



Are short AUDIT screeners effective in identifying unhealthy drinking of varying severity? A prison population study

Pape Hilde^{a,*}, Rossow Ingeborg^b, Bukten Anne^{a,c}

^a University College of Norwegian Correctional Service, Norway

^b Norwegian Institute of Public Health, Norway

^c Norwegian Centre for Addiction Research, University of Oslo, Section for Clinical Addiction Research, Oslo University Hospital, Norway

ARTICLE INFO

Keywords:

AUDIT
Short versions
Screening
Validation study
Alcohol problems
Prison

ABSTRACT

Background: Whether brief versions of the Alcohol Use Disorder Identification Test (AUDIT) can be used as graded severity measures is largely unknown. We examined the performance of eight such brief screeners in a prison population, and compared their effectiveness in detecting hazardous drinking, harmful drinking, and possible alcohol dependence as classified by the full ten-item AUDIT.

Methods: The study sample included pre-prison drinkers who participated in the Norwegian Offender Mental Health and Addiction (NorMA) study (n = 758). We conducted receiver operating characteristic curve (ROC) analyses and estimated the area under the curve (AUROC) to assess the performance of AUDIT-C (three consumption items) and four-item versions that consisted of AUDIT-C and one additional item.

Results: AUDIT-C performed very well in detecting unhealthy drinking of varying severity (AUROCs of 0.933 or 0.935). Four-item versions performed even better. Of these, the well-established AUDIT-4 was superior in identifying harmful drinking (AUROC=0.969) and possible alcohol dependence (AUROC=0.976). For AUDIT-C, the optimal cut-points in terms of the highest combined sensitivity and specificity were ≥ 6 (hazardous drinking), ≥ 8 (harmful drinking) and ≥ 8 or ≥ 9 (possible dependence). The corresponding cut-points on AUDIT-4 were ≥ 6 , ≥ 9 and ≥ 10 . The highest cut-point whereby *all* cases of possible dependence were identified was ≥ 6 on AUDIT-C and ≥ 8 on AUDIT-4. At these cut-points, almost all individuals with harmful drinking were also detected.

Conclusions: AUDIT-C and AUDIT-4 were both highly effective in detecting hazardous drinking, harmful drinking and possible alcohol dependence. AUDIT-4 was superior, notably as a graded severity measure.

1. Introduction

Hazardous drinking and alcohol problems are prevalent among individuals who enter prison (Fazel et al., 2017; Newbury-Birch et al., 2016; Seal et al., 2018), which is unsurprising; alcohol is implicated in various types of crime, notably violence (Evans et al., 2021; Graham and West, 2001; Rossow and Bye, 2013), and pre-prison heavy drinking is predictive of recidivism (Dowden and Brown, 2002). In addition to a range of adverse health and psychosocial outcomes (Babor et al., 2010; Rehm et al., 2017), excessive alcohol use is associated with an increased suicide risk during incarceration (Fazel et al., 2008) and an elevated post-release mortality rate (Chang et al., 2015). It is thus imperative to identify those who have been drinking heavily, and to offer adequate interventions.

Many European countries screen individuals for alcohol problems upon entry into prison, but validated tools are rarely used (WHO, 2019a, 2019b). The prison service in England and Wales stands out in this respect, as universal screening based on the Alcohol Use Disorder Identification Test (AUDIT) has been put into practice. The AUDIT has exhibited high validity across nations and settings (Babor and Robaina, 2016; Meneses-Gaya et al., 2009) and been described as a gold standard in detecting drinking behavior that is potentially or currently harmful to health. Studies of prison populations also suggest high validity of this screening tool (Coulton et al., 2012; Thomas et al., 2014) and that the reliability may be higher when the screening occurs a few weeks after incarceration rather than upon entry into prison (Maggia et al., 2004).

The AUDIT was not merely designed to identify risky drinking behavior or alcohol problems, but also to suggest interventions that vary

* Correspondence to: University College of Norwegian Correctional Service, P.O. Box 1, 2001 Lillestrøm, Norway.
E-mail address: Hilde.Pape@krus.no (P. Hilde).

according to the severity of the problems (Babor, et al., 2001; Saunders et al., 1993). The ten items capture alcohol consumption (items 1–3), dependence symptoms (items 4–6) and alcohol-related harm (items 7–10), and the total score ranges from 0 to 40. Scores ≤ 7 denote a negative screen, that is, abstinence or low-risk drinking. The standard categorization of positive screens and the recommended interventions are as follows (Babor and Robaina, 2016):

- Hazardous alcohol use (scores 8–15): Brief intervention
- Harmful alcohol use (scores 16–19): Brief intervention, further monitoring, and diagnostic evaluation
- Possible alcohol dependence (scores ≥ 20): Specialist treatment

The assessment of individuals who enter prison typically covers a broad range of domains, and full versions of screening instruments may add undue burden to the intake procedure. Short versions of the AUDIT may also be preferable for other reasons. Brady et al. (2002) found that many items were perceived as “intrusive” or “prying into private lives” by those being screened. The consumption items tended to be perceived as less invasive and easier to comprehend than those capturing dependence symptoms and alcohol-related harm.

The three AUDIT consumption items are also essential to the instrument’s validity (Higgins-Biddle and Babor, 2018), and they constitute the well-established AUDIT-C. This brief screener has exhibited high effectiveness in detecting unhealthy drinking behavior as determined by external validation criteria such as diagnostic assessment of alcohol use disorder (Kriston et al., 2008; Toner et al., 2019). AUDIT-C has also shown to perform approximately as well as the full AUDIT in various contexts.

Other studies have examined the performance of abbreviated AUDIT screeners against the full screening instrument. The bulk of this research focuses on the effectiveness of the AUDIT-C in detecting a positive screen (AUDIT scores ≥ 8) (e.g. Morojele et al., 2017; Nehlin, et al., 2012; Neumann et al., 2012; Seth et al., 2015), and high levels of accuracy have generally been reported. However, when the screening solely identifies individuals with either hazardous drinking, harmful drinking, or possible dependence, the results provide no cues with respect to likely treatment needs.

To our knowledge, only two studies have examined whether brief AUDIT screeners may work as graded severity measures. One was based on a sample of Aboriginal Australians (Calabria et al., 2014), the other was a general population study from Japan (Osaki et al., 2014). Both studies focused on AUDIT-C, and both concluded that it performed well in identifying both hazardous, harmful, and possible dependent drinking (as classified by the full AUDIT).

One may expect that four-item AUDIT screeners perform even better than the three-item AUDIT-C. The AUDIT-4 includes one item about other people’s concern about one’s drinking (item 10) in addition to AUDIT-C, but validation studies indicate that it performs only slightly better than AUDIT-C (Gual et al., 2002; Lee et al., 2018; Meneses-Gaya et al., 2010; Wu et al., 2008). One may thus ask whether other combinations of AUDIT-C and one additional item are preferable. No studies seem to have addressed this question thus far.

A final issue of interest pertains to the effectiveness of brief AUDIT versions in screening prison populations. Caviness et al. (2009) examined the accuracy of AUDIT-C and item 3 alone (i.e. the frequency of consuming 6 + units) in a sample of incarcerated females and their results were encouraging – notably for AUDIT-C. However, a positive screen on the full AUDIT was the one and only reference standard. No additional studies seem to have explored the validity of brief AUDIT screeners among imprisoned persons. Similar studies of other groups with a high prevalence of alcohol problems are also few and far between.

Against this backdrop, we examined the effectiveness of AUDIT-C and of four-item versions that consisted of AUDIT-C and one additional item in identifying unhealthy drinking behavior of varying severity (as classified by the full AUDIT) in a prison population. We

compared the performance of these brief screeners and examined cut-points on AUDIT-C and on the best performing four-item version.

2. Material and methods

Data stemmed from the Norwegian Offender Mental Health and Addiction (NorMA) study (Bukten et al., 2015). The target sample comprised all incarcerated individuals in Norway at the time of the data collection (June 2013–July 2014), and there were no pre-defined exclusion criteria.

The questionnaire was translated into four languages, but some individuals could not read any of these. Other reasons for non-participation were temporarily absence from the prison unit and preclusion of study eligibility by prison authorities for security reasons. About 40% of the prison population responded ($n = 1499$), and the sample was representative with respect to many demographic variables (Bukten et al., 2015).

Participation was voluntary and based on written informed consent. It was explicitly pointed out that refraining from participation was not associated with any sanctions and that the responses to the self-report questionnaire were strictly confidential. The NorMA-study was approved by the Norwegian Committee of Research Ethics. Details about data collection and ethics are reported elsewhere (Bukten et al., 2015; Bukten et al., 2016).

2.1. Study sample

Eighty percent reported that they had consumed alcohol in the year before incarceration. From this group, we extracted a subsample of 758 respondents whose current imprisonment had lasted 12 months or less, and who had responded to all the ten AUDIT items.

2.2. Measures

2.2.1. Alcohol Use Disorder Identification Test (AUDIT)

The original AUDIT consumption items are formulated in present tense without a specified time frame, while all other items refer to the past year (Babor et al., 2001). However, a past-year reference period is frequently used for all the items (e.g. Bush et al., 1998; Cook et al., 2005; Towers et al., 2011). In our study, the AUDIT items were modified to

Table 1
The ten AUDIT items as formulated in the present study.

<i>Alcohol consumption</i>	
1	How often did you have a drink containing alcohol in the year before incarceration?
2	How many drinks containing alcohol did you have on a typical day when you are drinking in the year before incarceration?
3	How often did you drink 6 or more units on one occasion in the year before incarceration?
<i>Symptoms of dependence</i>	
4	How often in the year before incarceration were you not able to stop drinking once you had started?
5	How often in the year before incarceration did you fail to do what was normally expected from you because of drinking?
6	How often in the year before incarceration did you need a first drink in the morning to get yourself going after a heavy drinking session?
<i>Alcohol-related harm</i>	
7	How often in the year before incarceration did you have a feeling of guilt or remorse after drinking?
8	How often in the year before incarceration were you unable to remember what happened the night before because you had been drinking?
9	Have you or someone else been injured as a result of your drinking?
10	Has a relative or friend or a doctor or another health worker been concerned about your drinking, or suggested you to cut down?

Note: There were five response options for items 1–8 (scale range: 0–4), and three response options for items 9 and 10; “No” (scored 0), ‘Yes, but not in the year before incarceration’ (scored 2), and ‘Yes, in the year before incarceration’ (scored 4).

assess the 12 months prior to imprisonment (Table 1).

A positive screen (scores ≥ 8), at least harmful drinking (scores ≥ 16), and possible alcohol dependence (scores ≥ 20) were the classification standards. In addition to AUDIT-C (item 1 +2 +3), we constructed all possible versions that consisted of the AUDIT-C and one additional item. We used the term ‘AUDIT-10’ rather than ‘the full AUDIT’ when describing our results.

2.2.2. Demographics

We used data on gender, age, and current imprisonment length.

2.3. Statistical analyses

We conducted receiver operating characteristic (ROC) curve analysis (Hanley, 1989), and calculated the area under the curve (AUROC) to examine the performance of the brief AUDIT screeners in detecting unhealthy drinking behavior of varying severity (as classified by AUDIT-10). ROC curves plot the true positive rate (sensitivity) against the false positive rate (1-specificity) for all cut-points, and the AUROC ranges from 0.5 (no better than chance) to 1.0 (perfect match). The AUROC has been considered fair for values between 0.7 and 0.8, good for values between 0.8 and 0.9 and excellent for values above 0.90 (Cuparencu et al., 2020). Because we examined the performance of brief AUDIT screeners against the full AUDIT, and thus relied on internal reference standards, the likelihood of obtaining high AUROC-values was elevated.

We compared ROC curves using z-statistics for paired design (DeLong et al., 1988) to test whether the four-item AUDIT screeners performed significantly better than the AUDIT-C, and to identify the screener that most accurately identified possible alcohol dependence. Next, we examined cut-points for each of the three categories of unhealthy drinking, and applied Youden's (1950) J to identify the highest combined level of sensitivity and specificity $J = \frac{\%sensitivity + \%specificity}{100} - 1$. We also calculated the positive predictive value (PPV) and the negative predictive value (NPV) of various cut-points.

In this study, sensitivity refers to the percentage of the positive cases on the AUDIT-10 that were also identified as positive cases by the brief AUDIT screener. Specificity is the percentage of negative cases on the AUDIT-10 that were correctly identified as such by the brief screener. PPV is the percentage of the positive cases on the brief screener that were also positive cases on AUDIT-10, while NPV is the percentage of the negative cases on the brief screener who were also identified as such on AUDIT-10. Thus, the higher the PPV the lower is the occurrence of false positives, and the higher the NPV the lower is the occurrence of false negatives. Unlike sensitivity and specificity, the PPV and the NPV depend on the prevalence of “true” positive cases in the sample.

We used MedCalc Statistical Software 19.5.3 (ROC analyses) and SPSS 26 (cross-tabulations with χ^2 -test).

2.3.1. Sensitivity analyses

Screening for health and psychosocial problems typically occurs shortly after imprisonment. Therefore, we compared the percentage distribution across the four AUDIT-10 categories for individuals reporting short-term (<3 months) and those reporting longer-term (3–12 months) imprisonment. Next, we examined whether the main findings persisted when restricting the ROC analyses to the former group.

3. Results

3.1. Sample description

The vast majority (93%) of the respondents were males, 63% were 35 years old or younger, and 49% had been imprisoned less than 3 months (Table 2). Moreover, two thirds had an AUDIT-10 positive screen; 34%

Table 2
Descriptive statistics of the study sample¹.

		%	(n)
Gender	Males	92.6	(699)
	Females	7.4	(56)
Age	17–25 years	26.4	(184)
	26–35 years	37.0	(258)
	≥ 36 years	37.6	(255)
Imprisonment length	< 3 months	48.9	(371)
	3–6 months	27.3	(207)
	> 6 months ¹	23.7	(180)
AUDIT-10 categories (score range)	Low-risk drinking (1–7)	33.9	(257)
	Hazardous drinking (8–15)	33.6	(255)
	Harmful drinking (16–19)	10.4	(79)
	Possible dependence (≥ 20)	22.0	(167)

¹ Those incarcerated > 12 months were excluded

had hazardous drinking behavior (scores of 8–15), 10% had harmful drinking behavior (scores of 16–19), and 22% were possibly alcohol dependent (scores ≥ 20).

3.2. Performance of the brief AUDIT screeners

The AUROCs for detecting each of the three categories of unhealthy drinking exceeded 0.90 for all the brief versions of the AUDIT (Table 3). Moreover, the four-item versions performed significantly better than the AUDIT-C, yet there was one exception; AUDIT-C + item 9 was not more effective than AUDIT-C alone in detecting possible alcohol dependence (AUDIT-10 scores ≥ 20).

AUDIT-4 had the highest AUROC for identifying possible dependence. Additional analyses showed that it differed significantly from AUDIT-C + item 4 ($p = 0.011$), AUDIT-C + item 5 ($p = 0.002$), AUDIT-C + item 6 ($p < 0.001$), AUDIT-C + item 8 ($p = 0.003$) and AUDIT-C + item 9 ($p < 0.001$) in this respect. AUDIT-4 had also the highest AUROC for detecting at least harmful drinking (AUDIT-10 scores ≥ 16). Its effectiveness in doing so exceeded that of AUDIT-C + item 5 ($p = 0.022$), AUDIT-C + item 6 ($p < 0.001$), AUDIT-C + item 8 ($p = 0.010$) and AUDIT-C + item 9 ($p = 0.019$). Hence, we selected AUDIT-4 for further analyses.

3.3. Cut-points on AUDIT-C and AUDIT-4

Table 4 provides the sensitivity, the specificity, Youden's J, the positive predictive value and the negative predictive value for selected cut-points on the AUDIT-C and the AUDIT-4 for each of the three AUDIT-10 categories of unhealthy drinking. Starting with AUDIT-C, a cut-point of ≥ 6 had the highest Youden's J for detecting individuals with an AUDIT-10 positive screen (sensitivity: 86%, specificity: 88%). The PPV was 94% (i.e. 6% false positives) and the NPV was 77% (i.e. 23% false negatives). The highest J-value for at least harmful drinking was reached at a cut-point of ≥ 8 (sensitivity: 85%, specificity: 85%), closely followed by a cut-point of ≥ 7 (sensitivity: 94%, specificity: 76%). The former had higher PPV (74%) than the latter (65%), while the NPVs were 96% and 92%, respectively. The optimal cut-point for possible alcohol dependence was either ≥ 8 (sensitivity: 93%, specificity: 78%) or ≥ 9 (sensitivity: 80%, specificity: 91%). The PPV was much lower at a cut-point of ≥ 8 (54%) than at a cut-point of ≥ 9 (71%), while the NPVs were almost equally high (98% and 94%, respectively). The percentages scoring ≥ 6 , ≥ 8 and ≥ 9 on AUDIT-C were 61%, 38% and 25%, respectively. Table 4 also shows that the cut-point of ≥ 6 was the highest whereby all individuals with possible alcohol dependence were identified. This cut-point also captured almost all (98%) individuals with at least harmful drinking.

Moving to AUDIT-4, the highest J-value for a positive screen was observed at a cut-point of ≥ 6 (sensitivity: 89%, specificity: 88%). The PPV and the NPV were 93% and 81%, respectively. The optimal cut-point for at least harmful drinking was ≥ 9 (sensitivity: 89%,

Table 3

AUROC for detecting AUDIT-10 categories of unhealthy drinking behavior, and differences between the AUDIT-C and the four-item screeners. The screener with the highest observed AUROC appears in bold.

AUDIT-10 categories	AUROC (95% CI)	Comparison with AUDIT-C
Positive screen¹		
AUDIT-C (item 1 +2 +3)	0.935 (0.915–0.951)	–
AUDIT-4 (item 1 +2 +3 +10)	0.953 (0.935–0.967)	p < 0.001
AUDIT-C + item 4	0.950 (0.932–0.964)	p < 0.001
AUDIT-C + item 5	0.951 (0.937–0.964)	p < 0.001
AUDIT-C + item 6	0.946 (0.927–0.961)	p < 0.001
AUDIT-C + item 7	0.955 (0.938–0.969)	p < 0.001
AUDIT-C + item 8	0.953 (0.935–0.967)	p < 0.001
AUDIT-C + item 9	0.967 (0.952–0.979)	p < 0.001
At least harmful drinking²		
AUDIT-C (item 1 +2 +3)	0.933 (0.913–0.950)	–
AUDIT-4 (item 1 +2 +3 +10)	0.969 (0.955–0.981)	p < 0.001
AUDIT-C + item 4	0.959 (0.942–0.972)	p < 0.001
AUDIT-C + item 5	0.956 (0.939–0.970)	p < 0.001
AUDIT-C + item 6	0.946 (0.927–0.961)	p = 0.001
AUDIT-C + item 7	0.961 (0.944–0.973)	p < 0.001
AUDIT-C + item 8	0.954 (0.937–0.968)	p < 0.001
AUDIT-C + item 9	0.954 (0.937–0.968)	p = 0.001
Possible dependence³		
AUDIT-C (item 1 +2 +3)	0.935 (0.915–0.951)	–
AUDIT-4 (item 1 +2 +3 +10)	0.976 (0.963–0.986)	p < 0.001
AUDIT-C + item 4	0.961 (0.944–0.973)	p < 0.001
AUDIT-C + item 5	0.957 (0.940–0.970)	p < 0.001
AUDIT-C + item 6	0.948 (0.930–0.963)	p = 0.003
AUDIT-C + item 7	0.967 (0.951–0.978)	p < 0.001
AUDIT-C + item 8	0.958 (0.942–0.971)	p < 0.001
AUDIT-C + item 9	0.942 (0.923–0.958)	p = 0.281

¹AUDIT-10 scores ≥ 8 ²AUDIT-10 scores ≥ 16 ³AUDIT-10 scores ≥ 20

specificity: 92%), with a PPV of 84% and a NPV of 94%. Regarding possible alcohol dependence, a cut-point of ≥ 10 yielded the highest J-value (sensitivity: 96%, specificity: 82%). The PPV was 71% and the NPV was 99%. The percentages scoring ≥ 6 , ≥ 9 and ≥ 10 on AUDIT-4 were 63%, 34% and 30%, respectively. Moreover, a cut-point of ≥ 8 was the highest to identify *all* individuals with possible alcohol dependence. It also identified 97% of those with at least harmful drinking.

3.4. Sensitivity analyses

The percentage distribution across the four AUDIT-10 severity categories was not significantly different for respondents who had been imprisoned < 3 months and those who had spent 3–12 months in prison ($p = 0.085$). Moreover, all main findings were replicated when we restricted the ROC analyses to the former group ($n = 371$). Specifically,

the AUROCs for the eight brief AUDIT screeners consistently exceeded 0.90, almost all the four-item versions performed significantly better than AUDIT-C, and AUDIT-4 had the highest AUROC for identifying at least harmful drinking and possible alcohol dependence. It may be noted that a majority (58%) of the respondents in these analyses had been imprisoned 30 days or less.

4. Discussion

Our study included individuals who had consumed alcohol in the year before incarceration, and two thirds had a positive screen on the full AUDIT. In this high-prevalence sample, AUDIT-C performed very well in identifying both hazardous drinking, harmful drinking, and possible alcohol dependence (as classified by the full AUDIT). Four-item versions that consisted of the AUDIT-C and one additional item performed even better. Of these, AUDIT-4 was superior in detecting harmful drinking and possible dependence.

We also identified optimal cut-points in terms of the highest combined sensitivity and specificity (i.e. Youden's J). Regarding AUDIT-C, they were ≥ 6 for a positive screen, ≥ 8 for at least harmful drinking, and ≥ 8 or ≥ 9 for possible alcohol dependence. Thus, the cut-points for the two most severe drinking categories were indistinctive. The corresponding cut-points on AUDIT-4 were ≥ 6 (positive screen), ≥ 9 (at least harmful drinking) and ≥ 10 (possible dependence). However, cut-points with the highest Youden's J are not necessarily the "best" choice, as we will discuss later.

4.1. Comparisons with other studies

Our study focused on issues that barely have been addressed in previous research, but two studies have examined the effectiveness of AUDIT-C as a graded severity measure (Calabria et al., 2014; Osaki et al., 2014). Both found that a cut-point of ≥ 6 captured (virtually) all individuals with the more severe alcohol problems (i.e. scores ≥ 16 on the full AUDIT). This result was replicated in our study.

The cut-points with the highest Youden's J for at least harmful drinking and for possible dependence differed between the above-mentioned studies, and from those in our study. This may reflect that the sample size and the proportion scoring high on the full AUDIT varied markedly between the studies. Thus, the higher they are, the better the accuracy of the brief screener.

One may expect that four-item versions of an instrument perform better than a three-item version because they are likely to capture more of the variance. Indeed, AUDIT-C and one additional item performed better than AUDIT-C alone, which agrees with previous comparisons of the effectiveness of AUDIT-C and AUDIT-4 (Gual et al., 2002; Lee et al., 2018; Meneses-Gaya et al., 2010; Wu et al., 2008). However, few – if any – studies have explored whether AUDIT-4 performs better than other combinations of AUDIT-C and one additional item. Whether AUDIT-4 may work as a graded severity measure is another issue that seems to have been overlooked.

Finally, only one previous study to our knowledge has examined the validity of brief AUDIT screeners in a prison population (Caviness et al., 2009). It was restricted to females, and a positive screen on the full AUDIT was the only reference standard. Moreover, two studies examined whether brief AUDIT screeners may work as graded severity measures, but none of them employed samples that were comparable to ours (Calabria et al., 2014; Osaki et al., 2014).

4.2. Choosing cut-points

Generally, the choice of cut-points depends on the screening purpose. If monitoring and crude assessment of the nature and the scale of alcohol problems in a population are the main purposes, relying on Youden's J seems reasonable. One should consider other cut-points when the aim is to identify individuals in likely need of professional

Table 4

Sensitivity, specificity, positive predictive value (PPV) and negative predictive values (NPV) for selected cut-points on AUDIT-C (scale:1–12) and AUDIT-4 (scale 1–16) for identifying unhealthy drinking behavior of varying severity (as classified by AUDIT-10). Optimal cut-points according to.

AUDIT-10 categories	Cut point	AUDIT-C						AUDIT-4					
		% of sample	Sens. %	Spec.%	J	PPV%	NPV%	% of sample	Sens.%	Spec.%	J	PPV%	NPV%
Positive screen ¹	≥ 4	78.8	95.8	54.5	0.50	80.4	87.0	81.5	98.4	51.4	0.50	79.0	94.3
	≥ 5	69.0	91.2	74.3	0.66	87.4	81.3	71.8	94.4	72.4	0.67	86.9	86.9
	≥ 6	60.9	86.2	88.3	0.75	93.5	76.7	63.3	89.4	87.6	0.77	93.3	80.9
	≥ 7	47.0	69.9	97.7	0.68	98.3	62.4	51.1	75.9	97.3	0.73	98.2	67.4
	≥ 8	37.6	56.9	100	0.57	100	54.3	44.5	67.3	100	0.67	100	61.0
At least harmful drinking ²	≥ 6	60.9	98.4	57.0	0.55	52.4	98.6	63.3	100	54.3	0.54	51.2	100
	≥ 7	47.0	93.9	75.6	0.70	64.9	96.3	51.1	98.8	71.9	0.71	62.8	99.2
	≥ 8	37.6	85.4	85.4	0.71	73.7	92.4	44.5	96.8	80.7	0.78	70.6	98.1
	≥ 9	24.9	67.5	95.6	0.63	87.8	85.9	34.3	88.6	91.8	0.80	83.8	94.4
	≥ 10	–	–	–	–	–	–	29.6	80.9	95.1	0.76	88.8	91.2
	≥ 11	–	–	–	–	–	–	23.7	69.9	98.4	0.68	95.6	87.2
Possible dependence ³	≥ 6	60.9	100	50.1	0.51	36.1	100	–	–	–	–	–	–
	≥ 7	47.0	98.2	67.5	0.66	46.1	99.3	51.1	100	62.8	0.68	43.2	100
	≥ 8	37.6	92.8	78.0	0.71	54.4	97.5	44.5	100	71.2	0.71	49.6	100
	≥ 9	24.9	80.2	90.7	0.71	70.9	94.2	34.3	98.8	83.9	0.83	63.5	99.6
	≥ 10	17.5	61.1	94.8	0.56	76.7	89.6	29.6	95.8	89.2	0.85	71.4	98.7
	≥ 11	–	–	–	–	–	–	23.7	88.0	94.4	0.82	81.7	96.5
	≥ 12	–	–	–	–	–	–	19.1	76.7	97.1	0.74	88.3	93.6

¹AUDIT-10 scores ≥ 8 ²AUDIT-10 scores ≥ 16 ³AUDIT-10 scores ≥ 20

Youden’s J in bold. Cut-points with sensitivities or specificities below 50% are not displayed.

help due to their harmful drinking behavior, as discussed below.

It has been argued that sensitivity should be prioritized if the purpose is to detect individuals with severe alcohol problems (Baggio and Iglesias, 2020; Rodríguez-Martos and Santamarina, 2007). However, when it is considered imperative to capture virtually all cases of harmful drinking and possible dependence, some – or maybe many – will inevitably be false positives. It is equally obvious that the larger the occurrence of false positives, the less effective is the screener.

Regarding AUDIT-C, a cut-point of ≥ 6 was the highest whereby all possible alcohol dependent individuals and almost all (98%) with at least harmful drinking behavior were detected. This cut-point also had the highest Youden’s J for identifying a positive screen. Six in ten (61%) had scores of ≥ 6, of whom 49% were false positives with respect to harmful drinking or possible dependence.

The results for AUDIT-4 showed that a cut-point of ≥ 8 was the highest that identified all cases of possible alcohol dependence and almost all (97%) individuals with at least harmful drinking behavior. Less than half (45%) of the respondents had scores in this range. Of these, 29% were false positives regarding the two categories of unhealthy drinking.

Thus, for the purpose of identifying virtually all individuals with more severe alcohol problems, AUDIT-4 was clearly preferable. In contrast to AUDIT-C, our results also suggested that AUDIT-4 may work as a crude severity measure. Specifically, in addition to the cut-point of ≥ 8, a cut point of ≥ 6 was optimal (cf. Youden’s J) in identifying an AUDIT positive screen. If high sensitivity is considered less important for less severe alcohol problems, scores of 6 or 7 would thus be indicative of hazardous drinking.

4.3. Limitations

A serious limitation of our study is the lack of external validation criteria. Rumpf et al. (2002) examined the performance of the full and a short version of AUDIT against diagnostic measures in general populations samples and found that the full version performed somewhat better in detecting alcohol dependence. In prison populations and other groups with high prevalence of alcohol problems, it is particularly important to use a diagnosis of alcohol use disorder as reference standard.

The original AUDIT items in our study were modified to assess the year before incarceration. Especially for those who had been incarcerated quite a few months, the responses may have been hampered by

inaccurate recall – which is likely to strengthen a tendency to under-report one’s drinking (Greenfield and Kerr, 2008). The “true” AUDIT scores may thus have been higher than those observed, implying that the positive and the negative predictive values (cf. Table 4) are potentially misleading. On the other hand, all main results were replicated when we restricted the ROC analyses to individuals who had been imprisoned less than 3 months. Anyhow, it would have been advantageous if the respondents had been recruited consecutively a couple of weeks after entry into prison (cf. Maggia et al., 2004).

According to several validation studies of the AUDIT or its brief versions, the optimal cut-points are higher for males than for females (e.g. DeMartini and Carey, 2012; Levola and Aalto, 2015; Osaki et al., 2014; Reinert and Allen, 2007; Verhoog et al., 2020). The vast majority (93%) of the respondents in our sample were males, and the number of females was too low to examine such potential gender differences. Previous analyses of the full NorMA sample showed that the percentage distribution across the four AUDIT categories was gender invariant (Pape et al., 2020), yet the “best” cut-points on brief AUDIT screeners may still be gender specific. Thus, high AUDIT-C scores may be more strongly associated with alcohol-related harm among females, reflecting gender differences in alcohol metabolism and the quantity required to reach a high blood alcohol concentration (Thomasson, 2002). Hence, the cut-off scores that we suggested may not be recommendable for females.

Finally, it is possible that the responses to the brief AUDIT screeners would have been different if they had not been embedded in the full AUDIT.

4.4. Implications and suggestions for further research

The prison setting may potentially offer a golden opportunity to detect and treat individuals with harmful drinking behavior. However, according to MacAskill et al. (2011) “alcohol problems among prisoners are under-detected, under-recorded and under-treated”. The screening may be restricted to a yes/no-question (Parkes et al. 2011) or to few unvalidated items (Obstbaum et al., 2016). The failure to use standardized assessment tools is staggering, but three- or four-item screeners are feasible and probably more acceptable than more time-consuming alternatives. This, in turn, underscores the potential importance of our study.

The results indicated that AUDIT-4 may work as a graded severity measure. However, further assessment of individuals with a positive

screen (scores ≥ 6) is important to make well-founded decisions regarding adequate interventions. If limited resources or other obstacles preclude the possibility to do so, individuals whose AUDIT-4 scores are indicative of harmful drinking or alcohol dependence (scores ≥ 8) should be prioritized.

Our study suggested that AUDIT-C and AUDIT-4 are about equally useful as an initial filter whereby those with a positive screen are singled out and asked the remaining AUDIT questions. This is particularly feasible if the single-item scores are recorded electronically and added up successively. If such a two-step strategy is implemented, about 60% of the pre-prison drinkers should be asked all the ten AUDIT questions (cf. Table 4). The size of this group will be markedly smaller (48%) if AUDIT-4 is applied and the selection of individuals for further assessment is restricted to those with at least harmful drinking behavior (scores ≥ 8).

Finally, additional studies of the performance of brief AUDIT screeners in prison populations should be conducted. Such studies should address the issue of gender specific cut-points and employ external reference standards – including diagnostic assessment of alcohol use disorder. More knowledge is also warranted about acceptable modes of screening for alcohol problems, as perceived by those being screened as well as the prison staff. The extent to which short AUDIT versions may work as graded severity measures should also be pursued, both in studies of incarcerated persons and in other population groups.

4.5. Conclusions

AUDIT-C performed very well in detecting unhealthy pre-prison drinking behavior of varying severity. AUDIT-C and one additional item performed even better, of which AUDIT-4 was superior in detecting harmful drinking and possible dependence.

CRediT authorship contributions statement

Hilde Pape performed all the statistical analyses and wrote the first draft of the manuscript - mainly in co-operation with Ingeborg Rossow. Anne Bukten designed the NorMA-study, and prepared the files for statistical analyses. All authors were involved in the process of completing the paper.

Author disclosure

All authors have read and approved the resubmitted version of the manuscript. The article is our original work. It has not received prior publication and is not currently under consideration for publication elsewhere.

Conflict of interest

None.

Acknowledgements

The NorMA-study was funded by the South-Eastern Norway Regional Health Authority and The Norwegian Centre for Addiction Research (SERAF). This paper was supported by The University College of Norwegian Correctional Service and The Norwegian Institute of Public Health. The authors wish to thank all those who participated in the study, and the prison management and the helpful staff members at the local prison units for assisting in collecting data.

References

- Babor, T., Caetano, R., Casswell, S., Edwards, G., Giesbrecht, N., Graham, K., et al., 2010. *Alcohol: No Ordinary Commodity: Research and Public Policy*, Second ed. Oxford University Press, Oxford.
- Babor, T.F., de la Fuente, J.R., Saunders, J., Grant, M., 2001. *The Alcohol Use Disorders Identification Test: Guidelines for use in Primary Health Care*, Second ed. World Health Organization, Department of Mental Health and Substance Dependence, Geneva.
- Babor, T.F., Robaina, K., 2016. The Alcohol Use Disorders Identification Test (AUDIT): a review of graded severity algorithms and national adaptations. *Int. J. Alcohol Drug Res.* 5 (2), 17–24.
- Baggio, S., Iglesias, K., 2020. Commentary on Campbell and Strickland (2019): caution is needed when using self-reported alcohol use disorder screening tools. *Addict. Behav.* 100, 106115.
- Brady, M., Sibthorpe, B., Bailie, R., Ball, S., Sumnerdodd, P., 2002. The feasibility and acceptability of introducing brief intervention for alcohol misuse in an urban Aboriginal medical service. *Drug Alcohol Rev.* 21 (4), 375–380.
- Bukten, A., Lund, I.O., Rognli, E.B., Stavseth, M.R., Lobmaier, P., et al., 2015. The Norwegian offender mental health and addiction study—design and implementation of a national survey and prospective cohort study. In: *Substance Abuse: Research and Treatment*, 9. SART.
- Bukten, A., Stavseth, M.R., Skurtveit, S., Kunø, N., Lobmaier, P., et al., 2016. Rusmiddelbruk og helsesituasjon blant innsatte i norske fengsel. Resultater fra The Norwegian Offender Mental Health and Addiction Study (NorMA). Seraf, Oslo.
- Bush, K., Kivlahan, D.R., McDonell, M.B., Fihn, S.D., Bradley, K.A., 1998. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. *Arch. Intern. Med.* 158 (16), 1789–1795.
- Calabria, B., Clifford, A., Shakeshaft, A.P., Conigrave, K.M., Simpson, L., et al., 2014. Identifying Aboriginal-specific AUDIT-C and AUDIT-3 cutoff scores for at-risk, high-risk, and likely dependent drinkers using measures of agreement with the 10-item Alcohol Use Disorders Identification Test. *Addict. Sci. Clin. Pract.* 9 (1), 17.
- Caviness, C.M., Hatgis, C., Anderson, B.J., Rosengard, C., Kiene, S.M., et al., 2009. Three brief alcohol screens for detecting hazardous drinking in incarcerated women. *J. Stud. Alcohol Drugs* 70 (1), 50–54.
- Chang, Z., Lichtenstein, P., Larsson, H., Fazel, S., 2015. Substance use disorders, psychiatric disorders, and mortality after release from prison: a nationwide longitudinal cohort study. *Lancet Psychiatry* 2 (5), 422–430.
- Cook, R.L., Chung, T., Kelly, T.M., Clark, D.B., 2005. Alcohol screening in young persons attending a sexually transmitted disease clinic: comparison of AUDIT, CRAFFT, and CAGE instruments. *J. Gen. Intern. Med.* 20 (1), 1–6.
- Coulton, S., Newbury-Birch, D., Cassidy, P., Dale, V., Deluca, P., et al., 2012. Screening for alcohol use in criminal justice settings: an exploratory study. *Alcohol Alcohol.* 47 (4), 423–427.
- Cuparencu, C., Rinnan, Å., Silvestre, M.P., Poppitt, S.D., Raben, A., et al., 2020. The anserine to carnosine ratio: an excellent discriminator between white and red meats consumed by free-living overweight participants of the PREVIEW study. *Eur. J. Nutr.* 60, 179–192.
- DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L., 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 837–845.
- DeMartini, K.S., Carey, K.B., 2012. Optimizing the use of the AUDIT for alcohol screening in college students. *Psychol. Assess.* 24 (4), 954.
- Dowden, C., Brown, S.L., 2002. The role of substance abuse factors in predicting recidivism: a meta-analysis. *Psychol. Crime Law* 8 (3), 243–264.
- Evans, I., Heron, J., Murray, J., Hickman, M., Hammerton, G., 2021. The influence of alcohol consumption on fighting, shoplifting and vandalism in young adults. *Int. J. Environ. Res. Public Health* 18 (7), 3509.
- Fazel, S., Cartwright, J., Norman-Nott, A., Hawton, K., 2008. Suicide in prisoners: a systematic review of risk factors. *J. Clin. Psychiatry* 69 (11), 1721–1731.
- Fazel, S., Yoon, I.A., Hayes, A.J., 2017. Substance use disorders in prisoners: an updated systematic review and meta-regression analysis in recently incarcerated men and women. *Addiction* 112 (10), 1725–1739.
- Graham, K., West, P., 2001. Alcohol and crime: examining the link. In: Heather, N., Peters, T.J., Stockwell, T. (Eds.), *International Handbook of Alcohol Dependence and Problems*. Wiley and Sons Ltd, New York, pp. 439–470.
- Greenfield, T.K., Kerr, W.C., 2008. Alcohol measurement methodology in epidemiology: recent advances and opportunities. *Addiction* 103 (7), 1082–1099.
- Gual, A., Segura, L., Contel, M., Heather, N., Colom, J., 2002. Audit-3 and audit-4: effectiveness of two short forms of the alcohol use disorders identification test. *Alcohol Alcohol.* 37 (6), 591–596.
- Hanley, J.A., 1989. Receiver operating characteristic (ROC) methodology: the state of the art. *Crit. Rev. Diagn. Imaging* 29 (3), 307–335.
- Higgins-Biddle, J.C., Babor, T.F., 2018. A review of the Alcohol Use Disorders Identification Test (AUDIT), AUDIT-C, and USAUDIT for screening in the United States: past issues and future directions. *Am. J. Drug Alcohol Abuse* 44 (6), 578–586.
- Kriston, L., Hölzel, L., Weiser, A.-K., Berner, M.M., Härter, M., 2008. Meta-analysis: are 3 questions enough to detect unhealthy alcohol use? *Ann. Intern. Med.* 149 (12), 879–888.
- Lee, J.H., Kong, K.A., Lee, D.H., Choi, Y.H., Jung, K.Y., 2018. Validation and proposal for cut-off values of an abbreviated version of the Alcohol Use Disorder Identification Test using the Korean National Health and Nutrition Examination survey. *Clin. Exp. Emerg. Med.* 5 (2), 113–119.
- Levola, J., Aalto, M., 2015. Screening for at-risk drinking in a population reporting symptoms of depression: a validation of the AUDIT, AUDIT-C, and AUDIT-3. *Alcohol Clin. Exp. Res.* 39 (7), 1186–1192.

- Maggia, B., Martin, S., Crouzet, C., Richard, P., Wagner, P., et al., 2004. Variation in audit (alcohol used disorder identification test) scores within the first weeks of imprisonment. *Alcohol Alcohol*. 39 (3), 247–250.
- MacAskill, S., Parkes, T., Brooks, O., Graham, L., McAuley, A., et al., 2011. Assessment of alcohol problems using AUDIT in a prison setting: more than an 'aye or no' question. *BMC Public Health* 11 (1), 1–12.
- Meneses-Gaya, C., Zuardi, A.W., Loureiro, S.R., Crippa, J.A.S., 2009. Alcohol Use Disorders Identification Test (AUDIT): an updated systematic review of psychometric properties. *Psychol. Neurosci* 2, 83–97.
- Meneses-Gaya, C., Zuardi, A.W., Loureiro, S.R., Hallak, J.E., Trzesniak, C., et al., 2010. Is the full version of the AUDIT really necessary? Study of the validity and internal construct of its abbreviated versions. *Alcohol Clin. Exp. Res.* 34 (8), 1417–1424.
- Morojele, N.K., Nkosi, S., Kekwaletswe, C.T., Shuper, P.A., Manda, S.O., et al., 2017. Utility of brief versions of the Alcohol Use Disorders Identification Test (AUDIT) to identify excessive drinking among patients in HIV care in South Africa. *J. Stud. Alcohol Drugs* 78 (1), 88–96.
- Nehlin, C., Fredriksson, A., Jansson, L., 2012. Brief alcohol screening in a clinical psychiatric population: special attention needed. *Drug Alcohol Rev.* 31 (4), 538–543.
- Neumann, T., Linnen, H., Kip, M., Grittner, U., Weiß-Gerlach, E., et al., 2012. Does the Alcohol Use Disorders Identification Test—consumption identify the same patient population as the full 10-item Alcohol Use Disorders Identification Test? *J. Subst. Abuse Treat.* 43 (1), 80–85.
- Newbury-Birch, D., McGovern, R., Birch, J., O'Neill, G., Kaner, H., et al., 2016. A rapid systematic review of what we know about alcohol use disorders and brief interventions in the criminal justice system. *Int. J. Prison Health* 12 (1), 57–70.
- Osaki, Y., Ino, A., Matsushita, S., Higuchi, S., Kondo, Y., et al., 2014. Reliability and validity of the alcohol use disorders identification test—consumption in screening for adults with alcohol use disorders and risky drinking in Japan. *A. Pac. J. Cancer Prev.* 15 (16), 6571–6574.
- Obstbaum, Y., Tyni, S., Mattila, A.K., Vartiainen, H., et al., 2016. Not all substance dependence problems are recognized as risks—comparing a medical health study with prison assessments. *Eur. J. Crim. Policy Res.* 22 (1), 189–210.
- Pape, H., Rossow, I., Bukten, A., 2020. Alcohol problems among prisoners: subgroup variations, concurrent drug problems, and treatment needs. *Eur. Addict. Res.* 1–10.
- Parkes, T., MacAskill, S., Brooks, O., Jepson, R., Atherton, L., McGhee, S., et al., 2011. *Prison Health Needs Assessment for Alcohol Problems*. NHS Health, Scotland (Available at). (<http://www.healthscotland.com/documents/4932.aspx>).
- Rehm, J., Gmel Sr, G.E., Gmel, G., Hasan, O.S., Imtiaz, S., et al., 2017. The relationship between different dimensions of alcohol use and the burden of disease—an update. *Addiction* 112 (6), 968–1001.
- Reinert, D.F., Allen, J.P., 2007. The alcohol use disorders identification test: an update of research findings. *Alcohol. Clin. Exp. Res.* 31 (2), 185–199.
- Rodríguez-Martos, A., Santamariña, E., 2007. Does the short form of the Alcohol Use Disorders Identification Test (AUDIT-C) work at a trauma emergency department? *Subst. Use Misuse* 42 (6), 923–932.
- Rossow, I., Bye, E.K., 2013. The problem of alcohol-related violence: an epidemiological and public health perspective. In: McMurrin, M. (Ed.), *Alcohol-Related Violence: Prevention and Treatment*. Wiley-Blackwell, Chichester, pp. 3–18.
- Rumpf, H.J., Hapke, U., Meyer, C., John, U., 2002. Screening for alcohol use disorders and at-risk drinking in the general population: psychometric performance of three questionnaires. *Alcohol Alcohol*. 37 (3), 261–268.
- Saunders, J.B., Aasland, O.G., Babor, T.F., De la Fuente, J.R., Grant, M., 1993. Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction* 88 (6), 791–804.
- Seal, D.W., Yancey, S., Reddy, M., Kinner, S.A., 2018. Alcohol use among incarcerated individuals. In: Stuart, A., Rich, J.D.J. (Eds.), *Drug Use in Prisoners: Epidemiology, Implications, and Policy Responses*. Oxford University Press, New York, pp. 57–70.
- Seth, P., Glenshaw, M., Sabatier, J.H., Adams, R., Du Preez, V., et al., 2015. AUDIT, AUDIT-C, and AUDIT-3: drinking patterns and screening for harmful, hazardous and dependent drinking in Katutura, Namibia. *PLoS One* 10 (3), e0120850.
- Thomas, E., Degenhardt, L., Alati, R., Kinner, S., 2014. Predictive validity of the AUDIT for hazardous alcohol consumption in recently released prisoners. *Drug Alcohol Depend.* 134, 322–329.
- Thomasson, H.R., 2002. Gender differences in alcohol metabolism. In: Galanter, M. (Ed.), *Recent Developments in Alcoholism*, vol 12. Springer, Boston. (https://doi.org/10.1007/0-306-47138-8_9).
- Toner, P., Böhnke, J.R., Andersen, P., McCambridge, J., 2019. Alcohol screening and assessment measures for young people: a systematic review and meta-analysis of validation studies. *Drug Alcohol Depend.* 202, 39–49.
- Towers, A., Stephens, C., Dulin, P., Kostick, M., Noone, J., et al., 2011. Estimating older hazardous and binge drinking prevalence using AUDIT-C and AUDIT-3 thresholds specific to older adults. *Drug Alcohol Depend.* 117 (2–3), 211–218.
- Verhoog, S., Dopmeijer, J.M., de Jonge, J.M., van der Heijde, C.M., Vonk, P., et al., 2020. The use of the alcohol use disorders identification test—consumption as an indicator of hazardous alcohol use among university students. *Eur. Addict. Res.* 26 (1), 1–9.
- WHO, 2019a. *WHO. Health in Prisons: Fact Sheet for 38 European Countries*. WHO Regional office for Europe, Copenhagen (2019. Available at). (<http://www.euro.who.int/en/health-topics/health-determinants/prisons-and-health/publications/2019/health-in-prisons-fact-sheets-for-38-european-countries-2019>).
- WHO, 2019b. *Screening for Harmful Use of Alcohol. Health in Prisons European Database 2019* (Available at). (http://apps.who.int/gho/data/node.prison.Screening_Harmful_Use_Alcohol?lang=en).
- Wu, S.-I., Huang, H.-C., Liu, S.-I., Huang, C.-R., Sun, F.-J., et al., 2008. Validation and comparison of alcohol-screening instruments for identifying hazardous drinking in hospitalized patients in Taiwan. *Alcohol Alcohol*. 43 (5), 577–582.
- Youden, W.J., 1950. Index for rating diagnostic tests. *Cancer* 3 (1), 32–35.