years or more	0	1	4

Table 5: Self-reporting of concussion by Academic status

Table 6 displays the type of physical activities the students were involved in.

	NA	YES	NO
NA	0	0	1
NCAA-sanctioned D1 athlete	0	0	2
Club sport at UW-Madison	0	2	4
Intramural activities/sports at UW-Madison	1	7	10
Uses UW-Madison physical activity/exercise facilities	4	15	15
Does not participate in any activities	2	6	4
Others	1	1	1

Table 6: Self-reporting of concussion by type of activity

In both, tables 5 and 6, there was not enough representation in all the cells to conduct statistical hypothesis testing. Table 7 displays the summary when asked if this was an athletic injury that occurred playing any level of UW sport (NCAA sanctioned D1, UW Club or UW RecSports/Intramurals)

	NA	YES	NO
No	3	27	32
Yes, NCAA-sanctioned D1	0	0	2
Yes, UW Club	0	1	0
Yes, UW RecSports/Intramurals	0	3	3

Table 7: Self-reporting of concussion by reason

Only 9 out of 76 students reported that the injury had occurred playing some level of UW sport. When asked the number of hours after the injury occurred when they reported, following are the distributions by gender for those who reported within 4 days. In addition to these students, there were two students who reported after one week (7 days) and one student after two weeks. These were excluded from the boxplot in figure 2 to better illustrate the distribution of the rest of the students.
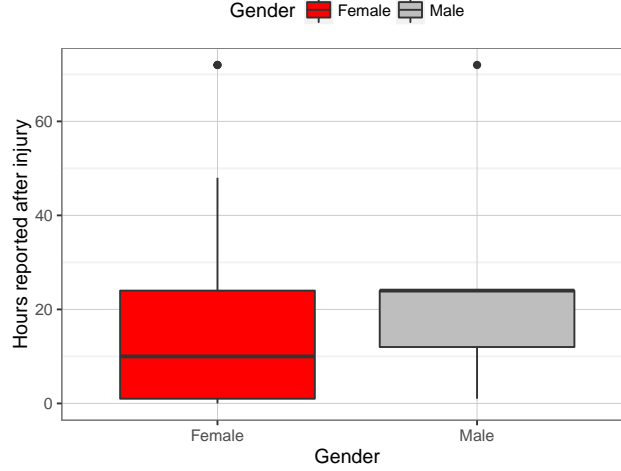
Figure 2: Distribution of Hours reported after injury, by Gender

Of the students who self-reported, when asked who they reported to, only 6 students reported to "parents" and 6 students reported to "others". Of the students who did not self-report, following are the reasons whey behind their not reporting.

	responses
Didnât know who to report it to	2
Didnât realize seriousness	9
Didnât think it was an injury	6
Didnât want to miss school	1
Other (specify)	5

Table 8: Why did they not report

Statistical Model

Of all the factors, Age and Gender (although not statistically significant), showed some difference in self-reporting behavior. Below is a plot which illustrates the possible interaction between Age and Gender. Older females and younger males seem to have self-reported. We build a model to predict the odds of self-reporting, from the student's age and gender. We fit a logistic regression model:

$$Y = \beta_0 + \beta_1 X + \beta_2 G \tag{1}$$

where, $Y = 1$, if self-reported, otherwise, $Y = 0$, X is the age of the student and G is the gender.
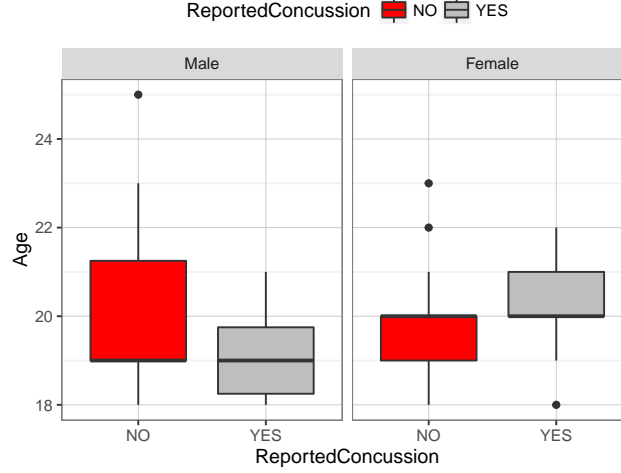
Figure 3: Distribution of Age, by Gender and self-reporting behavior

Logistic regression was used to estimate the probability of self-reporting. Two predictor variables - age of student and gender of student, were used in the analysis, with simultaneous entry of predictors. As shown in table 9, the overall predictive model was statistically significant (likelihood ratio chi-square = 10.15 [4], $p < 0.05$). The independent variables age and the interaction of age with gender were both significant in predicting the likelihood of self-reporting. On an average, older students were more likely to self-report. Females, on an average, were more likely to self-report. However, older males were less likely to self-report (OR = -2.19). Although the overall model and the two predictors were statistically significant, the classification results indicated modest success, with an overall rate of correct classification of 64.71%. The overall effect size was also modest, with the Nagelkerke R^2 equal to 10.83%.

Predictor	b(SE)	Wald	Odds Ratio	Lower OR	Upper OR
Constant	0.09 (0.31)	0.28	1.09	0.6	2.02
Age	0.66 (0.3)	2.23*	1.93	1.14	3.68
Male	-1.01 (0.68)	-1.49	0.36	0.07	1.24
Age:Male	-1.03 (0.47)	-2.19*	0.36	0.12	0.81

Table 9: Logistic Regression Results Predicting the Probability of Self Reporting

The coefficient of age was 0.659 ($p < 0.05$), which implied that with age the log odds of reporting increased, i.e., for every one year of age, the odds of self-reporting increased by 1.933 (which was $e^{0.659}$). For males, the odds of self-reporting was 0.363 (log odds = -1.012) compared to females, although the p-value was large. This was possibly due to the imbalance in the data (50 females vs 18 males). The interaction term between age and gender was -1.03 (p-value < 0.05), which implied that for males, the odds of self reporting reduces by 0.356 for every one year of age. This was consistent with the pattern in figure 3. The prediction model can be used for prediction: the probability of a 18 year old female to self-report was 20%, and male was 47%. But for a 22 year

old female it was 78%, but male was 17%.

Limitations

The summary and statistical findings, although interesting, have some limitations. The sample size was 76, and out of them 8 students did not respond to the question on self-reporting, further reducing the sample size to 68 students. Further analysis is necessary to assert if these non-respondents could be asserted as "missing-at-random". More respondents with information like the reason for injury, their activities, etc, would improve the predictive power of this model. A more balanced sample, in terms of the predictors would improve the robustness of the results. For example, if more males responded to the survey it would improved the gender balance.