Drug and Alcohol Dependence 197 (2019) 102–107

Contents lists available at ScienceDirect

Drug and Alcohol Dependence

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

Full length article

Development and validation of a risk predictive model for student harmful drinking—A longitudinal data linkage study

Duc Anh Ngo[a,c,*], Saumitra V. Rege[c], Nassima Ait-Daoud[b], Christopher P. Holstege[a,c]

[a] Department of Student Health, Division of Student Affairs, University of Virginia, 400 Brandon Avenue, Charlottesville, VA 22908, United States
[b] Department of Psychiatry and Neurobehavioral Sciences, University of Virginia School of Medicine, PO BOX 800223, Charlottesville, VA 22908-0699, United States
[c] Department of Emergency Medicine, University of Virginia School of Medicine, PO BOX 800699, Charlottesville, VA 22908-0699, United States

ARTICLE INFO

Keywords:
Alcohol intoxication
Emergency department
Risky drinking
Risk prediction
Data linkage

ABSTRACT

Background: This study aimed to develop a predictive model to quantify the risk of student harmful drinking associated with emergency department (ED) visits and/or campus-wide incidents reported to campus authorities in a U.S. public university.

Methods: Six-year (2010/11–2015/16) student enrollment data were linked to subsequent harmful drinking events defined as either alcohol intoxication associated with ED visits or alcohol-related incidents reported to authorities within 1 year following the annual (index) enrollment. Multivariable logistic regression analysis was used to develop a risk predictive model based on the first 3-year student cohort (n = 93,289), which was then validated in the following 3-year student cohort (n = 85,876).

Results: A total of 2609 students in the derivation cohort and 2617 students in the validation cohort had at least 1 harmful drinking event within 1 year following the index enrollment, providing an incidence of 2.8% and 3.1%, respectively. Student demographics (gender, age, ethnicity, parental tax dependency), academic level, Greek life member, transfer students, first-time enrolled students, having been diagnosed with depression or injury, and violence involvement were statistically significant predictors. C-statistics of the model were 0.86 in both cohorts, with excellent calibration and no evidence of over- or under-prediction observed from calibration plots.

Conclusions: By linking routinely collected student data, a robust risk predictive model was developed and validated to quantify absolute risk of harmful drinking for every student. This model can provide a useful tool for clinicians or health educators to make real time decision to plan target interventions for students at elevated risk.

1. Introduction

Alcohol misuse is well-recognized as a significant public health concern among college students. In 2017, for example, the past month use of alcohol (62% vs 56.4%) and mixing alcohol with energy drinks (31.5% vs 26.7%) was higher among US college students as compared to non-college peers (National Institute on Drug Abuse, 2018). In addition, numerous studies have linked student harmful alcohol use to a host of negative personal and campus-wide problems including poor academic performances, high-risk sexual behavior, sexual assault, violence and aggression, serious injuries, driving while intoxicated, and traffic fatalities (Ngo et al., 2018a; NIAAA, 2013; Ansari et al., 2013; Beck et al., 2010; Hingson et al., 2009; Wechsler et al., 2002). Considering these issues, most universities have implemented strict alcohol policies on college campuses. However, despite these policies, alcohol misuse and its associated problem behaviors are frequently cited reasons for disciplinary infractions and student visits to the emergency departments (ED).

Although harmful alcohol use is common among the student populations, variation in the risk of this problem behavior exists, dependent on a host of personal, inter-personal, and campus-related factors. It is well-established that student race/ethnicity (i.e., white students), first year students (Wechsler et al., 1995), and students affiliated with fraternities/sororities (Lo and Globetti, 1995; Presley et al., 2002) or students participating in athletic activities (Kwan et al., 2014; Lisha and Sussman, 2010; Turrisi et al., 2006) have higher rates of episodic heavy drinking than other students. The type of residence and the college size also affect the level of binge drinking (NIAAA, 2002). Psychiatric and emotional disorders, including psychological distress (Livingston et al., 2016), social anxiety (Brook and Willoughby, 2016; Dawson et al., 2005), or depression

* Corresponding author at: Department of Student Health, University of Virginia, 400 Brandon Avenue, Charlottesville, VA 22903, United States.
E-mail address: dan9k@virginia.edu (D.A. Ngo).

https://doi.org/10.1016/j.drugalcdep.2019.01.016
Received 19 September 2018; Received in revised form 22 January 2019; Accepted 23 January 2019
Available online 19 February 2019
0376-8716/

have been found to commonly occur together with alcohol use disorders among college students (Baer, 2002). The variability in the risk of student alcohol misuse signifies the need to identify students at higher risk to target intervention strategies to modify excessive drinking behaviors in order to curb problem drinking and unintended consequences among this emerging adult population.

For decades, electronic records have widely been used to collect and store student data and information. For example, enrollment data captures a broad spectrum of student information and characteristics; many of these are established risk factors for harmful drinking, including demographics (age, gender, and ethnicity), academic levels and programs, organizational affiliations (e.g., Greek life members), and extracurricular activities (e.g., athletic participation) (Ngo et al., 2018b). The university incident report system records a wide-range of alcohol related problems, including underage drinking, alcohol intoxication, high-risk sexual behavior, mental health problems, sexual assault, violence and aggression, injuries, and traffic fatalities (Bernat et al., 2014). This provides a great opportunity to link available data to create a single, comprehensive, and integrated student dataset (Dusetzina et al., 2014), which allows for longitudinal follow up of students over the course of study and development of robust multivariable predictive models to facilitate risk stratification and improve the identification of students at increased risk of harmful alcohol use.

While a number of risk assessment tools (Garcia Carretero et al., 2016; Blank et al., 2015) were developed, these tools have typically relied on questionnaire-based interviews and have focused on screening for binge drinking. This survey-based screening method is known to have major limitations associated with recall bias that could lead to misclassification of student drinking amount. Growing evidence also indicates that self-reported alcohol consumption is unreliable at high levels of drinking and cannot capture comprehensively multiple dimensions of student risky drinking (White et al., 2006; Devos-Comby and Lange, 2008). As this, available screening tools may not necessarily identify student harmful alcohol use beyond the binge threshold, particularly when it results in the need for an ED visit or is brought to attention of university authorities. In addition, it is not feasible to implement survey-based screening on a large scale at the population level.

The aim of this study, therefore, was to develop and validate a novel risk predictive model to predict absolute risk of student harmful drinking defined as either acute alcohol intoxication associated with ED visits or alcohol-related incidents recorded in the university incident report system. The primary purpose is to derive a risk score for every student based on established risk markers captured in available student datasets rather than to identify novel risk markers. We hypothesized that the proposed risk model would be able to identify high-risk students as soon as when they first enrolled, creating the opportunity to initiate and target early interventions.

2. Methods

2.1. Study Population, Study Design, and Data Sources

Our study population consisted of students aged 16–49 enrolled in a U.S. public university from 2009/10 to 2015/16 academic years. We used an open cohort design, which allows students to enter the study population throughout the study period rather than on a fixed entry date. Data for this study were created by linking the following 5 student electronic datasets.

Student Information System (SIS) is the university's student registry database that records a wide range of student information associated with enrollment and gives every student a university unique student identification number (UID). *Student Health Record Dataset* is a census of all student visits to the Student Health Center (SHC) clinic. The *ED's Patient Registration System* is a reporting system of students visiting the university hospital ED. These students were matched to their UID based on the medical record number (MRN) given to those who had a prior visit to the university health system, or first and last names if an MRN was not available among those without a prior visit to the university health system. *Clinical Data Repository* is an electronic data repository of patient admissions and visits to all clinics and departments in the university health system, which contains clinical records including ICD diagnostic codes associated with each ED visit. A complete description of these 4 datasets was published elsewhere (Ngo et al., 2018b). *The Incident Management Response System (IMRS)* is an electronic dataset recording information on the date, time, location, and types of incidents that involved alcohol, drug, or violence occurring on or off ground on campus or elsewhere in the city/county where students live. Major types of alcohol-related incidents included underage drinking and alcohol possession, being drunk in public places, negative behaviors under influence of alcohol (e.g., violence, assault, interpersonal conflicts, property damage), and psychological disorders or mental illness related to alcohol use (e.g., suicidal ideation).

2.2. Cohort Construction and Unit of Analysis

The study cohort consisted of all enrolled students who entered the cohort at their first enrollment (index enrollment) in each academic year during the period studied. The unit of analysis is the student annual enrollment (not individual student). An individual student may be evaluated multiple times for the risk of harmful alcohol use during the study period, corresponding to the number of years the student enrolled.

2.3. Data Linkage

The SIS records of the study cohort were linked using the student UID and academic term when the student enrolled with the 3 subsets of students: those who visited the university SHC clinic, those who visited the university hospital ED, and those with alcohol-related incidents recorded in the IMRS. Through this process, we obtained a comprehensive, integrated, de-identified dataset containing full records of student demographic and academic characteristics, organizational affiliation and extra-curricular activities, and clinical histories (of those who visited the SHC clinic) linked to a subset of students who engaged in risky drinking that resulted in ED visits or was reported to the university authorities.

2.4. Study Outcome and Follow Up

The study outcome was the **first** harmful drinking event defined as either an ED visit with alcohol intoxication or an alcohol-related incident recorded in the IMRS within one year since the index enrollment (entry date). ED visits with alcohol intoxication were identified using International Classification of Disease (ICD)-9 (305.0, 303.0) or ICD-10 (F10120, F10121, F10129) codes (NIAAA, 2013) documented in the patient ED clinical records. The physician's diagnosis of alcohol intoxication was primarily based on clinical presentation, a documented elevated alcohol level, and/or the patient self-report of significant drinking before the ED visit. Alcohol-related incidents were those flagged as involving alcohol use recorded in the IMRS.

Follow up started at the date when the student first enrolled (index enrollment) in each academic year during their course of study and ended at the earliest date of the following: the end date of the student's last semester, the date when the student was transferred to another university or withdrew, or day 365th following the index enrollment.

2.5. Identification of Risk Markers

We compiled a list of potential risk markers which have previously been established to increase the risk of excessive alcohol use among college students. Essentially, potential risk markers included the following variables:

A Demographic variables: age (16–19, 20–24, 25–29, 30–49), gender, race/ethnicity, veteran status, transfer/exchange students, citizenship status (domestic or foreign students), parental tax dependency (defined as 'one who is listed as a dependent on the federal or state income tax return of his parents or legal guardian') (University of Virginia, 2019)
B Organizational participation and sport-related activities: university-affiliated athletic teams, fraternities/sororities (e.g., Greek house member)
C Academic affiliation: school, academic level, academic year, transfer students
D Use of Counseling and Psychological Services (CAPS)
E Clinical risk markers: past year alcohol-related visits to the SHC clinic, having been given a diagnostic code for depression, anxiety, or injury
F Violence involvement

Risk markers listed in A, B, C were ascertained from SIS and risk markers listed in D, E were abstracted from the Student Health Record Dataset based on ICD codes within 1 year before the index enrollment. As high as 79.5% of the total enrolled students register at the SHC clinic as their primary care provider. Violence involvement was identified from IMRS in the year preceding the index enrollment.

2.6. Model Derivation and Development

A retrospective cohort including students whose index enrollment occurred in the first 3 academic years (2010/11–2012/13) was assembled to derive a model predicting the 1 year risk of student harmful drinking incidents. Step-wise multivariable logistic regression models were performed to provide coefficients and Odds Ratios (ORs) for each risk marker that was found significantly associated with the occurrence of **first** harmful drinking event within 1 year following the entry date. The analysis started with univariable analysis to identify the risk markers that were likely to be associated with the outcome ($p < 0.25$), and these variables were subsequently fitted into the initial multivariable model. Of the variables that were not significantly associated with the outcome, the one with the highest p-value was removed and the model was refitted with the remaining variables. Only variables that were significantly associated with the outcome were retained in the final model. The significance level was defined at 5% ($p < 0.05$) for non-clinical risk markers and at 10% ($p < 0.1$) for clinical risk markers (i.e., past year alcohol-related SHC visits, depression, anxiety, injury) to take into account the clinical importance of these factors in relation to student risky drinking.

The coefficient for each variable from the final model was used to create a risk predictive equation. This equation was then used to generate a risk score for each student, which was then converted to a 1–100 point scale, with higher scores indicating a higher probability of having a subsequent harmful drinking events in the year following the index enrollement.

The predictive performance of the model was evaluated by C-statistics (95% CI) as a measure of discrimination where values of 0.9–1.0 indicate outstanding discrimination, 0.8-0.9, excellent discrimination, 0.7-0.8, good discrimination, and < 0.7, poor discrimination (Lisha and Sussman, 2010). Hosmer-Lemeshow test and calibration plot, a plot assessing agreement between predicted and observed risks cross tenths of predicted risk, were used to assess model calibration (Hosmer et al., 2013; Nieboer et al., 2016; Steyerberg et al., 2010). Calibration slope (95% CI) was calculated. A bootstrapping technique was used to adjust for overfitting and overoptimistic model performance. An optimism-corrected C-statistics (95% CI) using 100 bootstrap samples created with replacement was reported. The sensitivity (SSV), specificity (SPV), positive predictive value (PPV), and negative predictive value (NPV) for selected percentiles of predicted risk values over the 1 year follow-up were calculated.

2.7. Model Validation

The final model was retrospectively tested on the independent validation cohort that consists of students whose index enrollment occurred in the next 3 academic years (2013/14–15/16). We repeated analysis performed on the derivation cohort to evaluate the model performance on the validation cohort.

All analyses used SAS statistical software, version 9.4. The reporting of the present study follows the Transparent Reporting of multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement (Collins et al., 2015). Ethical approval was provided by the University Institutional Review Board.

3. Results

3.1. Characteristics of the Study Population

The study population consisted 183,823 students aged 16–49 enrolled during the 6-year period of the study. After excluding 4658 (2.5%) due to missing data in one or more covariates, 179,165 students were available for analysis. Over the entire study period, 5226 students were identified to have at least 1 harmful drinking event of whom 924 (17.7%) had the first event associated with ED visits due to alcohol intoxication.

The derivation cohort comprised of 93,289 students, 2609 of whom had at least 1 harmful drinking event within 1 year from the date of index enrollment, and the validation cohort comprised of 85,876 students, 2617 of whom had at least 1 harmful drinking event within 1 year after the index enrollment. The incidence was 2.8% in the

Table 1
Characteristics of the study population.

Variable	Derivation cohort		Validation cohort	
	n	%	n	%
Gender				
Female	52,575	56.4	47,353	55.1
Male	40,714	43.6	38,523	44.9
Age				
16-19	23,469	25.2	24,556	28.6
20-24	36,030	38.6	35,503	41.3
25-29	13,295	14.3	11,428	13.3
30-49	20,495	22.0	14,389	16.8
Ethnicity				
Asian	7,870	8.4	8,021	9.3
Black or African American	5,973	6.4	4,970	5.8
Hispanic	4,225	4.5	4,263	5.0
Multi-Race	2,238	2.4	3,031	3.5
Non-Resident Alien	7,040	7.6	7,153	8.3
White	57,404	61.5	52,115	60.7
Other	8,539	9.2	6,323	7.4
Citizenship				
International	7,040	7.6	7,153	8.3
US	86,249	92.5	78,723	91.7
Parental tax dependency (Yes)	43,669	46.8	48,925	57.0
First-time enrollment (Yes)	37,133	39.8	31,771	37.0
Academic level				
Graduate	42,281	45.3	32,790	38.2
Undergraduate	51,008	54.7	53,086	61.8
Transfer students (Yes)	46,68	5	5,500	6.4
Greek life (Yes)	10,435	11.2	12,839	15.0
Athlete (Yes)	3,807	4.1	4,145	4.8
Past alcohol-related clinic visits (Yes)	153	0.2	159	0.2
Depression (Yes)	882	1.0	786	0.9
Anxiety (Yes)	1,201	1.3	824	1.0
Having used CAPS* (Yes)	1,697	1.8	2,859	3.3
Having an injury (Yes)	1,169	1.3	1,152	1.3
Involving in violence (Yes)	543	0.6	583	0.7

* Counseling and Psychological Services.

Table 2
Adjusted OR (95%CI) for the risk of student hazardous drinking in derivation and validation cohorts.

Predictor	Derivation cohort OR (95%CI)	Validation cohort OR (95%CI)
Gender (Referent = Female)	2.00 (1.84, 2.17)	1.87 (1.72, 2.03)
Age		
16-19	2.07 (1.51, 2.82)	2.68 (1.82, 3.96)
20-24	0.96 (0.71-1.29)	1.29 (0.88, 1.88)
25-29	Referent	Referent
30-49	0.10 (0.05, 0.17)	0.15 (0.07, 0.29)
Ethnicity (Referent = Asian)		
African American	1.40 (1.12, 1.75)	1.38 (1.11, 1.72)
Hispanic	1.73 (1.39, 2.16)	1.51 (1.22, 1.86)
Multiracial	1.67 (1.30, 2.15)	1.41 (1.11, 1.78)
White	1.63 (1.39, 1.90)	1.41 (1.22, 1.64)
Other	1.13 (0.93, 1.39)	0.96 (0.79, 1.18)
Parental tax dependency (Referent = No)	2.16 (1.85, 2.52)	2.44 (2.04, 2.92)
Greek life member (Referent = No)	2.15 (1.96, 2.35)	2.06 (1.89, 2.25)
Transfer student (Referent = No)	0.75 (0.62, 0.90)	0.58 (0.48, 0.71)
Academic level (Referent = Graduate)	4.54 (3.60, 5.73)	4.40 (3.40, 5.69)
First time enrollment (Referent = No)	2.98 (2.70, 3.29)	4.67 (4.20, 5.19)
Depression (Referent = No/Unknown)	1.42 (0.94, 2.16)	1.74 (1.11, 2.72)
Injury (Referent = No/Unknown)	1.80 (1.30, 2.48)	1.58 (1.10, 2.29)
Violence (Referent = No/Unknown)	1.85 (1.25, 2.73)	1.58 (1.00, 2.51)

derivation cohort and 3.1% in the validation cohort, respectively (Data not shown).

The baseline characteristics of students in derivation and validation cohorts are presented in Table 1. Both cohorts demonstrated similar distributions of socio-demographic characteristics, sport activity participation, having been diagnosed with depression and/or anxiety, having an injury, involving in violence in the year preceding the index enrollment. Exception included there were less students with parental tax dependency (47% vs. 57%) and less students having used CAPS (1.8% vs. 3.3%) in the derivation cohort. In both cohorts, very few students (0.2%) were given a diagnostic code for alcohol use in their prior visits to the SHC clinic.

3.2. Predictive Risk Markers

Table 2 shows the results of multivariable logistic regression analysis for the derivation and validation cohorts. Of the 16 variables tested (Table 1), 11 were significantly associated with the risk of student harmful drinking in the final multivariable model. With regards to demographic factors, males (vs. females), students under 20 (vs. 25–29) years of age, White, African American, Hispanic, multiracial (vs. Asian) students, students having parental tax dependency, and students affiliated with fraternities/sororities were more likely than others to engage in harmful drinking behavior ($p < 0.05$). Regarding academic involvement, undergraduate students and first-time enrolled students were more likely whereas students who were transferred from an earlier institution were less likely than others to experience such harmful drinking events ($p < 0.05$). In terms of clinical and behavioral risk markers, students who were diagnosed with depression ($p < 0.1$) or had a prior visit due to injuries to the SHC clinic ($p < 0.05$), and students who were reported to IMRS for violence involvement in the year before the index enrollment ($p < 0.05$) were at higher risk. Student citizenship (American vs. international), student athletes, use of CAPS, past year alcohol-related clinic visits, and having been diagnosed with anxiety were not statistically significant predictors ($p > 0.1$).

3.3. Model Development

In the derivation cohort, the C- statistic was 0.856 (95% CI: 0.849, 0.862). The agreement between the observed and predicted proportion of events showed excellent apparent calibration (Fig. 1, left) with a calibration slope of 1.070 (95% CI: 1.067, 1.073). The optimism-corrected C-statistics of the risk score in the derivation cohort, by the bootstrapping technique, was 0.856 (95% CI: 0.855, 0.857). For top 1% of students at higher risk, SSV, SPV, and PPV of the predicted risk scores were 11.8%, 99.2%, and 30.3%, respectively. A reduction of the SPV from 99.2% to 50.7% corresponding to top 1% and 50% of students at high risk increased SSV from 11.8% to 94.9%, but it reduced PPV from 30.3% to 5.3% (Table 3).

3.4. Model Validation

Applying our final risk model to the independent validation cohort gave a C-statistics of 0.857 (95% CI: 0.850, 0.864), and excellent calibration (Fig. 1, right) with calibration slope only slightly above 1 (1.070 (95% CI: 1.067, 1.073)). SSV and SPV of the derivation and validation cohorts at different risk cores thresholds were the same or similar. For example, top 5% of students at highest risk (e.g., those with scores of 95 or higher), derivation cohort had 23.7% SSV and 95.9% SVP, and the validation cohort had 23.7% SSV and 96.8% SPV, respectively (Table 3).

4. Discussion

To our knowledge, this is the first study to date that developed and validated a novel risk predictive model to calculate the absolute risk of student harmful drinking based on a complete student population. Overall the prediction model had strong discrimination power and excellent calibration. The model incorporates 11 predictive markers, including sociodemographic characteristics, academic, campus-related, behavioral, and clinical factors, most of them were well-documented in previous studies to increase the risk of student problem drinking. It is important, however, to note that being a member of a university athletic team was not a statistically significant predictor. A possible explanation is that athletic students are better supervised or required to abstain from drinking during sport seasons or risk the loss of their athletic scholarship. Also, athletic students in this particular campus might be reluctant to engage in risky drinking due to policies or disciplinary actions pertaining to student athletes' alcohol use or specific athletic leadership.

Epidemiological research has documented that while binge drinking has been decreasing among adolescents and young adults, the use of alcohol-related hospital services – particularly of ED visits has been increasing (White et al., 2018; NIAAA, 2013). Similar to a recent study

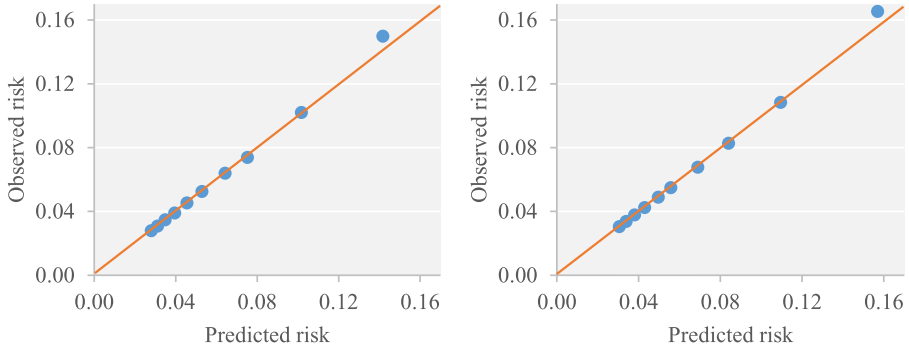

Fig. 1. Model calibration in the derivation cohort (left) and validation cohort (Right).

Table 3
Performance for predicting the risk of student hazardous drinking in the derivation and validation cohorts.

Risk score	Derivation cohort					Validation cohort				
	Total high risk students	SSV	SPV	PPV	NPV	Total high risk students	SSV	SPV	PPV	NPV
Top 1%	1021	11.8	99.2	30.3	97.5	1122	14.2	99.1	33.2	97.4
Top 5%	4664	23.7	95.9	20.5	98.1	3506	23.7	96.8	23.7	97.8
Top 10%	9678	55.6	90.9	15.0	98.6	9491	60.0	90.5	16.5	98.6
Top 25%	23524	77.7	76.3	8.6	99.2	21469	78.4	76.7	9.6	99.1
Top 50%	47137	94.9	50.7	5.3	99.7	44809	94.0	49.1	5.5	99.6

by White et al. that further confirmed that ED visits involving acute alcohol consumption among young people aged 18 years or older rose significantly between 2006 and 2014 (White et al., 2018), our current study demonstrates a rising trend in the annual incidence (from 2.65% in 2010/11 to 3.27% in 2014/15 and 2.91% in 2015/16 (data not shown)) of student harmful alcohol use other than binge drinking, which included acute intoxication associated with ED visits. In addition, several studies of non-student populations in other Western countries (Bertholet et al., 2014; Haberkern et al., 2010; O'Farrell et al., 2004; Verelst et al., 2012) also reported rising trends in prevalence of ED visits with alcohol intoxication. These conflicting epidemiological trends indicate that available intervention strategies focused on binge drinking might not been effective in reducing student harmful alcohol beyond the binge drinking thresholds. There is a strong need to reorient and expand the focus of interventions to address the rising burden of ED visits or campus-wide incidents due to heavy alcohol consumption. Screening for high risk students would be the first important step in addressing this apparent need.

The risk prediction model developed from this study offers a complementary tool to screen student dangerous alcohol use beyond the binge threshold. The majority of students seek care at the university SHC clinic where information on student demographic, campus-related, and clinical factors is routinely collected and readily available to primary care physicians. Hence, a risk screening tool can easily be developed and adapted in such a setting to facilitate physicians to identify high risk students for better targeting of clinic-based intervention efforts and make expedited referral to available interventions and services. Given potential stigma associated with alcohol misuse among the student populations, care should be taken in communicating with students who are identified and labelled as at high risk to prevent unintended psychological consequences.

Our risk predictive model has numerous strengths. It was based on absolute risks derived and validated in 2 large and independent cohorts of a complete student population with limited missing data. In addition, it is built by comprehensively linking student administrative and primary healthcare datasets with ED clinical data and the university incident management system to create an integrated dataset. Such dataset allowed for more complete ascertainments of student harmful drinking incidents and captured a wide range of risk markers which may otherwise not be available when only a single data source was used. In particular, by linking enrollment data with subsequent harmful drinking incidents, follow up was complete for all students since they first enrolled until the end of their study at the university. In addition, measures of student harmful alcohol use were ascertained from ED's physician diagnosis and incident reports, which is not subject to self-reporting or recall bias often occurring in sample-based surveys. The modeling approach was straightforward, which can easily be applied and adapted to further external validation in other universities with a similar affiliated health system and hospital ED where both routinely collected student administrative and clinical data are available and linked.

Our study is not without limitations. First, records of alcohol-related incidents in the IMRS was typically based on bystanders' reports, which might be subject to misclassification of student harmful drinking, especially when the incident involved a group of students where not all of them were heavy drinkers. In addition, it might be possible that not all alcohol- as well as non-alcohol-related incidents were reported and subsequently recorded in the IMRS, resulting in under-reporting of both the outcome and risk markers ascertained from this dataset (i.e., violence). Second, our earlier validation study found that only 66% of student ED visits with alcohol intoxication were captured by ICD diagnostic codes (Ngo et al., 2018c). Third, data on clinical risk markers (e.g., depression, anxiety) were only available among students visiting the SHC clinic, thus excluding those with these conditions who did not present to this clinic. Fourth, the findings are limited to a public university campus with an affiliated health system, specific campus policies and incident recording systems, which may not be generalizable to other universities. However, the observed relationships with key student characteristics (e.g., age, gender, Greek life member) were consistent with earlier studies, indicating that risk markers identified reflect generalizable risk patterns of harmful alcohol use among student populations and the predictive model can reliably identify students at higher risk.

The predictive model can be used to develop a complementary tool which can be automatically populated on the university computer systems to facilitate risk assessment for every student, especially for those who already had contact with the health system through the SHC clinic. Stratification of students by risk scores can help planning

interventions, as high risk students tend to require more intense interventions and more resources. Based on routinely collected student data, a risk score for every student can automatically be generated and available to health educators at the time of enrollment and to primary care physicians when students visit the SHC clinic so that they would refer students with higher risk scores to interventions. In addition, a proactive management and treatment of clinical risk markers (i.e., depression) not only benefit general health of students with these conditions but also have impact to lower the risk of subsequent harmful drinking and related consequences as well as to reduce the need for alcohol-related ED visits.

This study provides an example of utility and usefulness of linking routinely collected student data to develop and validate a predictive model that estimates the probability of student harmful drinking incidents within the next year after enrollment. Based on a complete student population, the model was able to derive a risk score for every individual student each time a student enrolled. The risk model developed from the current study can easily be externally validated in student populations from other universities, offering a useful tool for healthcare providers and health educators to make real time decision on student absolute risk of harmful drinking and plan targeted interventions or make expedited referral to outreach education or preventive services for students at elevated risk.

Conflict of Interest

All authors declare that they have no conflicts of interest.

Contributors

CH conceived the dataset. AN conceived the study design, extended data linkages, performed the analyses, and drafted the manuscript. All authors contributed to interpretation of the results and writing the manuscript. All authors approved the final version of the manuscript before submission. None of the original material contained in this manuscript has been submitted for consideration nor will any of it be published elsewhere

References

Ansari, W.E., Stock, C., Mills, C., 2013. Is alcohol consumption associated with poor academic achievement in university students? Int. Prev. Med. 4, 1175–1188.

Baer, J., 2002. Student factors: understanding individual variation in college drinking. J. Stud. Alcohol S14, 40–53.

Beck, K.H., Kasperski, S.J., Caldeira, K.M., Vincent, K.B., O'Grady, K.E., Arria, A.M., 2010. Trends in alcohol-related traffic risk behaviors among college students. Alcohol Clin. Exp. Res. 34, 1472–1478.

Bernat, D.H., Lenk, K.M., Nelson, T.F., Winters, K.C., Toomey, T.L., 2014. College law enforcement and security department responses to alcohol-related incidents: a national study. Alcohol Clin. Exp. Res. 38, 2253–2259.

Bertholet, N., Adam, A., Faouzi, M., Boulat, O., Yersin, B., Daeppen, J.B., Clerc, D., 2014. Admissions of patients with alcohol intoxication in the Emergency Department: a growing phenomenon. Swiss Med. 144, w13982.

Blank, M.L., Connor, J., Gray, A., Tustin, K., 2015. Screening for hazardous alcohol use among university students using individual questions from the Alcohol Use Disorders Identification Test-Consumption. Drug Alcohol Rev. 34, 540–548.

Brook, C.A., Willoughby, T., 2016. Social anxiety and alcohol use across the university years: adaptive and maladaptive groups. Dev. Psychol. 105, e1–2.

Collins, G.S., Reitsma, J.B., Altman, D.G., Moons, K.G., 2015. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 350, g7594.

Dawson, D.A., Grant, B.F., Stinson, F.S., Chou, P.S., 2005. Psychopathology associated with drinking and alcohol use disorders in the college and general adult populations. Drug Alcohol Depend. 77, 139–150.

Devos-Comby, L., Lange, J.E., 2008. Standardized measures of alcohol-related problems: a review of their use among college students psychology of addictive behaviors. Psychol. Addict. Behav. 22, 349–361.

Dusetzina, S.B., Tyree, S., Meyer, A.M., 2014. Linking Data for Health Services Research: A Framework and Instructional Guide [Internet]. Agency for Healthcare Research and Quality (US), Rockville (MD). Sep [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK253313/ (Accessed December 2016).

Garcia Carretero, M.A., Novalbos Ruiz, J.P., Martinez Delgado, J.M., O'Ferrall Gonzalez, C., 2016. Validation of the alcohol use disorders identification test in university students: AUDIT and AUDIT-C. Adicciones 28, 194–204.

Haberkern, M., Exadaktylos, A.K., Marty, H., 2010. Alcohol intoxication at a university hospital acute medicine unit–with special consideration of young adults: an 8-year observational study from Switzerland. Emerg. Med. J. 27, 199–202.

Hingson, R.W., Zha, W., Weitzman, E.R., 2009. Magnitude of and trends in alcohol-related mortality and morbidity among U.S. College students ages 18-24, 1998-2005. J. Stud. Alcohol Drugs Suppl. 16, 12–20.

Hosmer, D., Lemeshow, S., Studrdivant, R.X., 2013. Applied Logistic Regression. John Wiley Sons, Inc.

Kwan, M., Bobko, S., Faulkner, G., Donnelly, P., Cairney, J., 2014. Sport participation and alcohol and illicit drug use in adolescents and young adults: a systematic review of longitudinal studies. Addict. Behav. 39, 497–506.

Lisha, N.E., Sussman, S., 2010. Relationship of high school and college sports participation with alcohol, tobacco, and illicit drug use: a review. Addict. Behav. 35, 399–407.

Livingston, N., Christianson, N., Cochran, B., 2016. Minority stress, psychological distress, and alcohol misuse among sexual minority young adults: a resiliency-based conditional process analysis. Addict. Behav. 63, 125–131.

Lo, C.C., Globetti, G., 1995. The facilitating and enhancing roles Greek associations play in college drinking. Int. J. Addict. 30, 1311–1322.

National Institute on Drug Abuse, 2018. Drug and Alcohol Use in College-Age Adults in 2017 [Online]. Available: https://www.drugabuse.gov/related-topics/trends-statistics/infographics/drug-alcohol-use-in-college-age-adults-in-2017 (Accessed January 2019). .

Ngo, D.A., Ait-Daoud, N., Rege, S.V., Ding, C., Gallion, L., Davis, S., Holstege, C.P., 2018a. Differentials and trends in emergency department visits due to alcohol intoxication and co-occurring conditions among students in a U.S. Public university. Drug Alcohol Depend. 183, 89–95.

Ngo, D.A., Ait-Daoud, N., Rege, S.V., Ding, C., Gallion, L., Davis, S., Holstege, C.P., 2018b. Trends in incidence and risk markers of student emergency department visits with alcohol intoxication in a U.S. Public university – a longitudinal data linkage study. Drug Alcohol Depend. 188, 341–347.

Ngo, D.A., Ait-Daoud, N., Rege, S.V., Ding, C., Gallion, L., Davis, S., Holstege, C.P., 2018c. Validity of diagnostic coding of alcohol intoxication associated with student university hospital emergency department visits. J. Addict. Med. 12, 499–500.

NIAAA, 2002. High Risk Drinking in College: What We Know and What We Need to Learn. U.S. Department of Health and Human Services.

NIAAA, 2013. Alcohol-related emergency department visits and hospitalizations and their co-occurring drug-related, mental health, and injury conditions in the United States: Findings from the 2006-2010 nationwide emergency department sample (NEDS) and nationwide inpatient samples (NIS). Bethesda, MD. .

Nieboer, D., van der Ploeg, T., Steyerberg, E.W., 2016. Assessing discriminative performance at external validation of clinical prediction models. PLoS One 11, e0148820.

O'Farrell, A., Allwright, S., Downey, J., Bedford, D., Howell, F., 2004. The burden of alcohol misuse on emergency in-patient hospital admissions among residents from a health board region in Ireland. Addiction 99, 1279–1285.

Presley, C.A., Meilman, P.F., Leichliter, J.S., 2002. College factors that influence drinking. J. Stud. Alcohol S14, 82–90.

Steyerberg, E.W., Vickers, A.J., Cook, N.R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M.J., Kattan, M.W., 2010. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 21, 128–138.

Turrisi, R., Mallett, K.A., Mastroleo, N.R., Larimer, M.E., 2006. Heavy drinking in college students: who is at risk and what is being done about it? J. Gen. Psychol. 133, 401–420.

University of Virginia, 2019. Codes of Virginia [Online]. Available: http://www.virginia.edu/sacs/references/VaCode/Va.Code23-7.4.pdf (Accessed January 2019). .

Verelst, S., Moonen, P.J., Desruelles, D., Gillet, J.B., 2012. Emergency department visits due to alcohol intoxication: characteristics of patients and impact on the emergency room. Alcohol Alcohol 47, 433–438.

Wechsler, H., Dowdall, G.W., Davenport, A., Castillo, S., 1995. Correlates of college student binge drinking. Am. J. Public Health 85, 921–926.

Wechsler, H., Lee, J.E., Nelson, T.F., Kuo, M., 2002. Underage college students' drinking behavior, access to alcohol, and the influence of deterrence policies. Findings from the Harvard School of Public Health College Alcohol Study. J. Am. Coll. Health 50, 223–236.

White, A.M., Kraus, C.L., Swartzwelder, H.S., 2006. Many college freshmen drinks at levels far beyond the bing threshold. Alcohol. Clin. Exp. Res. 6, 1006–1010.

White, A.M., Slater, M.E., Ng, G., Hingson, R., Breslow, R., 2018. Trends in alcohol-related emergency department visits in the United States: results from the nationwide emergency department sample, 2006 to 2014. Alcohol. Clin. Exp. Res. 42, 352–359.