

# Identifying an accurate self-reported screening tool for alcohol use disorder: evidence from a Swiss, male population-based assessment

Stéphanie Baggio<sup>1,2</sup> , Bastien Trächsel<sup>3</sup>, Valentin Rousson<sup>3</sup>, Stéphane Rothen<sup>4</sup>, Joseph Studer<sup>5</sup> , Simon Marmet<sup>5</sup>, Patrick Heller<sup>1,6</sup>, Frank Sporkert<sup>7</sup>, Jean-Bernard Daeppen<sup>5</sup>, Gerhard Gmel<sup>5,8,9,10</sup> & Katia Iglesias<sup>11</sup>

## ABSTRACT

**Background and aims** Short screenings for alcohol use disorder (AUD) are crucial for public health purposes, but current self-reported measures have several pitfalls and may be unreliable. The main aim of our study was to provide empirical evidence on the psychometric performance of self-reports currently used. Our research questions were: compared with a gold standard clinical interview, how accurate are (1) self-reported AUD, (2) self-reported alcohol use over time and (3) biomarkers of alcohol use among Swiss men? Finally, we aimed to identify an alternative screening tool. **Design** A single-center study with a cross-sectional design and a stratified sample selection. **Setting** Lausanne University Hospital (Switzerland) from October 2017 to June 2018. **Participants** We selected participants from the French-speaking participants of the ongoing Cohort Study on Substance Use and Risk Factors ( $n = 233$ ). The sample included young men aged on average 27.0 years. **Measurements** We used the Diagnostic Interview for Genetic Studies as the gold standard for DSM-5 AUD. The self-reported measures included 11 criteria for AUD, nine alcohol-related consequences, and previous 12 months' alcohol use. We also assessed biomarkers of chronic excessive drinking (ethyl glucuronide and phosphatidylethanol). **Findings** None of the self-reported measures/biomarkers taken alone displayed both sensitivity and specificity close to 100% with respect to the gold standard (e.g. self-reported AUD: sensitivity = 92.3%, specificity = 45.8%). The best model combined eight self-reported criteria of AUD and four alcohol-related consequences. Using a cut-off of three, this screening tool yielded acceptable sensitivity (83.3%) and specificity (78.7%). **Conclusions** Neither self-reported alcohol use disorder nor heavy alcohol use appear to be adequate to screen for alcohol use disorder among young men from the Swiss population. The best screening alternative for alcohol use disorder among young Swiss men appears to be a combination of eight symptoms of alcohol use disorder and four alcohol-related consequences.

**Keywords** Alcohol, community-based sample, epidemiology, machine learning, psychometrics, public health.

Correspondence to: Stéphanie Baggio, Division of Prison Health, Geneva University Hospitals and University of Geneva, Chemin du Petit Bel Air 2, 1226 Thônex, Switzerland. E-mail: stephanie.baggio@hcuge.ch

Submitted 11 April 2019; initial review completed 21 June 2019; final version accepted 14 October 2019

## INTRODUCTION

Alcohol use disorder (AUD) is conceptualized as a complex syndrome, including physiological, behavioral and cognitive symptoms developed after repeated alcohol use [1–3]. The observable symptoms are considered to be valid indicators of the underlying disorder [4]. In the current definition from the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-5) [4], these symptoms are related to three main areas: loss of control over alcohol use and compulsive use, detrimental consequences (i.e. functional impairment and hazardous use) and pharmacological symptoms. A diagnosis of AUD can be made if a patient shows at least two symptoms of the 11 listed in

the DSM-5 diagnostic of AUD; however, AUD is also regarded as a dimensional construct. It ranges from mild to severe AUD, according to the number of symptoms endorsed. Heavy alcohol use is not considered as a symptom in the DSM-5.

Short and reliable screening tools of AUD for population-based assessments are crucial for public health purposes. Indeed, we need to establish prevalence rates at the population level, to develop adequate prevention and treatment plans, to monitor substance use and to optimize substance use policies. Traditionally, a reliable diagnosis requires a rigorous evaluation in a face-to-face exchange with an experienced clinician. However, clinical interviews are costly and time-consuming, so they are not suitable for

population-based assessments. Short quantitative measures of AUD are therefore needed.

### Current gaps in AUD screening

Unfortunately, current self-reported measures based on the DSM criteria have several pitfalls. Although the criteria defined in the current DSM-5 definition addresses several gaps identified in previous versions of the DSM [5,6] and screeners have been successfully re-calibrated according to DSM-5 criteria [7–9], some major issues remain. First, DSM diagnostic criteria may be misinterpreted, especially by young people [10]. Young inexperienced drinkers may misinterpret several symptoms, such as ‘tolerance’, ‘drink longer/larger than intended’ and ‘withdrawal’ [11,12]. In addition, the criteria may not be understood in the same way among age groups [10,13]. A second issue with self-reported AUD screenings is their lack of specificity [14], which causes a large proportion of false positives (i.e. people who screen positive even if they do not have the disorder). Other pitfalls have been reported; for example, depressed individuals tend to over-report AUD symptoms, so in some cases self-reported AUD may be even less reliable [15]. However, comparisons of self-reported AUD with clinical diagnoses used as gold standard are scarce.

### Alternative: doing it by numbers?

Heavy alcohol use has been described as a possible suitable criterion for AUD [14,16], as it is responsible for most of the burden of disease and consequences associated with AUD [16]. However, previous studies report a moderate overlap between use and disorder [16,17], suggesting that addiction reflects more than just the magnitude of use [18–20]. The recommended risk thresholds are defined with two drinks or more per day [21] and, more recently, with 100 g of pure ethanol (10 standard drinks) per week [22]. The risk threshold of heavy use is 40–60 g/day [2]. These thresholds have been tested against health issues, and empirical evidence in relationship with AUD is still needed.

Another way to measure (heavy) alcohol use involves biochemical markers [23]. These measures can be useful to assess chronic excessive drinking. As they do not rely on self-reports or clinical evaluation [24], they do not suffer the typical problems of substance use misreporting [25]. The most commonly used biomarkers are ethyl glucuronide (EtG) and phosphatidylethanol (PEth). These biomarkers are assessed as continuous variables, but guidelines are also available to identify alcohol overuse. Chronic excessive drinking is defined as  $\geq 30$  pg/mg of pure ethanol per day over several months for EtG [24] and  $\geq 210$  ng/ml for PEth [26]. A cut-off of 9 pg/mg for EtG [27] has also been described as a marker of at-risk

alcohol use and a cut-off  $\leq 112$  ng/ml [28] as a marker for abstinence or moderate drinking for PEth.

### Aims of the study

The main aim of our study was to provide empirical evidence on the psychometric performance of self-reports currently used in population-based assessments. More precisely, we sought to provide insights into the current theoretical debate on the relevance of alcohol use as a suitable criterion for AUD. The specific aims were the following: compared to a gold standard clinical interview, how accurate are (1) self-reported AUD, (2) self-reported alcohol use over time and (3) biomarkers of alcohol use? Finally, we aimed to identify an alternative screening tool (aim 4).

## METHODS

### Design and setting

This study was a single-center study with a cross-sectional design and a stratified sample selection. We collected data at the Lausanne University Hospital (Switzerland) from October 2017 to June 2018. The Swiss National Science Foundation funded the study (no. 10001C\_173418/1), which was approved by the Ethic Committee of the Canton of Vaud (no. 2017–00776). The protocol is published elsewhere [29].

### Participants and procedures

We recruited participants from the ongoing Cohort Study on Substance Use and Risk Factors (C-SURF) [30]. This cohort collects data on young Swiss men transitioning into adulthood [30,31]. Participants were initially enrolled in army recruitment centers in 2010 (average age = 20 years), but that setting was for enrollment only; the study itself occurred outside the army setting. Because conscription is mandatory for all young Swiss men, the C-SURF sample is representative for French-speaking young Swiss men.

We proposed participation in this study to all French-speaking participants of C-SURF who had answered to the second follow-up questionnaire (from 2016 to 2018) and who had a valid e-mail address ( $n = 2668$ ). In September 2017, we invited them to complete the 10-question version of the Alcohol Use Disorder Identification Test (AUDIT) [32] on-line. A total of 1371 participants completed the questionnaire within 4 weeks (response rate = 51.4%). Then, participants were selected using a stratified sampling strategy. Participants were selected among two strata: those with a high AUDIT score ( $\geq 13$ ) and those with a low score ( $< 13$ ) [33]. Of the 1371 respondents, all participants in the high-score stratum were included (not enough participants to perform a random selection in that

subset,  $n = 193$ ), and a random selection was used in the low-score stratum ( $n = 137$ ). Finally, three psychologists called the selected participants to invite them to participate in the study. An appointment was scheduled at the Lausanne University Hospital for those who agreed to participate ( $n = 233$ , response rate = 70.6%, similar response rates in both strata: low-score stratum = 68.9%, high-risk stratum = 72.0%). The study flow-chart is presented in Fig. 1.

Young people were eligible for study participation if they: (1) were initially conscripted at the army recruitment center in Lausanne, (2) had a valid e-mail address, (3) completed the AUDIT, (4) were alcohol users and (5) provided informed consent to participate.

During the 90-minute appointment, participants completed a computer-assisted questionnaire, a structured interview to assess AUD with a highly skilled psychologist who received a 1-month training to assess AUD and provided biological samples (hair and blood). Both participants and psychologists were blinded to the participants' strata (low or high AUDIT-scores). All participants received a compensation of CHF 100.00 (approximately US\$101 or €90) for their participation (gift voucher) and for their traveling expenses.

#### Sample size calculation

As described in the study protocol, we used several proportion non-inferiority tests to ensure a sufficient sample size in both groups (with and without AUD) [29]. We set the following parameters: alpha = 5%, power = 80% and margin of equivalence = 10% [34]. We tested sensitivity values between 20 and 80% for self-reported AUD and we used sensitivity = 50% because it was the worst scenario in term of sample size. Based on the theoretical work on heavy alcohol use [14,16] we expected that the sensitivity of heavy alcohol use would be higher, so we set sensitivity of heavy alcohol use = 55%. Finally, we set the correlation between AUD and heavy alcohol use at  $r = 0.35$ . This was in line with previous studies reporting a moderate correlation

between AUD and alcohol use [16,17]; this correlation was more conservative than the one in the C-SURF study ( $r = 0.5$ ). Based on these parameters, we needed 86 participants in each group.

#### Measures

##### Diagnostic of AUD

We assessed the gold standard with the Diagnostic Interview for Genetic Studies (DIGS) [35]. This structured interview enables a comprehensive and reliable assessment of AUD. It has a high inter-rater agreement and a good concordance with clinical diagnoses from medical records [35]. As the DIGS was not yet adapted to the DSM-5 criteria, we replaced the DSM-IV question regarding legal problems (removed in the DSM-5) by a question assessing craving (added into the DSM-5, and defined as having a strong desire or urge to drink). Otherwise, we used all dimensions assessed in the DSM-IV. We assessed AUD (at least mild, cut-off score = 2) for the previous 12 months.

##### Self-reported AUD

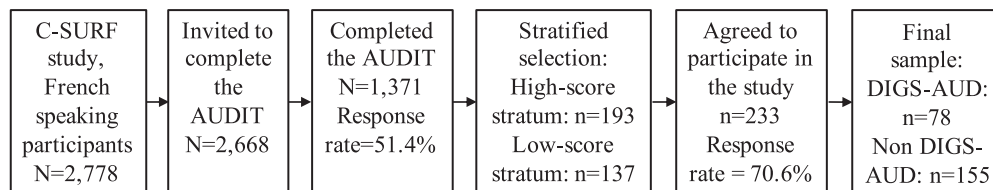
The participants self-reported the presence or absence of the 11 criteria of DSM-5 AUD during the previous 12 months [32,36]. These questions are listed elsewhere [10]. We combined two questions for the second AUD criterion (hazardous alcohol use). All criteria are detailed in Table 1. We used the recommended cut-off of two symptoms or more to define the presence or absence of at least mild AUD.

##### Alcohol-related consequences

We recorded the presence or absence of nine alcohol-related consequences, as detailed in Table 1 [32,37,38].

##### Twelve-month alcohol use

We assessed the previous 12-month alcohol use with an extended quantity–frequency questionnaire [39]. Questions address weekdays and weekends separately. We then



C-SURF: Cohort study on substance use and risk factors; AUDIT: Alcohol Use Disorder Identification Test;

DIGS: Diagnostic interview for genetic studies; AUD: alcohol use disorder.

**Figure 1** Flow diagram of the study.

**Table 1** Descriptive statistics for alcohol-related variables in the combined study sample of young Swiss men ( $n = 233$ ).

Variables		Descriptive statistics
DIGS diagnosis of alcohol use disorder (gold standard) <sup>a</sup>		
No		66.5 (155)
Yes		33.5 (78)
Self-reported alcohol use disorder <sup>a</sup>		
A1	Neglect role because of alcohol use	32.2 (75)
A2	<i>Hazardous use of alcohol</i>	64.4 (150)
A3	Social/interpersonal problems due to alcohol use	11.6 (27)
A4	<i>Tolerance</i>	21.0 (49)
A5	<i>Withdrawal</i>	11.6 (27)
A6	Use larger/longer than intended	66.5 (155)
A7	<i>Unable to control/quit alcohol use</i>	10.3 (24)
A8	<i>Time spent getting/using/recovering alcohol</i>	33.1 (77)
A9	<i>Activities given up because of alcohol use</i>	5.2 (12)
A10	<i>Physical/psychological problems due to alcohol use</i>	12.9 (30)
A11	<i>Craving</i>	8.2 (19)
Self-reported AUD sum score (0–11) <sup>b</sup>		2.77 (2.26)
Self-reported AUD diagnosis ( $\geq 2$ )		67.0 (156)
Self-reported alcohol-related consequences <sup>a</sup>		
C1	I drank alcohol or took drugs or medicine in order to get over any of the bad secondary effects of drinking alcohol	23.6 (55)
C2	I had a mental blackout after drinking alcohol	47.2 (110)
C3	<i>While drinking alcohol, I did something that I badly regretted later</i>	46.4 (108)
C4	<i>I had unplanned sex because I was drunk</i>	24.0 (56)
C5	I had sex without a condom because I was drunk	12.0 (28)
C6	<i>I had an accident or I got injured because I was drunk</i>	19.7 (46)
C7	<i>I came into a conflict with the police or with the authorities more than once because of my alcohol use</i>	4.3 (10)
C8	I came into an argument or into a fight while drinking alcohol or straight after	15.9 (37)
C9	I damaged property, because I was drunk	8.2 (19)
Sum score (0–9) <sup>b</sup>		2.01 (1.88)
Alcohol use (no. of drinks per week) <sup>b</sup>		
Previous 12-month use		13.90 (13.78)
Biomarkers of alcohol use		
Ethylglucuronide (EtG) ( $n = 227$ ) <sup>b</sup>		23.99 (53.76)
Chronic excessive alcohol use according to EtG ( $\geq 30$ np/mg) <sup>a</sup>		23.4 (53)
Phosphatidylethanol (PEth) ( $n = 231$ ) <sup>b</sup>		263.49 (304.61)
Chronic excessive alcohol use according to PEth ( $\geq 210$ ng/ml) <sup>a</sup>		42.9 (99)

DIGS: Diagnostic interview for genetic studies; AUD: alcohol use disorder. Items in italic type were included in the 12-item measure selected by machine learning. <sup>a</sup>Percentages and  $n$  under brackets are reported; <sup>b</sup>means and standard deviations under brackets are reported.

converted the resulting measures into a number of drinks per week. One standard drink contains 10 g of pure alcohol.

#### Biomarkers of chronic excessive drinking

We collected a lock of hair to measure EtG and capillary blood to measure PEth. We extracted EtG using micropulverized extraction and SPE purification. The extracts were submitted to LC–MS/MS analysis (6500 Qtrap Sciex, Framingham, MA, USA). Chromatographic separation was achieved using a Merck Chromolith<sup>®</sup>, Darmstadt, Germany RP-C18e column (150 × 3 mm). We determined the PEth homolog 16:0/18:1 concentration using methanolic extraction, followed by liquid chromatography with tandem mass spectrometry (LC–MS/MS) measurement (5000 Qtrap Sciex) [40].

#### Analytical strategy

We first computed descriptive statistics (either percentages or mean scores) for all variables, as well as Spearman's correlations between the various alcohol-related measures.

#### Receiver operating characteristic curves (ROC)

Next, we used ROC curves to test whether seven measures were good indicators of the DIGS-AUD gold standard: (1) self-reported AUD, (2) 12-month alcohol use, (3) EtG, (4) PEth, (5) self-reported AUD and alcohol-related consequences (20 questions) and additionally, (6) alcohol-related consequences and (7) AUDIT. We derived the area under the curve (AUC), sensitivity, specificity and percentage of correctly classified participants for each



analysis/cut-off. The best cut-off was selected using the Youden's *J* statistic: maximum (sensitivity + specificity – 1) [41]. For sensitivity and specificity, values close to 100% are ideal [42].

#### *Machine learning*

Following the most promising results (the fifth model with 20 items), we used a machine-learning procedure to select the best model maximizing performance and parsimony. We considered the 1 048 575 possibilities (i.e. all subsets including between one and 20 items). We selected the best model using the Akaike information criterion (AIC), as this method prevented overfitting to some extent by penalizing the log-likelihood of the model using twice the number of parameters included in the model. However, AIC could be drastically reduced if all the predictors in the model are forced to have exactly the same weight, which is equivalent to considering a single sum score (i.e. the number of 'yes' replies within a given subset of items). We then computed a ROC curve to select the best model. As the selection of items was based on the same data, these results might be too optimistic. To correct for this optimism, we used a bootstrap procedure [43]. We considered 1000 bootstrap resamples from the original data set, in each case selecting the best sum of the 1 048 575 possible subsets of items using the AIC criterion, then calculating the corresponding AUC for the bootstrap sample and in the original sample. The difference was taken as an estimate of the optimism, which was deduced from the initial AUC estimate.

#### *Non-response bias*

We also investigated non-response bias, as non-respondents may differ on alcohol use. We compared respondents and non-respondents on alcohol-related variables, using data from the second follow-up questionnaire of C-SURF. Overall, the relative bias between respondents and non-respondents was small (see Supporting information, Table S1).

#### *Additional/sensitivity analyses*

We conducted several additional and sensitivity analyses. These included: ROC curves combining self-reported AUD symptoms and binary variables of alcohol use (using various theoretical and empirically derived cut-offs), as well as different definitions of alcohol use during periods of various length (also using previous waves of the C-SURF study). As we did not identify a screener with acceptable psychometric properties, these analyses are not presented in this paper. We also ran sensitivity analyses with a cut-off of 3 on the DIGS-AUD, which yielded similar results. As the prevalence rate of AUD and polysubstance use (use of any illicit substance other than cannabis) were over-represented in the sample, we also conducted

sensitivity analyses using random subsets of the sample to test whether the final screening tool performed equally in a sample closer to a general population sample. The results from the subsets and the full sample were similar; they are available upon request.

The EtG measure had various time-frames according to the length of the hair collected; these time-frames did not correspond to the 12-month period used in the DIGS-AUD. As a sensitivity analysis, we computed the ROC curve for participants with hair segments  $\geq 6$  cm (i.e. segments that provided information on the previous 6-month alcohol use,  $n = 45$ ). The results were similar to those of the whole sample (including  $n = 182$  participants with hair segments  $< 6$  cm, data available upon request).

## **RESULTS**

### **Preliminary analyses**

Participants were aged, on average,  $27.00 \pm 1.44$  years. Descriptive statistics are reported in Table 1. Overall, 33.5% of the participants were diagnosed with an AUD according to the DIGS-AUD and 67.0% according to self-reported AUD. Participants reported drinking an average of 13.90 drinks per week (range = 0.35–91). A total of 23.4% were classified as having a chronic excessive alcohol use using EtG and 42.9% using PEth. Correlations between alcohol-related variables are reported in Supporting information, Table S2.

### **ROC curves for self-reported measures and biomarkers (aims 1, 2 and 3)**

The performance of the various self-reported measures of alcohol is reported in Table 2. All corresponding ROC curves are reported in Supporting information, Figs S1–S4. The first panel reports the performance of self-reported AUD (aim 1). For the recommended cut-off of two or more symptoms, the sensitivity was 92.3% and the specificity was 45.8%. Taken separately, most symptoms had an acceptable specificity ( $\geq 80\%$ ) and a low sensitivity ( $\leq 59\%$ ), except the two most common symptoms which had the reverse pattern (sensitivity = 82%, specificity  $\leq 45\%$ ; for more detail, see the first panel of Supporting information, Table S3).

The second and third panels of Table 2 report the performances of previous 12-month alcohol use (aim 2). The cut-off with the highest percentage of correctly classified participants was 21 drinks per week or more (sensitivity = 38.5%, specificity = 93.6%). The highest sensitivity/specificity was achieved with the cut-off of 10.5 drinks per week or more (sensitivity = 83.3%, specificity = 60.0%).

The fourth and fifth panels of Table 2 report the performances of biomarkers of chronic excessive drinking (aim 3). The recommended cut-offs (EtG: 30 pg/mg and

**Table 2** Performance of screening questionnaires compared to diagnosis of alcohol use disorder (gold standard).

Measure	Criterion	Sensitivity	Specificity	Correctly classified
Self-reported alcohol use disorder AUC = 0.844	Everybody	100	0.00	33.48
	≥ 1	100	20.65	47.21
	≥ 2	92.31	45.81	61.37
	≥ 3	84.62	70.32	75.11
	≥ 4	60.26	87.10	78.11
	≥ 5	46.15	94.84	78.54
	≥ 6	32.05	96.77	75.11
	≥ 7	17.95	100	75.53
	≥ 8	11.54	100	70.39
	≥ 9	6.41	100	68.67
	≥ 10	5.13	100	67.81
	Nobody	3.85	100	66.52
Previous 12-month alcohol use (no. of drinks per week) AUC = 0.785	Everybody	100	0.00	33.48
	≥ 7	91.03	44.52	60.09
	≥ 10	85.90	55.48	65.67
	≥ 14	60.26	69.03	66.09
	≥ 21	38.46	93.55	75.11
	≥ 27	28.21	97.42	74.25
	≥ 34	17.95	100	70.82
	≥ 44	12.82	100	67.38
	≥ 60	5.13	100	68.24
	Nobody	2.56	100	66.52
Ethyl glucuronide (EtG) AUC = 0.700	Everybody	100	0.00	33.48
	≥ 10	73.68	60.93	65.20
	≥ 19	53.95	76.16	68.72
	≥ 20	51.32	76.82	68.28
	≥ 30	34.21	82.12	66.08
	> 60	10.53	93.38	65.64
	≥ 100	7.89	96.69	66.96
	Nobody	0.00	98.01	65.20
Phosphatidylethanol (PEth) AUC = 0.617	Everybody	100	0.00	33.33
	> 210	53.25	61.69	58.57
	≥ 745	9.09	95.45	66.67
	Nobody	0.00	99.35	66.23

AUC = area under the receiver operating characteristic curve.

9 pg/mg, PEth: 210 ng/ml) all displayed poor performance. The cut-offs with the highest sensitivity/specificity were, respectively, 12 pg/ng or more for EtG (sensitivity = 71.1%, specificity = 66.9%) and 90 ng/ml for PEth (sensitivity = 83.1%, specificity = 39.6%). For EtG, this corresponded to more than an at-risk alcohol use (defined as EtG = 9 pg/ng) and to a moderate alcohol use for PEth (Peth ≤ 112 ng/ml).

Alcohol-related consequences alone performed better than the other self-reported measures and biomarkers, with a cut-off of 3, sensitivity = 64.1% and specificity = 76.8% (see Supporting information, Fig. S5). We also tested the psychometric properties of the AUDIT, which was used to select participants (see Supporting information, Fig. S6). The recommended cut-off of 13 yielded an acceptable sensitivity (94.4%) but a low specificity

(52.9%). The best cut-off was ≥ 15 (sensitivity = 69.2%, specificity = 75.5%).

#### Machine learning (aim 4)

The best alternative model combined the 11 symptoms of self-reported AUD and the nine alcohol-related consequences [area under the ROC curve (AUC) = 0.857]. These results are reported in Table 3 and Supporting information, Fig. S7. The best cut-off was ≥ 6 (sensitivity = 76.9%, specificity = 81.9%).

We tested how and how well it would be possible to predict the diagnosis of AUD from a subset of the 20 symptoms/consequences of this model. The best model included eight symptoms and four consequences (AUC = 0.872, AIC = 194.8); the selected symptoms/consequences are highlighted in italic in Table 1. A summary of the associations of the items with the DIGS-AUD and of the best models are reported in Supporting information, Tables S3 and S4. We then derived the performance of these 12 symptoms/consequences, as reported in Table 4 and Supporting information, Fig. S8. The cut-off that maximized sensitivity and specificity was ≥ 3 (sensitivity = 83.3%, specificity = 78.7%).

On average, the AUC in the bootstrap sample was 0.031 above that of the original sample. This quantity was taken as an estimate of our optimism, which was then deduced from our initial AUC of 0.872. The corrected estimate of AUC was 0.841.

**Table 3** Performance of self-reported alcohol use disorder and alcohol-related consequences (20 items) compared to diagnosis of alcohol use disorder (gold standard).

Cut-off	Sensitivity	Specificity	Correctly classified
Everybody	100	0.00	33.48
≥ 1	100	16.13	44.21
≥ 2	97.44	31.61	53.65
≥ 3	97.44	43.87	61.80
≥ 4	89.74	57.42	68.24
≥ 5	83.33	72.26	75.97
≥ 6	<b>76.92</b>	<b>81.94</b>	<b>80.26</b>
≥ 7	62.82	87.10	78.97
≥ 8	44.87	91.61	75.97
≥ 9	37.18	97.42	77.25
≥ 10	26.92	98.71	74.68
≥ 11	21.79	98.71	72.96
≥ 12	15.38	99.35	71.24
≥ 13	11.54	100	70.39
≥ 14	7.69	100	69.10
≥ 15	2.56	100	67.38
≥ 16	1.28	100	66.95
Nobody	0.00	100	66.52

The preferred model is highlighted in bold type.

**Table 4** Performance of the 12 items of self-reported alcohol use disorder and alcohol-related consequences included in the best-fitting model compared to diagnosis of alcohol use disorder (gold standard).

<i>Cut-off</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>Correctly classified</i>
Everybody	100	0.00	33.48
≥ 1	100	27.10	51.50
≥ 2	93.59	48.39	63.52
<b>≥ 3</b>	<b>83.33</b>	<b>78.71</b>	<b>80.26</b>
≥ 4	73.08	87.74	82.83
≥ 5	43.59	94.84	77.68
≥ 6	25.64	98.06	73.82
≥ 7	14.10	100	71.24
≥ 8	8.97	100	69.53
≥ 9	5.13	100	68.24
≥ 10	2.56	100	67.38
≥ 11	1.28	100	66.95
Nobody	0.00	100	66.52

The preferred model is highlighted in bold type.

## DISCUSSION

This study aimed to provide empirical evidence on the psychometric performance of self-reported AUD and to identify an alternative screening tool among young men.

### Aim 1: self-reported AUD

The study first showed that the 11 self-reported criteria of AUD did not correspond well to the gold standard when using the recommended threshold. Indeed, using this cut-off of two symptoms or more, the sensitivity was acceptable (92.3%), but there were 54.2% of false positives. We obtained a higher specificity with higher cut-offs, but the percentage of false negatives was unacceptable. Therefore, self-reported AUD did not appear as a valid screening tool for DIGS-AUD. Taken separately, no symptom performed really badly, meaning that the participants did not clearly misunderstand any of them. The commonly misunderstood symptoms (e.g. tolerance, withdrawal and symptoms related to loss of control [11,12]) did not have particularly low psychometric performance and were not low-threshold criteria.

Overall, this result suggested that self-reported AUD should not be used for a binary diagnosis, as participants were likely to be misclassified as false positives (or false negatives if a higher threshold is used). Using a sum-score of symptoms would provide a more accurate picture of AUD in population-based assessments, if analyzed using appropriate count models [44]. This approach is in line with studies showing that AUD and substance use disorders in general are dimensional constructs [45,46]. However, this approach does not address the question of differences in

severity for each symptom [47], nor the need of a valid binary classification that can be used to establish prevalence rates at the population level.

### Aims 2 and 3: alcohol use over time

Unfortunately, the proposition of using the number of drinks [14,16] was not a reliable alternative. Among the measures of alcohol use, neither self-reports nor biomarkers displayed acceptable psychometric performance. Heavy use was associated with an acceptable specificity, but it also had a high percentage of false negatives; therefore, it seemed that, in our sample of young men, heavy alcohol use did not capture AUD, despite previous suggestions [14,16]. This result was in line with previous findings on the moderate overlap between alcohol and AUD or substance use and substance use disorder [16,17]. As reported in previous studies, the concept of addictive disorder as measured in the DIGS did not appear as a magnitude of use [18–20].

Also, these findings are not really surprising, as alcohol use is not included in the current DSM-5 definition of AUD, which instead focuses on loss of control over alcohol use, compulsive use, detrimental consequences and pharmacological symptoms [4]. In addition, our sample was limited to young men, who are not representative of the whole population of drinkers. They are more likely to drink for social reasons and they may be transitioning away from their period of heaviest drinking. Heavy alcohol use might be a reasonably good indicator in other groups (such as women and older populations).

### Aim 4: alternative screening tool

Beyond these findings, we identified a suitable alternative: adding the nine alcohol-related consequences to the 11 criteria of self-reported AUD. A cut-off of 6 or more yielded the highest performance (sensitivity = 76.9%, specificity = 81.9%) of all self-reported measures. These results were in line with previous findings suggesting that diagnostic validity improves when harm indicators (e.g. health problems and loss of important activities) are included [48].

However, 20 symptoms/consequences are somewhat lengthy for a population-based assessment and a short screening, so we used machine learning to select the optimal model, combining good psychometric performance and parsimony. The resulting final model with 12 symptoms/consequences and a cut-off of 3 displayed the best sensitivity (83.3%, false negatives = 16.7%) and specificity (78.7%, false positives = 21.3%) of all models tested. Additionally, it is possible to use alternative cut-offs to maximize sensitivity or specificity, as desired. For example, to screen for AUD in clinical settings (e.g. to identify at-risk

patients in order to propose them a complete diagnostic interview), a cut-off of two or more would yield a sensitivity of 93.6%, which corresponds to a small percentage of false negatives (6.4%). On the contrary, to maximize specificity (in other words, to identify people who do not have an AUD), a cut-off of five or more would be an optimal estimation of AUD in population-based assessment conducted on a similar population. In our analyses, sensitivity and specificity had an equal weight, but cut-offs could be modified depending on which error (false positive or false negative) is worse in a given situation.

In the final model, we included eight symptoms of self-reported AUD and four alcohol-related consequences (see the English and French versions in Supporting information, Tables S5 and S6). These items combined several aspects of AUD: loss of control (i.e. inability to reduce or stop drinking, excessive time drinking or recovering from its after-effects and craving), pharmacological effects (i.e. tolerance and withdrawal) and detrimental consequences [i.e. being engaged in dangerous situations, having (mental) health problems, giving up other activities, doing things badly regretted afterwards, accident/injury, unplanned sex and conflict with authorities]. Consequences related to alcohol use played an important role in the screening tool. Again, these findings were in line with the conclusions of a previous study reporting that both dysfunction and harm components are needed to gain validity [48]. Some of these consequences might be related to the specific characteristics of our sample (young Swiss men), particularly unplanned sex, accidents or injuries and conflict with authorities. A previous narrative review reported that substance-related consequences introduce bias and degrade the validity of the diagnoses and, thus, that negative consequences should be avoided in diagnostic systems [49]. Our findings contrasted with this recommendation, with consequences being important to achieve good psychometric performance in the final screening tool. It should be kept in mind that the actual definition of AUD, as assessed in our clinical interview, encompasses detrimental consequences (on social functioning, health and related to hazardous use). Therefore, it was not surprising that the screening tool also included consequences, which are parts of the actual definition of AUD. The DSM-5 tried to achieve a better understanding of substance use disorders among young adults [50] and our findings suggested that including consequences helps to achieve this objective. Focusing on the consequences associated with alcohol use and AUD is a perspective of harm reduction, which is meant to minimize the health risks and detrimental effects of substance use [51].

Three symptoms of the original 11-item AUD screening tool were not included in the final screening; namely, role neglect, social/interpersonal problems and use larger/longer than intended. The last item was endorsed

by a large majority of participants (66.5%), so it might be insensitive for this population likely to binge drink [52]. Social roles and relationships might be not yet affected among young adults, because they did yet not have all the responsibilities of adults [53].

### Limitations

This study had some shortcomings. First, our study needs to be replicated. Indeed, young men constitute a specific group of alcohol drinkers, for which heavy alcohol use may not indicate a disorder. In our large initial sample, only 14% were classified as having an AUD according to the AUDIT. Further studies should focus on women and older groups, including clinical populations. Furthermore, we used a sample of young Swiss men who were already enrolled into a cohort study. This cohort could have differed from the whole population, as they had participated in this survey for years. A second limitation was related to the different time-frames of the diagnostic of AUD (previous 12 months) and biomarkers. The EtG values depended on the length of the participants' hair, and provided information for various time-periods. This is important, because shorter recall periods do not always reflect typical drinking patterns [39]. However, when we restricted the analyses to participants with hair segments of 6 cm or more, the results were similar to those of the whole for the EtG in our study. Another limitation was that we did not reach the desired sample size ( $n = 86$ ) for the group with AUD ( $n = 78$ ). However, the desired sample size was based on a moderate correlation between AUD and alcohol use ( $r = 0.35$ ) and in our sample, the correlation was higher ( $r = 0.417$ ). Therefore, this small deviation from the planned sample size was not an important issue. Another issue was related to the sample size calculation. There are no (statistical or clinical) standards for defining the margin of equivalence. A previous study in another field reported that it was difficult to obtain significant results with a margin of less than 10% [34], so we used this value in the study. Our findings might be used as a starting-points for defining the margin of equivalence in other studies. The last shortcoming was that, due to the lack of an objective gold standard, it is difficult to validate the diagnostic instruments [35]. However, the DIGS has been described as a valid assessment of AUD [35] and it allowed a standardized assessment according to the current definition of AUD.

### CONCLUSION

To conclude, neither self-reported DSM-5 criteria nor heavy alcohol use constituted valid screening tools for AUD. The best alternative was to combine eight symptoms of AUD and four alcohol-related consequences. These findings provided important insights in some current debates



in the addiction field: AUD was more than just heavy alcohol use, and consequences were important to define a reliable assessment. The final screening tool should be tested in others populations to achieve a robust empirical evidence regarding its psychometric performance and accuracy for population-based assessments. However, alcohol use should not be neglected, as it is responsible of a large burden of disease and detrimental health consequences.

#### Declaration of interests

None.

#### Acknowledgements

This study was supported by Swiss National Research Foundation (no. 10001C\_173418/1).

#### Author's affiliations

Division of Prison Health, Geneva University Hospitals and University of Geneva, Thônex, Switzerland,<sup>1</sup> Department of Forensic Psychiatry, Institute of Forensic Medicine, University of Bern, Bern, Switzerland,<sup>2</sup> Division of Biostatistics, Center for Primary Care and Public Health (Unisanté), University of Lausanne, Lausanne, Switzerland,<sup>3</sup> Addiction Division, Geneva University Hospitals and University of Geneva, Geneva, Switzerland,<sup>4</sup> Addiction Medicine, Department of Psychiatry, Lausanne University Hospital, Lausanne, Switzerland,<sup>5</sup> Adult Psychiatry Division, Department of Mental Health and Psychiatry, Geneva University Hospitals, Thônex, Switzerland,<sup>6</sup> Centre of Legal Medicine, Forensic Toxicology and Chemistry Unit, Lausanne and Geneva Universities, Lausanne, Switzerland,<sup>7</sup> Addiction Switzerland, Lausanne, Switzerland,<sup>8</sup> Centre for Addiction and Mental Health, Toronto, ON, Canada,<sup>9</sup> University of the West of England, Bristol, UK<sup>10</sup> and School of Health Sciences, HES-SO University of Applied Sciences and Arts of Western Switzerland, Fribourg, Switzerland<sup>11</sup>

#### References

1. American Psychiatric Association (APA) *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. Washington, DC: APA; 2000.
2. Rehm J. How should prevalence of alcohol use disorders be assessed globally? *Int J Methods Psychiatr Res* 2016; **25**: 79–85.
3. World Health Organization (WHO) *ICD-10 Classification of Mental and Behavioral Disorder: Clinical Descriptions and Diagnostics Guidelines*. Geneva: Switzerland: WHO; 1992.
4. American Psychiatric Publishing *Diagnostic and Statistical Manual of Mental Disorders*, 5th edn. Arlington, VA: American Psychiatric Publishing; 2013.
5. Baggio S., Studer J., Dupuis M., Gerhard G. Subthreshold problem drinkers in DSM-5 alcohol use disorder classification. *Am J Addict* 2016; **25**: 408–15.
6. Hasin D. S., O'Brien C. P., Auriacombe M., Borges G., Bucholz K., Budney A., et al. DSM-5 criteria for substance use disorders: recommendations and rationale. *Am J Psychiatry* 2013; **170**: 834–51.
7. Lundin A., Hallgren M., Forsman M., Forsell Y. Comparison of DSM-5 classifications of alcohol use disorders with those of DSM-IV, DSM-III-R, and ICD-10 in a general population sample in Sweden. *J Stud Alcohol Drugs* 2015; **76**: 773–80.
8. Bartoli F., Carrà G., Crocamo C., Clerici M. From DSM-IV to DSM-5 alcohol use disorder: an overview of epidemiological data. *Addict Behav* 2015; **41**: 46–50.
9. Mohler-Kuo M., Foster S., Gmel G., Dey M., Dermota P. DSM-IV and DSM-5 alcohol use disorder among young Swiss men. *Addiction* 2015; **110**: 429–40.
10. Marmet S., Studer J., Bertholet N., Grazioli V. S., Daeppen J.-B., Gmel G. Interpretation of DSM-5 alcohol use disorder criteria in self-report surveys may change with age. A longitudinal analysis of young Swiss men. *Addict Res Theory* 2019; **0**: 1–9.
11. Karriker-Jaffe K. J., Witbrodt J., Greenfield T. K. Refining measures of alcohol problems for general population surveys. *Alcohol Clin Exp Res* 2015; **39**(2): 363–70.
12. Slade T., Teesson M., Mewton L., Memedovic S., Krueger R. F. Do young adults interpret the DSM diagnostic criteria for alcohol use disorders as intended? A cognitive interviewing study. *Alcohol Clin Exp Res* 2013; **37**: 1001–7.
13. Lane S. P., Steinley D., Sher K. J. Meta-analysis of DSM alcohol use disorder criteria severities: structural consistency is only 'skin deep'. *Psychol Med* 2016; **46**: 1769–84.
14. Nutt D. J., Rehm J. Doing it by numbers: a simple approach to reducing the harms of alcohol. *J Psychopharmacol (Oxf)* 2014; **28**: 3–7.
15. Baggio S., Iglesias K., Studer J., Dupuis M., Daeppen J.-B., Gmel G. Is the relationship between major depressive disorder and self-reported alcohol use disorder an artificial one? *Alcohol Alcohol* 2015; **50**: 195–9.
16. Rehm J., Marmet S., Anderson P., Gual A., Kraus L., Nutt D. J., et al. Defining substance use disorders: do we really need more than heavy use? *Alcohol Alcohol* 2013; **48**: 633–40.
17. Saha T. D., Stinson F. S., Grant B. F. The role of alcohol consumption in future classifications of alcohol use disorders. *Drug Alcohol Depend* 2007; **89**: 82–92.
18. Baggio S., Dupuis M., Studer J., Spilka S., Daeppen J.-B., Simon O., et al. Reframing video gaming and internet use addiction: empirical cross-national comparison of heavy use over time and addiction scales among young users. *Addiction* 2016; **111**: 513–22.
19. Baggio S., N'Goran A. A., Deline S., Studer J., Dupuis M., Henchoz Y., et al. Patterns of cannabis use and prospective associations with health issues among young males. *Addiction* 2014; **109**: 937–45.
20. Király O., Tóth D., Urbán R., Demetrovics Z., Maraz A. Intense video gaming is not essentially problematic. *Psychol Addict Behav* 2017; **31**: 807–17.
21. Department of Health *Alcohol guidelines review - Report from the Guidelines development group to the UK chief medical officers*. London, UK: Department of Health; 2016.
22. Wood A. M., Kaptoge S., Butterworth A. S., Willeit P., Warnakula S., Bolton T., et al. Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599 912 current drinkers in 83 prospective studies. *Lancet* 2018; **391**: 1513–23.
23. Cabarcos P., Álvarez I., Tabernero M. J., Bermejo A. M. Determination of direct alcohol markers: a review. *Anal Bioanal Chem* 2015; **407**: 4907–25.
24. Society of Hair Testing *Consensus for the Use of Alcohol Markers in Hair for Assessment of Both Abstinence and Chronic Excessive Alcohol Consumption*. Brisbane: Australia: Society of Hair Testing; 2016.
25. Japuntich S. J., Arditte Hall K. A., Joos C. M., Rasmusson A. M., Pineles S. L. Methods to reduce false reporting of substance abstinence in clinical research. *Int J Methods Psychiatr Res* 2018; **27**: e1603.
26. Hahn J. A., Anton R. F., Javors M. A. The formation, elimination, interpretation and future research needs of

- phosphatidylethanol (PEth) for research studies and clinical practice. *Alcohol Clin Exp Res* 2016; **40**: 2292–5.
27. Kharbouche H., Faouzi M., Sanchez N., Daeppen J. B., Augsburg M., Mangin P., *et al.* Diagnostic performance of ethyl glucuronide in hair for the investigation of alcohol drinking behavior: a comparison with traditional biomarkers. *Int J Leg Med* 2012; **126**: 243–50.
  28. Schröck A., Wurst F. M., Thon N., Weinmann W. Assessing phosphatidylethanol (PEth) levels reflecting different drinking habits in comparison to the alcohol use disorders identification test–C (AUDIT-C). *Drug Alcohol Depend* 2017; **178**: 80–6.
  29. Iglesias K., Sporkert F., Daeppen J.-B., Gmel G., Baggio S. Comparison of self-reported measures of alcohol-related dependence among young Swiss men: a study protocol for a cross-sectional controlled sample. *BMJ Open* 2018; **8**: e023632.
  30. Gmel G., Akre C., Astudillo M., Bähler C., Baggio S., Bertholet N., *et al.* The Swiss cohort study on substance use risk factors—findings of two waves. *Sucht* 2015; **61**: 251–62.
  31. Studer J., Baggio S., Mohler-Kuo M., Dermota P., Gaume J., Bertholet N., *et al.* Examining non-response bias in substance use research—are late respondents proxies for non-respondents? *Drug Alcohol Depend* 2013; **132**: 316–23.
  32. Knight J. R., Wechsler H., Kuo M., Seibring M., Weitzman E. R., Schuckit M. A. Alcohol abuse and dependence among U. S. college students. *J Stud Alcohol* 2002; **63**: 263–70.
  33. Meneses-Gaya C., Zuairi A. W., Loureiro S. R., Hallak J. E. C., Trzesniak C., de Azevedo M. J. M., *et al.* Is the full version of the AUDIT really necessary? Study of the validity and internal construct of its abbreviated versions. *Alcohol Clin Exp Res* 2010; **34**: 1417–24.
  34. Salazar A. J., Romero J. A., Bernal O. A., Moreno A. P., Velasco S. C., Díaz X. A. Noninferiority and equivalence evaluation of clinical performance among computed radiography, film, and digitized film for telemammography services. *Int J Telemed Appl* 2016; **2016**: 3642960.
  35. Berney A., Preisig M., Matthey M.-L., Ferrero F., Fenton B. T. Diagnostic interview for genetic studies (DIGS): inter-rater and test-retest reliability of alcohol and drug diagnoses. *Drug Alcohol Depend* 2002; **65**: 149–58.
  36. Grant B. E., Dawson D. A., Stinson F. S., Chou P. S., Kay W., Pickering R. The alcohol use disorder and associated disabilities interview schedule-IV (AUDADIS-IV): reliability of alcohol consumption, tobacco use, family history of depression and psychiatric diagnostic modules in a general population sample. *Drug Alcohol Depend* 2003; **71**: 7–16.
  37. Dupuis M., Baggio S., Henchoz Y., Deline S., N'Goran A., Studer J., *et al.* Risky single occasion drinking frequency and alcohol-related consequences: can abstinence during early adulthood lead to alcohol problems? *Swiss Med Wkly* 2014; **144**: w14017.
  38. Wechsler H., Davenport A., Dowdall G., Moeykens B., Castillo S. Health and behavioral consequences of binge drinking in college. A national survey of students at 140 campuses. *JAMA* 1994; **272**: 1672–7.
  39. Gmel G., Studer J., Deline S., Baggio S., N'Goran A., Mohler-Kuo M., *et al.* More is not always better—comparison of three instruments measuring volume of drinking in a sample of young men and their association with consequences. *J Stud Alcohol Drugs* 2014; **75**: 880–8.
  40. Déglon J., Lauer E., Sporkert F., Valois N., Favrat B., Thomas A., *et al.* Analyse du phosphatidylethanol Sur micro-prélèvement de sang séché: nouvel outil pour le suivi de la consommation d'alcool. Phosphatidylethanol analysis for micro-sampling dried blood: A new tool for monitoring alcohol use *Toxicol Anal Clin* 2018; **30**: S42–S43.
  41. Youden W. J. Index for rating diagnostic tests. *Cancer* 1950; **3**: 32–5.
  42. van Stralen K. J., Stel V. S., Reitsma J. B., Dekker F. W., Zoccali C., Jager K. J. Diagnostic methods I: sensitivity, specificity, and other measures of accuracy. *Kidney Int* 2009; **75**: 1257–63.
  43. Steyerberg E. W., Harrell F. E., Borsboom G. J., Eijkemans M. J., Vergouwe Y., Habbema J. D. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001; **54**: 774–81.
  44. Baggio S., Iglesias K., Rousson V. Modeling count data in the addiction field: some simple recommendations. *Int J Methods Psychiatr Res* 2018; **27**.
  45. Kerridge B. T., Saha T. D., Gmel G., Rehm J. Taxometric analysis of DSM-IV and DSM-5 alcohol use disorders. *Drug Alcohol Depend* 2013; **129**: 60–9.
  46. Liu R. T. Substance use disorders in adolescence exist along continua: Taxometric evidence in an epidemiological sample. *J Abnorm Child Psychol* 2017; **45**: 1577–86.
  47. Stevens J. E., Steinley D., McDowell Y. E., Boness C. L., Trull T. J., Martin C. S., *et al.* Toward more efficient diagnostic criteria sets and rules: the use of optimization approaches in addiction science. *Addict Behav* 2019; **94**: 57–64.
  48. Wakefield J. C., Schmitz M. F. The harmful dysfunction model of alcohol use disorder: revised criteria to improve the validity of diagnosis and prevalence estimates. *Addiction* 2015; **110**: 931–42.
  49. Martin C. S., Langenbucher J. W., Chung T., Sher K. J. Truth or consequences in the diagnosis of substance use disorders. *Addiction* 2014; **109**: 1773–8.
  50. Teesson M., Baillie A. J., Slade T. 'Truth or consequences'—what about truth or dare. *Addiction* 2014; **109**: 1782–4.
  51. Charlet K., Heinz A. Harm reduction—a systematic review on effects of alcohol reduction on physical and mental symptoms. *Addict Biol* 2017; **22**: 1119–59.
  52. Kuntsche E., Gmel G. Alcohol consumption in late adolescence and early adulthood—where is the problem? *Swiss Med Wkly* 2013; **143**: w13826.
  53. Atak H., Cok F. The Turkish version of inventory of the dimensions of emerging adulthood. *Int J Hum Soc Sci* 2008; **3**: 451–7.

## Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Figure S1.** ROC curve for self-reported alcohol use disorder against the diagnosis of alcohol use disorder (gold standard)

**Fig S2.** ROC curve for previous twelve-month alcohol use against the diagnosis of alcohol use disorder (gold standard)

**Fig S3.** ROC curve for Ethyl glucuronide against the diagnosis of alcohol use disorder (gold standard)

**Fig S4.** ROC curve for Phosphatidylethanol against the diagnosis of alcohol use disorder (gold standard)

**Fig S5.** ROC curve for self-reported alcohol-related consequences against the diagnosis of alcohol use disorder (gold standard)

**Fig S6.** ROC curve for self-reported AUDIT against the diagnosis of alcohol use disorder (gold standard)

**Fig S7.** ROC curve for self-reported alcohol use disorder and alcohol-related consequences (20 items) against the diagnosis of alcohol use disorder (gold standard)

**Fig S8.** ROC curve for the best alternative model (12 symptoms/consequences) against the diagnosis of alcohol use disorder (gold standard)

**Table S1.** Comparisons between respondents and non-

respondents on alcohol-related variables

**Table S2.** Spearman correlations between diagnosis of AUD (gold standard), self-reported measures of alcohol use, and biomarkers

**Table S3.** Summary of the associations between the 20 items and the diagnosis of AUD (gold standard)

**Table S4.** Summary of the best models including a subset of items of self-reported alcohol use disorder and alcohol-related consequences

**Table S5.** Final screening tool in English

**Table S6.** Final screening tool in French.