

Serveur Académique Lausannois SERVAL [serval.unil.ch](http://serval.unil.ch)

## Author Manuscript

Faculty of Biology and Medicine Publication

This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Published in final edited form as:

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

**Authors:** Baggio S, Trächsel B, Rousson V, Rothen S, Studer J, Marmet S, Heller P, Sporkert F, Daepfen JB, Gmel G, Iglesias K

**Journal:** Addiction (Abingdon, England)

**Year:** 2020 Mar

**Issue:** 115

**Volume:** 3

**Pages:** 426-436

**DOI:** [10.1111/add.14864](https://doi.org/10.1111/add.14864)

In the absence of a copyright statement, users should assume that standard copyright protection applies, unless the article contains an explicit statement to the contrary. In case of doubt, contact the journal publisher to verify the copyright status of an article.

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>

<sup>1</sup> Division of Prison Health, Geneva University Hospitals and University of Geneva, Chemin du Petit Bel Air 2, 1226 Thônex, Switzerland

<sup>2</sup> Department of Forensic Psychiatry, Institute of Forensic Medicine, University of Bern, Falkenplatz 16, 3012 Bern, Switzerland

<sup>3</sup> Division of Biostatistics, Center for Primary Care and Public Health (Unisanté), University of Lausanne, Biopôle 2, Route de la Corniche 10, 1010 Lausanne, Switzerland

<sup>4</sup> Addiction Division, Geneva University Hospitals and University of Geneva, rue Grand-Pré 70C, 1202 Geneva, Switzerland

<sup>5</sup> Addiction Medicine, Department of Psychiatry, Lausanne University Hospital, Rue du Bugnon 23, 1011 Lausanne, Switzerland

<sup>6</sup> Adult Psychiatry Division, Department of Mental Health and Psychiatry, Geneva University Hospitals, Chemin du Petit Bel Air 2, 1226 Thônex, Switzerland

<sup>7</sup> Centre of Legal Medicine, Forensic Toxicology and Chemistry Unit, Lausanne and Geneva Universities, Chemin de la Vulliette 4, 1000 Lausanne, Switzerland

<sup>8</sup> Addiction Switzerland, Avenue Louis-Ruchonnet 14, 1003 Lausanne, Switzerland

<sup>9</sup> Centre for Addiction and Mental Health, 1001 Queen Street West, Toronto ON – M6J 1H4, Canada

<sup>10</sup> University of the West of England, Coldharbour Ln, Stoke Gifford, Bristol BS16 1QY, UK

<sup>11</sup> School of Health Sciences, HES-SO University of Applied Sciences and Arts of Western Switzerland, Route des Arsenalux 16a, 1700 Fribourg, Switzerland

**Running head:** Brief assessment for alcohol use disorder

**Word count:** 4,426

**Declaration of competing interest:** The authors declare no competing interest. This study was supported by Swiss National Research Foundation (no. 10001C\_173418/1).

## **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 through 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 twelve 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 (AUD) nor heavy alcohol use appear to be adequate to screen for AUD among young men from the Swiss population. The best

screening alternative for AUD among young Swiss men appears to be a combination of eight symptoms of AUD and four alcohol-related consequences.

**Keywords:** alcohol; community-based sample; epidemiology; machine learning; public health; psychometrics.

# Identifying an accurate self-reported screening tool for alcohol use disorder: Evidence from a population-based assessment

## **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 as 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 out of the eleven listed in the DSM-5 diagnostic of AUD, however AUD is also thought 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 recalibrated 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 across 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 overreport AUD symptoms, so self-reported AUD may be even less reliable in some cases (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 grams of pure ethanol (ten 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

nor 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 through 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 among young Swiss men transitioning into adulthood (30, 31). Participants were initially enrolled in army recruitment centers in 2010 (average age = 20), but that setting was for enrolment only; the study itself occurred outside of 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 through 2018) and who had a valid email (n=2,668). In September 2017, we invited them to complete the ten-question version of the Alcohol Use Disorder Identification Test (AUDIT) (32) online. A total of 1,371 participants completed the questionnaire within four 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 1,371 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 flowchart is presented in Figure 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 email, 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 one-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.- (approx. US\$101 or 90€) for their participation (gift voucher) and for their travelling 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 in 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 twelve months.

*Self-reported AUD.* The participants self-reported the presence or absence of the eleven criteria of DSM-5 AUD during the previous twelve 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 twelve-month alcohol use with an extended quantity-frequency questionnaire (39). Questions address separately weekdays and weekends. We then 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). Chromatographic separation was achieved using a Merck Chromolith® RP-C18e column (150 x 3 mm). We determined the PEth homolog 16:0/18:1 concentration using methanolic extraction, followed by 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 correlations between the various alcohol-related measures.

*Receiver operating characteristic curves.* Next, we used receiver operating characteristic (ROC) curves to test whether seven measures were good indicators of the DIGS-AUD gold standard: 1) self-reported AUD, 2) twelve-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:  $\max(\text{sensitivity} + \text{specificity} - 1)$  (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 1 and 20 items). We selected the best model using the Akaike information criterion (AIC), as this method prevented to some extent overfitting 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" 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 1,000 bootstrap resamples from the original dataset, 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 Supplementary Table S1).

*Additional/sensitivity analyses.* We conducted several additional and sensitivity analyses. These include: 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 over 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 overrepresented 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 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 twelve-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 six-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 on average  $27.00 \pm 1.44$  years old. 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 Supplementary 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 Supplementary Figures S1 through 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 Supplementary Table S3).

The second and third panels of Table 2 report the performances of previous twelve-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 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=9pn/ng) and to a moderate alcohol use for PEth (PEth $\leq 112$  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 Figure S5). We also tested the psychometric properties of the AUDIT, which was used to select participants (see Figure S6). The recommended cut-off of 13 yielded an acceptable sensitivity (94.4%) but a low specificity (52.9%). The best cut-off was  $\geq 15$  (sensitivity=69.2%, specificity=75.5%).

#### *Machine learning (aim 4)*

The best alternative model combined the 11 symptoms of self-reported AUD and the 9 alcohol-related consequences (area under the ROC curve [AUC]=.857). These results are reported in Table 3 and in Supplementary Figure S7. The best cut-off was  $\geq 6$  (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=.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 Supplementary Tables S3 and S4. We then derived the performance of these 12 symptoms/consequences, as reported in Table 4 and Supplementary Figure S8. The cut-off that maximized sensitivity and specificity was  $\geq 3$  (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 .872. The corrected estimate of AUC was .841.

## **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 eleven self-reported criteria of AUD did not correspond well to the gold standard when using the recommended threshold. Indeed, 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 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, since 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).

Besides, 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 eleven criteria of self-reported AUD. A cut-off of 6 or more yielded the highest performance (sensitivity=76.9%, specificity=81.9%) out 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 rather long 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. Besides, 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 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 aftereffects, 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 after, 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. One should keep 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 not have yet 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 in a cohort study. This cohort could have differed from the whole population, as they

participated in this survey for years. A second limitation was related to the different time frames of the diagnostic of AUD (previous twelve 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 ( $r=.35$ ) and in our sample, the correlation was higher ( $r=.417$ ). 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 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.

## References

1. APA. Diagnostic and Statistical Manual of Mental Disorders, fourth ed, Washington, DC: American Psychiatric Association; 2000.
2. REHM J. How should prevalence of alcohol use disorders be assessed globally?, *Int J Methods Psychiatr Res* 2016; 25: 79-85.
3. WHO. ICD-10 classification of mental and behavioral disorder: Clinical descriptions and diagnostics guidelines, Geneva: Switzerland: World Health Organization; 1992.
4. ASSOCIATION A. P. Diagnostic and Statistical Manual of Mental Disorders, fifth ed, 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, *The American Journal on Addictions* 2016; 25: 408-415.
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, *AJP* 2013; 170: 834-851.
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: 773-780.
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: 429-440.
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, *Addiction Research & 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.
12. SLADE T., TEESON 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, *Alcoholism: Clinical and Experimental Research* 2013; 37: 1001-1007.
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: 1769-1784.
14. NUTT D. J., REHM J. Doing it by numbers: a simple approach to reducing the harms of alcohol, *J Psychopharmacol (Oxford)* 2014; 28: 3-7.

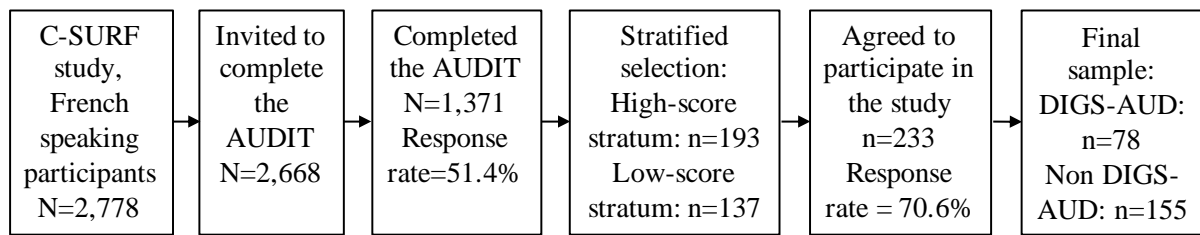
15. BAGGIO S., IGLESIAS K., STUDER J., DUPUIS M., DAEPPEN J.-B., GMEL G. Is the relationship between major depressive disorder and self-reported alcohol use disorder an artificial one?, *Alcohol Alcohol* 2015; 50: 195-199.
16. REHM J., MARMET S., ANDERSON P., GUAL A., KRAUS L., NUTT D. J. et al. Defining substance use disorders: do we really need more than heavy use?, *Alcohol Alcohol* 2013; 48: 633-640.
17. SAHA T. D., STINSON F. S., GRANT B. F. The role of alcohol consumption in future classifications of alcohol use disorders, *Drug and Alcohol Dependence* 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: 513-522.
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: 937-945.
20. KIRÁLY O., TÓTH D., URBÁN R., DEMETROVICS Z., MARAZ A. Intense video gaming is not essentially problematic, *Psychol Addict Behav* 2017; 31: 807-817.
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: 1513-1523.
23. CABARCOS P., ÁLVAREZ I., TABERNEIRO M. J., BERMEJO A. M. Determination of direct alcohol markers: a review, *Anal Bioanal Chem* 2015; 407: 4907-4925.
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., RASMUSSEN 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: 2292-2295.
27. KHARBOUCHE H., FAOUZI M., SANCHEZ N., DAEPPEN J. B., AUGSBURGER 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 Legal Med* 2012; 126: 243-250.
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 and Alcohol Dependence* 2017; 178: 80-86.
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: 251-262.

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 and Alcohol Dependence* 2013; 132: 316-323.
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: 263-270.
33. MENESES-GAYA C., ZUARDI A. W., LOUREIRO S. R., HALLAK J. E. C., TRZESNIAK C., DE AZEVEDO MARQUES 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: 1417-1424.
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, *International Journal of Telemedicine and Applications* 2016.
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 and Alcohol Dependence* 2002; 65: 149-158.
36. GRANT B. F., 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 and Alcohol Dependence* 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: 1672-1677.
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: 880-888.
40. DÉGLON J., LAUER E., SPORKERT F., VALOIS N., FAVRAT B., THOMAS A. et al. Analyse du phosphatidyléthanol sur micro-prélèvement de sang séché : nouvel outil pour le suivi de la consommation d'alcool, *Toxicologie Analytique et Clinique* 2018; 30: S42-S43.
41. YODEN W. J. Index for rating diagnostic tests, *Cancer* 1950; 3: 32-35.
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 International* 2009; 75: 1257-1263.
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: 774-781.
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 and Alcohol Dependence* 2013; 129: 60-69.
46. LIU R. T. Substance use disorders in adolescence exist along continua: Taxometric evidence in an epidemiological sample, *J Abnorm Child Psychol* 2017; 45: 1577-1586.

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.
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: 931-942.
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: 1773-1778.
50. TEESSON M., BAILLIE A. J., SLADE T. 'Truth or consequences'—what about truth or dare, *Addiction* 2014: 109: 1782-1784.
51. CHARLET K., HEINZ A. Harm reduction—a systematic review on effects of alcohol reduction on physical and mental symptoms, *Addiction Biology* 2017: 22: 1119-1159.
52. KUNTSCH 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, *International Journal of Human and Social Sciences* 2008: 3: 451-457.



Figure 1. Flow diagram of the study



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.

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

Variables	Descriptive statistics
<b>DIGS diagnosis of alcohol use disorder (gold standard)<sup>1</sup></b>	
No	66.5 (155)
Yes	33.5 (78)
<b>Self-reported alcohol use disorder<sup>1</sup></b>	
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>2</sup>	2.77 (2.26)
Self-reported AUD diagnosis ( $\geq 2$ )	67.0 (156)
<b>Self-reported alcohol-related consequences<sup>1</sup></b>	
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>2</sup>	2.01 (1.88)
<b>Alcohol use (no. of drinks per week)<sup>2</sup></b>	
Previous twelve-month use	13.90 (13.78)
<b>Biomarkers of alcohol use</b>	
Ethylglucuronide (EtG) (n=227) <sup>2</sup>	23.99 (53.76)
Chronic excessive alcohol use according to EtG ( $\geq 30$ np/mg) <sup>1</sup>	23.4 (53)
Phosphatidylethanol (PEth) (n=231) <sup>2</sup>	263.49 (304.61)
Chronic excessive alcohol use according to PEth ( $\geq 210$ ng/ml) <sup>1</sup>	42.9 (99)

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

<sup>1</sup> Percentages and n under brackets are reported.

<sup>2</sup> Means and standard deviations under brackets are reported.

Items in italic were included in the 12-item measure selected by machine learning.

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=.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 twelve-month alcohol use (no. of drinks per week) AUC=.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=.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=.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

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

<b>Cut-off</b>	<b>Sensitivity</b>	<b>Specificity</b>	<b>Correctly classified</b>
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
<b>≥ 6</b>	<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.

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)

<b>Cut-off</b>	<b>Sensitivity</b>	<b>Specificity</b>	<b>Correctly classified</b>
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.

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

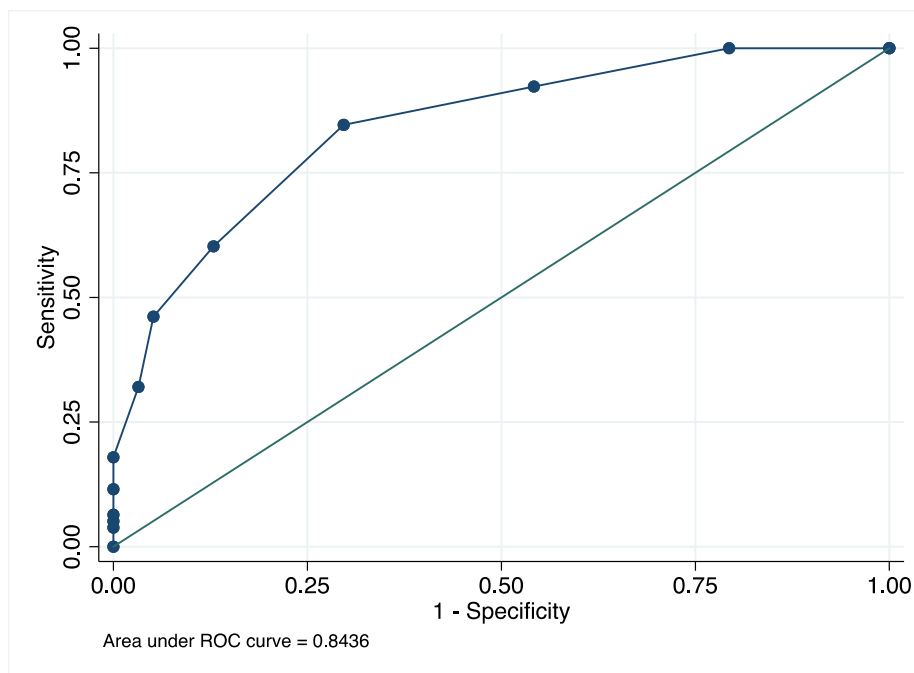


Figure S2. ROC curve for previous twelve-month alcohol use against the diagnosis of alcohol use disorder (gold standard)

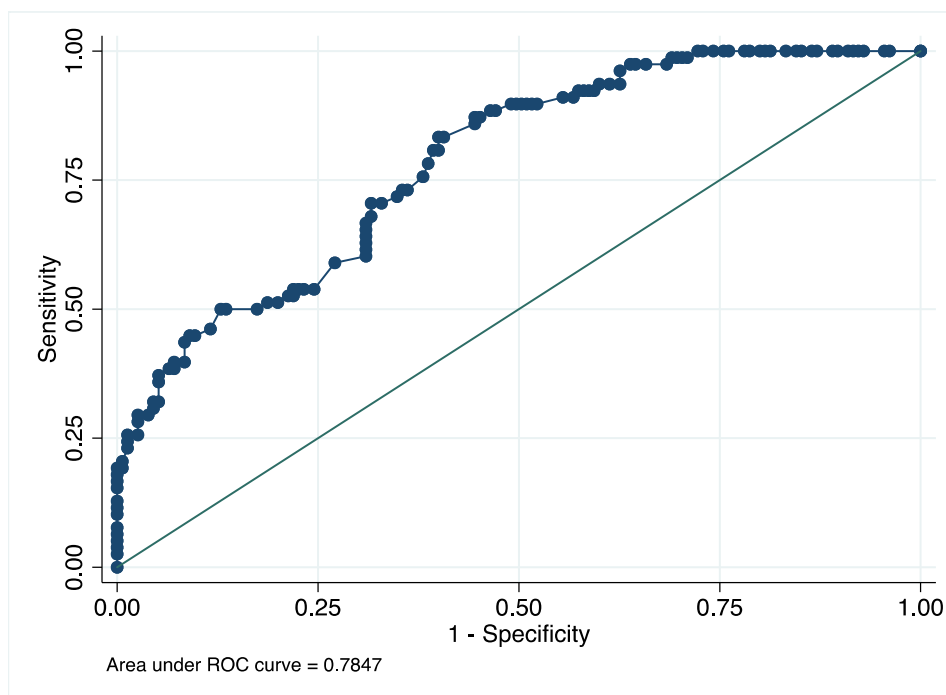


Figure S3. ROC curve for Ethyl glucuronide against the diagnosis of alcohol use disorder (gold standard)

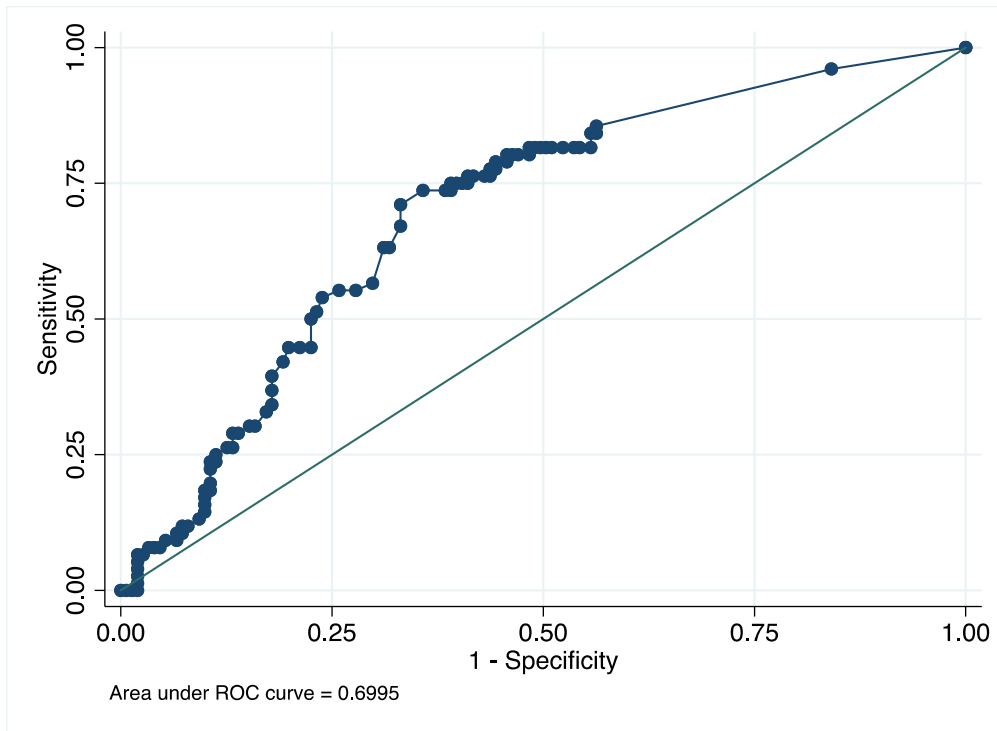


Figure S4. ROC curve for Phosphatidylethanol against the diagnosis of alcohol use disorder (gold standard)

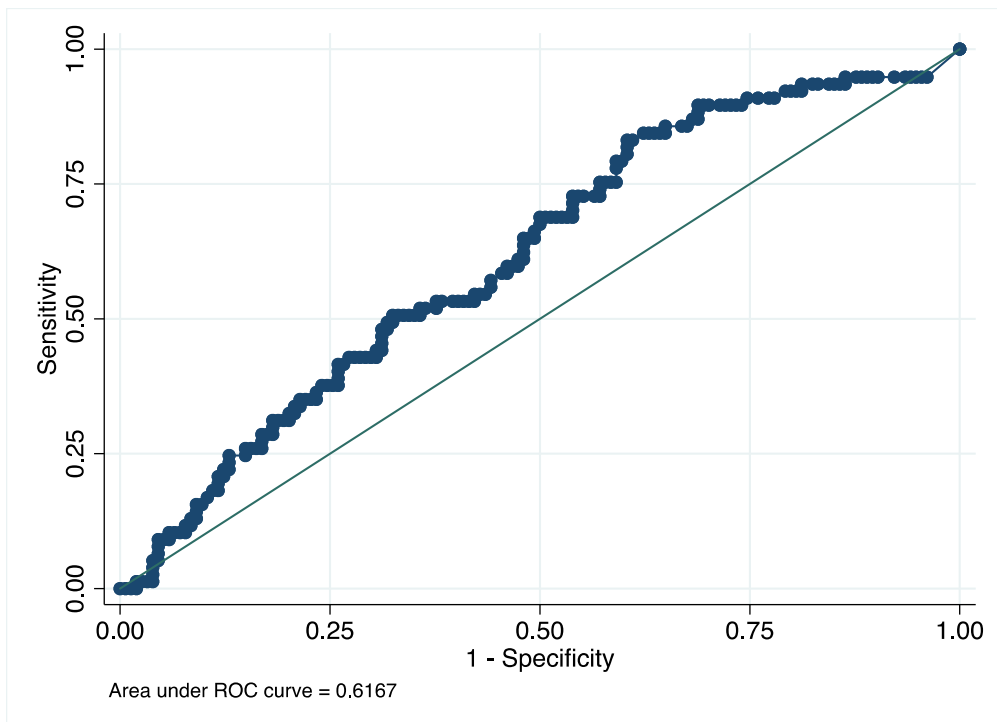


Figure S5. ROC curve for self-reported alcohol-related consequences against the diagnosis of alcohol use disorder (gold standard)

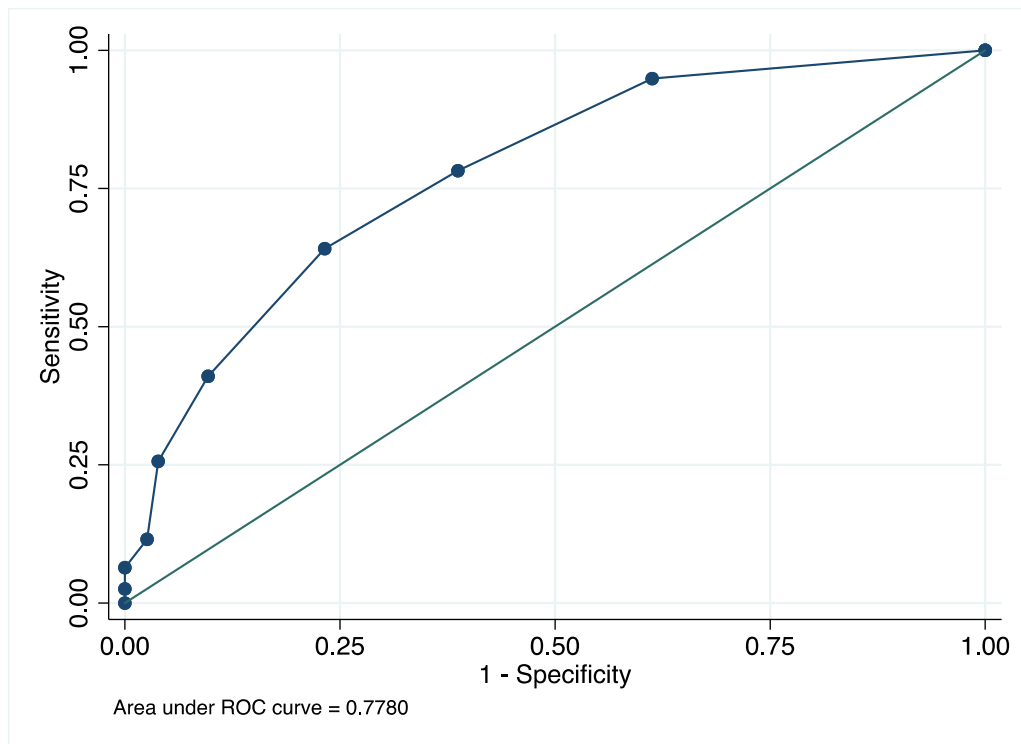


Figure S6. ROC curve for self-reported AUDIT against the diagnosis of alcohol use disorder (gold standard)

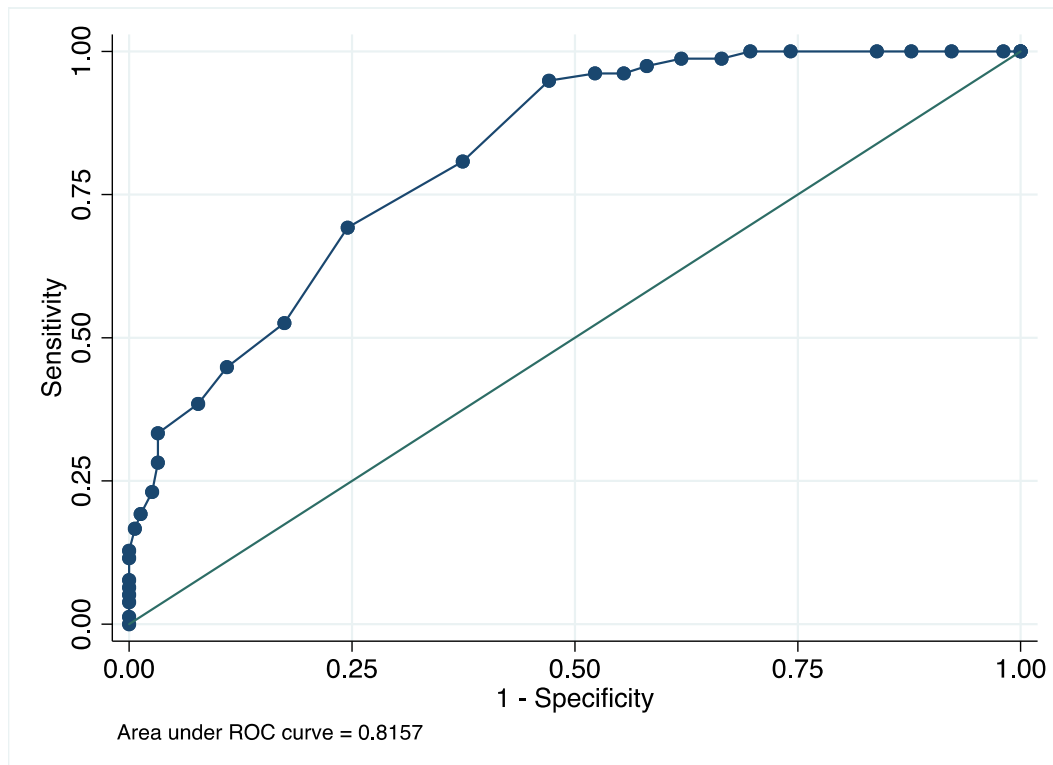




Figure S7. ROC curve for self-reported alcohol use disorder and alcohol-related consequences (20 items) against the diagnosis of alcohol use disorder (gold standard)

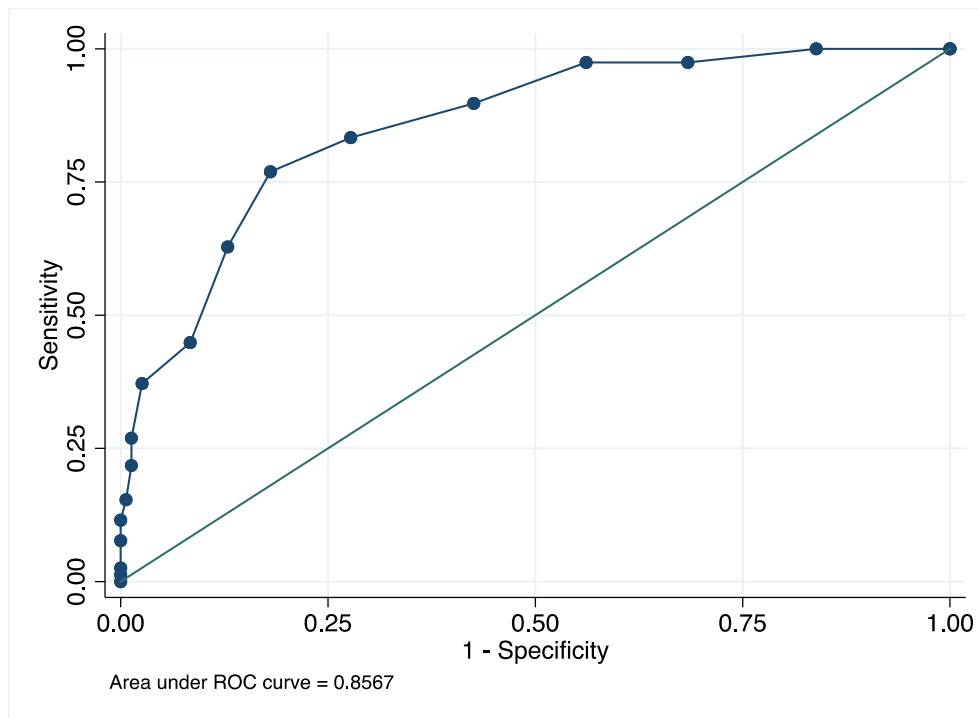


Figure S8. ROC curve for the best alternative model (12 symptoms/consequences) against the diagnosis of alcohol use disorder (gold standard)

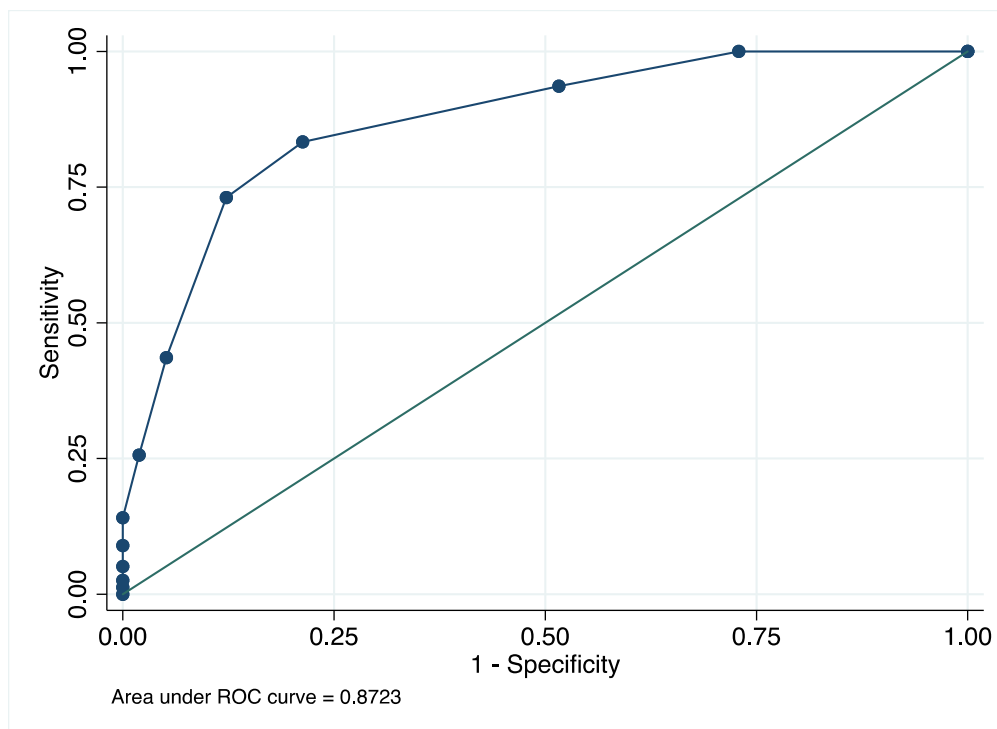


Table S1. Comparisons between respondents and non-respondents on alcohol-related variables

	Respondents	Non-respondents	p-value <sup>1</sup>	Relative bias <sup>2</sup>
n	1,371	1,293	-	-
Previous twelve-month alcohol use	7.87	7.05	.288	5.3%
Score AUDIT-C	4.88	4.69	.349	1.9%
Score self-reported AUD	1.30	1.15	.020	5.9%

AUDIT-C: Alcohol use disorder identification test, sum-score of the three first questions on alcohol use, AUD: alcohol use disorder

<sup>1</sup> p-values from negative binomial regressions.

<sup>2</sup> Relative bias =  $\frac{N_{nr}}{N_{tot}} \times (\bar{Y}_{nr} - \bar{Y}_r) \div \bar{Y}_{tot}$ , where  $\frac{N_{nr}}{N_{tot}}$  is the non-response rate,  $(\bar{Y}_{nr} - \bar{Y}_r)$  the difference between the means of respondents and non-respondents, and  $\bar{Y}_{tot}$  the total mean (see Studer et al., 2013).

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

	2	3	4	5	6	7
1. Diagnostic of alcohol use disorder <sup>1</sup>	.570	.382	.462	.184	.328	.475
2. Self-reported alcohol use disorder	-	.825	.641	.373	.387	.620
3. Self-reported alcohol use disorder (binary) <sup>1</sup>	-	-	.574	.361	.359	.529
4. Previous twelve-month alcohol use	-	-	-	.566	.595	.523
5. Ethyl glucuronide (EtG)	-	-	-	-	.595	.329
6. Phosphatidylethanol (PEth)	-	-	-	-	-	.414
7. Self-reported alcohol-related consequences	-	-	-	-	-	-

<sup>1</sup> Rank biserial correlations were used.

All correlations were statistically significant with  $p < .001$ .

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

<b>Question</b>	<b>Sensitivity</b>	<b>Specificity</b>	<b>AUC</b>	<b>OR</b>	<b>p-value</b>
A1. Role neglect	45/78=58%	125/155=81%	.692	5.68	<.001
A2. Hazardous use	64/78=82%	69/155=45%	.633	3.67	<.001
A3. Social problems	20/78=26%	148/155=95%	.606	7.29	<.001
A4. Tolerance	35/78=45%	141/155=91%	.679	8.20	<.001
A5. Withdrawal	21/78=27%	149/155=96%	.615	9.15	<.001
A6. Larger/longer	64/78=82%	64/155=41%	.617	3.22	.001
A7. Loss control	16/78=21%	147/155=95%	.577	4.74	.001
A8. Time spent	46/78=59%	124/155=80%	.695	5.75	<.001
A9. Give up activities	9/78=12%	152/155=98%	.548	6.61	.006
A10. Health problems	23/78=29%	148/155=95%	.625	8.84	<.001
A11. Craving	16/78=21%	152/155=98%	.593	13.08	<.001
C1. Get over	31/78=40%	131/155=85%	.621	3.60	<.001
C2. Mental blackout	54/78=69%	99/155=64%	.666	3.98	<.001
C3. Regret	54/78=69%	101/155=65%	.672	4.21	<.001
C4. Unplanned sex	33/78=42%	132/155=85%	.637	4.21	<.001
C5. No condom	15/78=19%	142/155=92%	.554	2.60	.019
C6. Accident/injury	28/78=36%	137/155=88%	.621	4.26	<.001
C7. Conflict authorities	7/78=9%	152/155=98%	.535	5.00	.022
C8. Argument/flight	22/78=28%	140/155=90%	.593	3.67	<.001
C9. Property damage	9/78=12%	145/155=94%	.525	1.89	.186

AUC: area under the curve, OR: odds-ratio

Labels for self-reported alcohol use disorder (A1-A11) and alcohol-related consequences (C1-C9) are reported in Table 1.

All items were significantly and positively associated with the gold standard (diagnostic interview for alcohol use disorder), except C9.

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

No. of items	Items included	AIC	AUC
1	A4 (tolerance)	262.7	.679
2	+A10 (health problems)	239.9	.749
3	+A8 (time spent)	228.3	.801
4	+C3 (regret) +A11 (craving) -A10	216.5	.832
5	+C4 (unplanned sex) +A5 (withdrawal) +A10 -C3 -A8	208.5	.828
6	+A2 (hazardous use)	203.1	.850
7	+A8	200.3	.860
8	+C3	197.7	.866
9	+A7 (loss control)	195.8	.868
10	+C6 (accident/injury)	195.3	.872
11	+A9 (give up activities)	195.0	.872
<b>12</b>	<b>+C7 (conflict authorities)</b>	<b>194.8</b>	<b>.872</b>
13	+C1 (get over)	195.3	.870
14	+A1 (role neglect)	196.5	.871
15	+A3 (social problems)	198.0	.869
16	+C2 (mental blackout)	199.3	.869
17	+C8 (argument/flight)	201.1	.867
18	+A6 (larger/longer)	202.8	.862
19	+C9 (property damage)	204.2	.860
20	+C5 (no condom)	206.5	.857

AIC: Akaike Information Criterion, AUC: area under the curve

Labels for self-reported alcohol use disorder (A1-A11) and alcohol-related consequences (C1-C9) are reported in Table 1.

The best model is highlighted in bold.

## S5 Table. Final screening tool in English

### **In the past 12 months...**

- A2-1 Did you more than once drive a car or another vehicle (such a car, bicycle, motorcycle or moped) shortly after you had had several dinks with alcohol?
- A2-2 Did you find yourself more than once in a situation that increased your chances of getting injured (sing machines, walking or doing sport in a dangerous area or around heavy traffic) after you had been drinking too much alcohol?
- A4 Did you find you needed a lot more alcohol to become high o drunk than you used to?
- A5 Did you start feeling nervous or shaky for a full day or more after you had cut down on your drinking?
- A7 Did you try to cut down on your drinking, but couldn't?
- A8 Did you find yourself spending a great deal of time obtaining, using, or recovering from the effects of alcohol?
- A9 Did you give up activities you care about (e.g., school, work or being with friends and family) because of your drinking?
- A10 Did you continue drinking even though you were aware that alcohol had repeatedly caused you anxiety, depression or health problems?
- A11 Have you had such a strong desire or urge to drink that you could not help drinking?

### **In the last 12 months, it happened that...**

- C3 While drinking alcohol, I did something that I badly regretted later.
- C4 I had unplanned sex because I was drunk.
- C6 I had an accident or I got injured because I was drunk.
- C7 I came into a conflict with the police or with the authorities more than once because of my alcohol use.

## S6 Table. Final screening tool in French

### **Au cours des 12 derniers mois, vous est-il arrivé...**

- A2-1 De conduire plus d'une fois un véhicule (par exemple voiture vélo, vélomoteur, scooter ou moto) après avoir bu quelques verres ?
- A2-2 De vous être trouvé plus d'une fois dans une situation dans laquelle vous auriez pu vous blesser après avoir trop bu (par exemple avec des outils au travail ou sur des machines, en faisant du sport ou d'autres situations dangereuses telles que marcher au milieu de la route) ?
- A4 D'avoir l'impression que vous devez consommer beaucoup plus d'alcool qu'auparavant pour en ressentir les effets et être saoul ?
- A5 D'avoir tremblé ou de vous être senti nerveux pendant toute une journée ou plus après avoir arrêté votre consommation ?
- A7 D'avoir essayé de réduire votre consommation, mais sans succès ?
- A8 D'avoir passé beaucoup de temps soit à vous saouler, à être saoul, ou à dessaouler ?
- A9 D'avoir abandonné des activités importantes telles que l'école, le travail ou les rencontres entre amis ou avec la famille à cause de votre consommation d'alcool ?
- A10 D'avoir continué votre consommation d'alcool bien que vous saviez qu'elle avait généré des problèmes récurrents de santé ou des problèmes psychiques tels que l'anxiété ou la dépression ?
- A11 D'avoir ressenti un besoin ou un désir si fort que vous ne pouviez pas faire autrement que de boire de l'alcool ?

### **Au cours des 12 derniers mois, avez-vous expérimenté les conséquences suivantes ?**

- C3 J'ai fait quelque chose en buvant que j'ai regretté plus tard.
- C4 J'ai eu une relation sexuelle qui n'était pas prévue parce que j'avais bu.
- C6 J'ai eu un accident ou je me suis blessé parce que j'avais bu.
- C7 J'ai eu plus d'une fois des ennuis avec la police ou d'autres représentants de l'ordre publique à cause de ma consommation d'alcool.