Concurrent validity of an Estimator of Weekly Alcohol Consumption (EWAC) based on the Extended AUDIT

Contents

1 Introduction

Globally, alcohol consumption is responsible for 5% of all deaths and disease burden [1]. The burden of alcohol consumption goes far beyond the burden of *alcohol use disorders*, defined in the International Statistical Classification of Diseases (ICD-10 F10.1/F10.2; ICD-11 QE10/6C40.1).

A number of strategies exist in order to either prevent [2], or treat and reduce [3] harm from alcohol consumption. Some recommend systematic screening of alcohol use disorders using a validated diagnostic tool. More recently, others have advocated monitoring alcohol consumption like many other risk factors (eg body mass index or cholesterol), making individuals more conscious of their current intake and associated risk [4, 5]. This approach is intended to enable timely lifestyle interventions (and, if unsuccessful, pharmacological interventions), in order to prevent the emergence of alcohol use disorders. Simultaneously, public health

recommendations to the public [6, 7] now promote the reduction of alcohol consumption, in the absence of a 'safe' level below which alcohol does not increase disease risk [8].

In contrast with alcohol use disorder screening tools, alcohol consumption assessment tools are not well embedded in clinical practice. Potential reasons include: lack of diagnostic validation, difficulty to self-assess alcohol consumption, and time pressures. This creates a gap in the range of resources destined to help individuals understand, monitor and take control of their alcohol consumption with–or without–the involvement of primary care physicians.

The objective of this paper is to develop and validate a fast and easy-to-complete Estimator of Weekly Alcohol Consumption (EWAC) by repurposing an existing screening tool. In many countries, the most widely used screening tool is the Alcohol Use Disorders Identification Test (AUDIT) [9], a 10-item questionnaire schedule widely employed in clinical practice and clinical research as a diagnostic test for alcohol use disorder. The short, 3-item consumption version known as AUDIT-C has good diagnostics accuracy for both: (a) interview-based clinical diagnoses; and (b) consumption in excess of maximum recommended intakes (eg 140g/week in Australia, 112 g/week in UK) in a variety of populations [10].

The present study aims to validate the AUDIT-C against continuous measures of alcohol consumption, as opposed to against a particular clinical diagnosis or consumption threshold. To this end, we adopt a variant of the AUDIT-C schedule known as the 'Extended AUDIT-C', which improves the granularity of the information collected with a greater range of response options on quantity and frequency (AUDIT items 1 and 2). The Extended AUDIT-C has been used in UK research as part of two trials [11, 12] and one continuous household survey [13] to capture greater information on the higher risk drinkers, based on the observation that AUDIT consumption items are right-truncated.

This paper is made up of three investigations. Study 1 estimates coefficients to apply to each of the AUDIT-C response items to compute an EWAC using data from a large English household survey, the Alcohol Toolkit Study (ATS). It then tests the concurrent validity of the EWAC against Graduated-Frequency (GF) across a number of demographic subpopulations in England. Study 2 tests the concurrent validity of the EWAC against the 28-day Timeline Followback (TLFB) in a clinical population (hospital clinic outpatients). Study 3 compares the population-wide total and empirical cumulative distribution of alcohol consumption in England using: the EWAC, the ATS graduated frequency schedule, the Health Survey for England, and alcohol sales official statistics.

2 Methods

2.1 Approach

This paper develops and validates an Estimator of usual Weekly Alcohol Consumption in units (EWAC) based on the Extended AUDIT-C. Neither the AUDIT-C nor the Extended AUDIT-C provide a measure of usual alcohol consumption, but the product of frequency of drinking (AUDIT item 1) and quantity of drinking (AUDIT item 2) can be used to estimate usual alcohol consumption, with adjustment for occasional heavy use (AUDIT item 3), following methods developed for quantity-frequency-variability instruments [14].

In practice, for every individual i, the EWAC is computed as the product of F_i and Q_i (AUDIT items 1 and 2 respectively) adjusted with the frequency of intense drinking V_i (AUDIT item 3):

$$\text{EWAC}_i = F_i Q_i + V_i b$$

where b denotes the average number of units of alcohol consumed in an intense drinking day. Coefficients F, Q, V and b are unknown. In this study, two sets of potential coefficients are evaluated:

- using the AUDIT response item interval midpoint: for example, 2.5 for '2 to 3 times per week' in AUDIT item 2
- using a statistical model to estimate the coefficients directly from gold standard data.

All results are reported in UK alcohol units (8g or 10mL of pure alcohol). Analyses are conducted in `R` [15] using packages `tidyverse` and `rstan` [16, 17]. Computer scripts for all analyses are available on an online repository [18].

2.2 Data sources

A longstanding obstacle in alcohol research and care is the absence of a diagnostic gold standard, which is not addressed even by the development of new biomarkers. Instead, a number of instruments exist which measure self-reported alcohol consumption with varying validity and reliability. A state-of-the art review by [19] is summarised in Table 1 below. Prospective diaries tend to record higher alcohol consumption by minimising recall bias, followed by GF, while lower levels seem to be recorded with QF measures [20, 21].

Table 1: Alcohol schedules: selective comparison

Schedule	Bias	Variance	Measures variability
Graduated frequency (GF)	Unclear	Low	Yes
Quantity-Frequency (QF)	Unclear	Low	No
Quantity-Frequency-Variability (QFV)	Unclear	Low	Yes
Yesterday' recall	Minimal	High	No
Timeline followback (TLFB)	Low	Low	Yes
Prospective diary	Low	Low	Yes

The present paper employs four sources of data presented in Table 2 below, together with references to methodological descriptions.

Table 2: Description of data sources

Source	Time	Study population	Measurements	Validation	Sample size	Design
Alcohol Toolkit Study (ATS) [13]	November 2015–October 2017 (waves 110–133)	Residents of private English households aged 16 years and over	Interviewer-led Extended AUDIT-C and GF	[20], [22]	40,832	Cross-sectional sample survey
Health Survey for England (HSE) [23]	2011	Residents of private English households aged 16 years and over	Interviewer-led beverage-specific QF	Consistent with a prospective diary [24] and yesterday recall [25].	8,610	Cross-sectional sample survey

Source	Time	Study population	Measurements	Validation	Sample size	Design
Hospital questionnaire study (HOSP)	February 2019–October 2019	Outpatient and day case hospital patients aged 18 years and over	Interviewer/self-administered Extended AUDIT-C and interviewer-led TLFB	TLFB records fewer drinking days and thus lower consumption than a prospective diary [26]. This recall bias increases with the number of days elapsed [27, 28]	130	Cross-sectional study with block randomisation to (a) self-administered AUDIT or (b) researcher-administered AUDIT.
Alcohol retail sales [29]	2014	English population aged 18 years and over	–	[29]	–	Ratio of all alcohol produced or processed in the UK, as well as alcohol imported into the UK for sale and consumption, over the mid-year population estimate

2.3 Study 1: EWAC estimation and concurrent validity

Study 1 evaluates the accuracy of the EWAC derived from a researcher-administered Extended AUDIT-C in community households. A pre-registered protocol for this study is available online [30]. It uses the GF schedule from the ATS as a gold standard. Two sets of coefficients are compared: AUDIT response item interval midpoint; and coefficients estimated using a statistical model (supplementary material 1-2).

First, the agreement between the EWAC and the GF is quantified by studying two types of deviations:

- bias is estimated using the mean deviation to the gold standard $\text{MD} = n^{-1} \sum_{i=1}^{n} (\text{EWAC}_\text{i} - \text{GF}_\text{i})$. We test the hypothesis that the MD is greater than 1 UK unit using a two-sided t-test.
- precision is estimated using the root mean squared deviation $\text{RMSD} = \sqrt{n^{-1} \sum_{i=1}^{n} (\text{EWAC}_\text{i} - \text{GF}_\text{i})^2}$, a measure of total error: it capture both bias and random deviation from the assumed gold standard. For example, an RMSD of 2 signifies that the EWAC is on average with $\pm$ 2 UK units of the gold standard. We test the hypothesis that the RMSD is greater than 2 UK units using a one-sided χ^2 homogeneity test.

Second, we examine whether the EWAC's validity varies across population subgroups.

- the simple deviation (EWAC $-$ GF) is regressed in a linear model to test for subgroup differences in MD

- the squared deviation $(\text{EWAC} - \text{GF})^2$ is regressed in a log-transformed linear model to test subgroup differences in MSD. Model coefficients are then transformed (square root of the exponential) into relative RMSD estimates interpreted as the ratio of the subgroup RMSD to the reference category RMSD, a ratio >1 indicating worse precision than in the reference category.

Both models include the following predictors: sex by age group; ethnic group; highest educational qualification; religion; smoking status. Additional models are fitted solely in respondents with an AUDIT-C score of 5 or more or an AUDIT score of 8 or more (under the ICD terminology, hazardous and harmful alcohol use), using extra variables measures during interview: favourite drink (beer; wine; spirits alone; mixed spirits; cider; other); and whether the respondent attempt to restrict alcohol intake in the last 12 months (eg by drinking less, choosing lower strength alcohol or using smaller glasses).

Third, we test whether the EWAC is superior to the traditional AUDIT-C score at predicting drinking in excess of 14 UK units per week. In the UK, a score of 5 or more is categorised as 'increasing risk' or 'higher risk', and regarded as indicating consumption in excess of 14 units per week. The study tests the hypothesis that the EWAC's receiver operating characteristics' area under the curve (AUC) is greater that the AUDIT-C score using a nonparametric paired AUC test [31].

The ATS data (n = 40,832) is affected by missing data: 35% of respondents (n = 14,408) reported never drinking alcohol in AUDIT item 1 and were not asked any further AUDIT or GF questions. These participants are excluded from the analysis. A further 4,020 respondents (0.2% of those reporting drinking in AUDIT item 1) did not have a valid GF alcohol consumption record and were also excluded. In total, 22,404 valid observations remain for the diagnostic analysis, in which missing GF data is assumed to be missing at random conditionally on the Extended AUDIT-C responses. In the subgroup analysis (Study 2 below), a further 530 repondents (0.5%) were assumed to have data missing at random and were excluded.

2.4 Study 2: Concurrent validity in hospital outpatients

Study 2 aims to confirm the robustness of findings in (a) a clinical population; (b) when the Extended AUDIT-C is self-administered; (c) using a different gold standard: the 28-day TLFB. Participants were recruited from a range of clinics at a large acute hospital in Southampton, UK: orthopaedics, rheumatology, young adults, managed care and mental health outpatients, as well as endoscopy day cases. A total of 130 participants consented to participating, and block-randomised to one of two groups:

(1) self-administered Extended AUDIT-C (n=59)
(2) researcher-administered Extended AUDIT-C (n=71).

Once this first questionnaire completed, both groups were asked to complete a 28-day alcohol TFLB questionnaire administered by the researcher.

To model the relationship between the EWAC and the TLFB, the number of units consumed on any day is assumed to follow a negative binomial distribution, the rate of which is determined by the latent usual alcohol consumption as well as the following variables: AUDIT mode of administration (self-administered; researcher-administered), week day, number of days elapsed since (1-7; 8-14; 15-21; 22-28). TLFB daily number of units consumed were rounded to the nearest integer and regressed in a negative binomial regression model against the aforementioned predictors, as well as the EWAC divided by 7 to obtain its daily equivalent. The corresponding regression coefficient measures the ratio of TLFB to EWAC.

2.5 Study 3: Aggregate concurrent validity

Study 3 examines the consistency in aggregate alcohol consumption estimates across England in residents aged 18 years and over. We plot plot the empirical cumulative distributions of alcohol consumption given by (1) the EWAC estimated from the ATS; (2) the quantity-frequency estimated in the ATS; (3) the beverage-specific estimators in HSE in 2011; (4) the prospective diary estimator in HSE 2011. In this analysis, survey weights are used: in (1-3), poststratification weights estimated using calibration and age-sex MYPES; in (4), similar postratification weights adjusted for self-selection into participation to the prospective diary data collection.

We report the percentage of total alcohol sales for England accounted for by each method, using both on-trade and off-trade 2014 sales estimates for England from [29].

3 Results

3.1 Study 1: EWAC estimation and validation

3.1.1 Overall bias and accuracy

The first step involved choosing a set of coefficients to compute the EWAC (supplementary materials 1–2).

The EWAC computed with the midpoint of the AUDIT item intervals has a Pearson's correlation $r = 0.69$, [95% CI: 0.69, 0.70]. It produces a mean deviation (MD) of 0.71 UK alcohol units/week, indicating a bias inferior to the preregistered ± 1-unit bias allowance ($p = 1.000$). The root mean squared deviation (RMSD) estimate of 12.3 units [95% CI: 11.2;13.2] is significantly greater than the pre-registered 2-unit total error allowance ($p < 0.001$), suggesting that the EWAC falls on average 12 units away from the GF gold standard.

Coefficients estimated empirically (statistical model reported in supplementary materials 2) provide a small improvement: $r = 0.71$ [0.71, 0.72] (Kendall's rank correlation $\tau = .63$) and MD = 0.18 ($p = 1.000$). With RMSD = 10.9 [95% CI: 9.8;12.0], precision remains statistically significantly greater than 2 ($p < 0.001$).

The RMSD masks a dispersed and skewed distribution of error. Table 3 shows that, for 50% of participants, the EWAC falls within ± 2 UK units of the GF weekly consumption estimate. In other terms, an interval estimate defined as the EWAC ± 2 units (eg '10 to 14 units') would cover the gold standard for half of individuals, while an interval estimate defined as the EWAC ± 3 units (eg '9 to 15 units') would cover the gold standard for 60% of individuals.

Table 3: Percentiles of the absolute deviation between EWAC and GF schedule (n = 22,404)

10%	20%	30%	40%	50%	60%	70%	80%	90%	95%	99%
0.4	0.7	1.0	1.5	2.1	3.0	4.2	6.2	10.6	17.0	38.7

Figure 1 compares individual EWAC and GF values. We note the departure of lines of best fit from the diagonal, demonstrating the EWAC's small positive bias (MD > 0) is not consistent. The plots indicate a slight positive bias for consumptions up to 10-14 units/week, then a slight negative bias above this threshold.

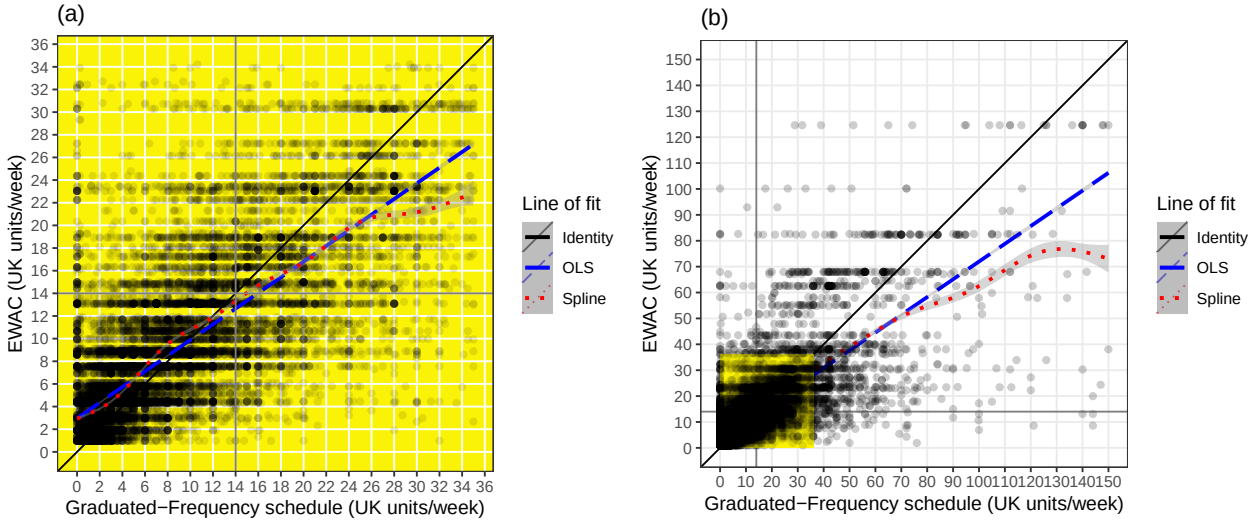


Figure 1: Plots of EWAC against GF in (a) low/increasing risk ATS respondents (n=21,338) and (b) all ATS respondents (n=22,373)

3.1.2 Subgroup accuracy

Next, the MD and RMSD are regressed against respondent characteristics in order to identify subgroups with heterogeneous bias or precision (supplementary material 3, Table 7). The reference category's key characteristics are: females aged between 25 and 34 years of White ethnicity without educational qualifications, who never smoked. The model predictors explain a very modest proportion of both MD and RMSD (R^2 statistics $< 2\%$). Nevertheless, specific subgroups do exhibit very different MD and RMSD.

Figure 7 summarises the MD for a selection of subgroups whose predicted MD was either above 1 or below -1; and whose coefficients had a p-value below 0.05. Respondents of Black, Other, and White Other ethnic groups had significantly overestimated EWACs: their MDs were respectively 4.8 units [95% CI: 2.1, 7.5]; 5.9 units [1.6, 10.1] and 1.6 [0.2, 3.0] in excess of the reference MD. The MDs of respondents aged 55 to 64 years or 75 years and over respectively had MDs 2.2 units [0.5; 3.9] and 4.2 units [0.9; 7.6] in excess of the reference MD. Similar results were found in increasing risk drinkers, without significant evidence of an effect of favourite drink or attempts to reduce alcohol intake in the past year (supplementary material 3, Table 8)

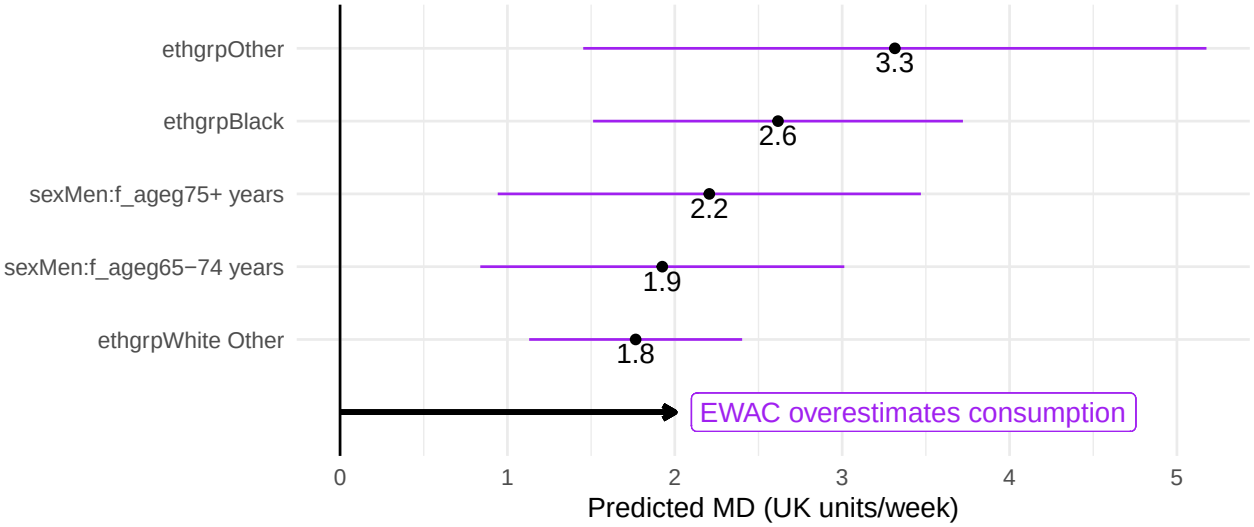


Figure 2: Forest plot of modelled MD for selected subgroups

Figure 3 shows those subgroups with an RMSD found to be significantly different from the rest of the population ($p < 0.05$); and estimated to be 20% greater or smaller than the RMSD of the reference category. This shows that RMSD is 58% [95% CI: 50; 67%] greater in current smokers, 34% [14; 56%] greater in respondents who stopped smoking in the past year, and 23% greater [17; 30%] in respondents who stopped smoking over a year ago. It is also 44% greater [29; 60%] greater in men and 34% greater [19; 50%] in respondents aged 16 to 24 years. Conversely, error is 35 to 70% smaller in White Other, Black and Other ethnic groups.

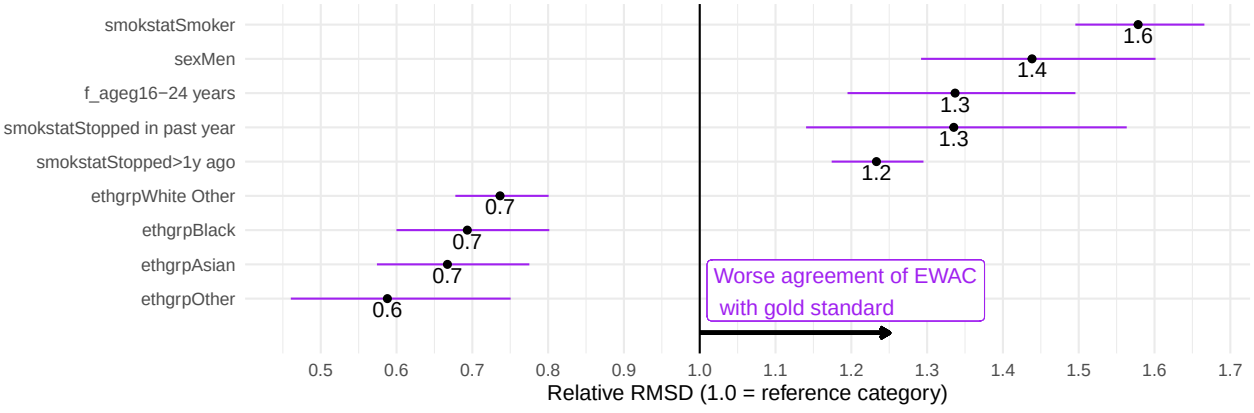


Figure 3: Forest plot of RMSD ratio (selected subgroups to reference category)

Figure 4 visualises the same analysis in increasing risk drinkers exclusively (AUDIT-C $\geq$ 5 or AUDIT $\geq$ 8). With a higher mean alcohol consumption, results differ from the overall picture presented in Figure 3. The RMSD of respondents favouring mixed spirits had an RMSD 23% [95% CI: 9.6; 38] smaller than the reference category. Educational qualifications seem to significantly improve the agreement between EWAC and the gold standard. School and degree-level qualifications reduced RMSD by 24% [12; 37%] and 37% [23; 52%] smaller respectively with unchanged MD, suggesting that respondents may have better recall and clarity over alcohol beverage content. Conversely, the RMSD of respondents who attempted to reduce their alcohol consumption was 23% [16; 30%] larger than the reference category.

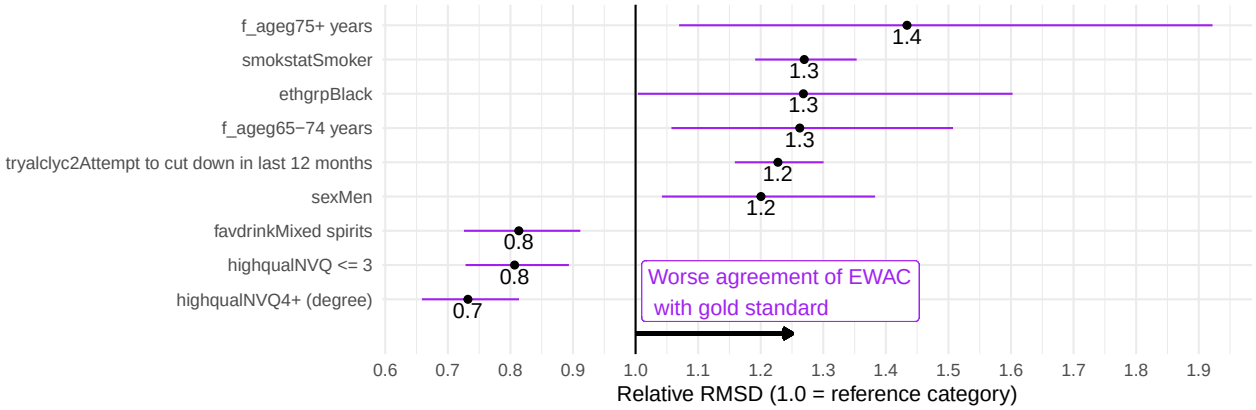


Figure 4: Forest plot of RMSD ratio (selected subgroups to reference category) in respondents with a hazardous/harmful alcohol use (AUDIT-C>=5 or AUDIT>=8

3.1.3 Receiver Operating Characteristics

The last step of this analysis was to examine the EWAC's ability to classify participants' consumption as being equal or in excess of 14 UK units/week. In comparison with the AUDIT-C score (ranging 0-12), EWAC increases the area under the curve by 5 percentage points (Table 4), a statistically significant improvement ($p < 0.001$). ROC curve are available from Supplementary Material 3. The cut-off that maximises the sum of specificity and sensitivity is 9.95 units/week. In constrast, using the nominal cut-off of 14 units/week yields a sensitivity of 0.705 and a specificity of 0.922.

In summary, using an EWAC threshold of 10 units or more over an AUDIT-C score of 5 or more provides equivalent sensitivity, but with an a steep gain in specificity of 13 percentage points.

Table 4: Receiver operating characteristics of AUDIT-C score and EWAC for consumption >= 14 UK units/week (n = 22,404)

Metric	AUC	95% CI	Best threshold	Sensitivity	Specificity
AUDIT-C score	0.871	[0.866; 0.876]	4.50	0.882	0.684
EWAC	0.921	[0.917; 0.925]	9.95	0.876	0.816

Note: The best threshold refers the cut-off value that maximises the sum of sensitivity and specificity.

3.2 Study 2: Concurrent validity in hospital patients

A total of 105 participants (81%) completed the Extended AUDIT-C, of whom 63 were classified as low-risk drinkers, 25 as increasing-risk alcohol users (AUDIT-C score of 5 to 7), and 17 as high-risk alcohol users (AUDIT-C score of 8 or more). A total of 13 participants did not provide TLFB information for at least one day missing out of 28, resulting in 5% (137/2940) of TLFB days were missing.

Using only 103 participants with a minimum of 7 days recorded on the TLFB, MD is estimated at -0.5 unit/week, which does not provide any evidence of bias greater than 1 unit/week ($p = 0.316$). As for error,

RMSD is estimated at 5.4 [95%CI: 3.5; 6.8] and is statistically significantly greater than 2 units ($p < 0.001$). A potential reason for RMSD being considerably smaller than in Study 1 is the distribution of alcohol consumption of this small pilot dataset. Respondents' alcohol intake as estimated by TLFB was low, with a mean of 7 units/week; a median of 2 units/week, and a 90th percentile of 23 units/week. Consequently, the probability of observing the strong deviations that can be observed in presence of high alcohol consumption data was low.

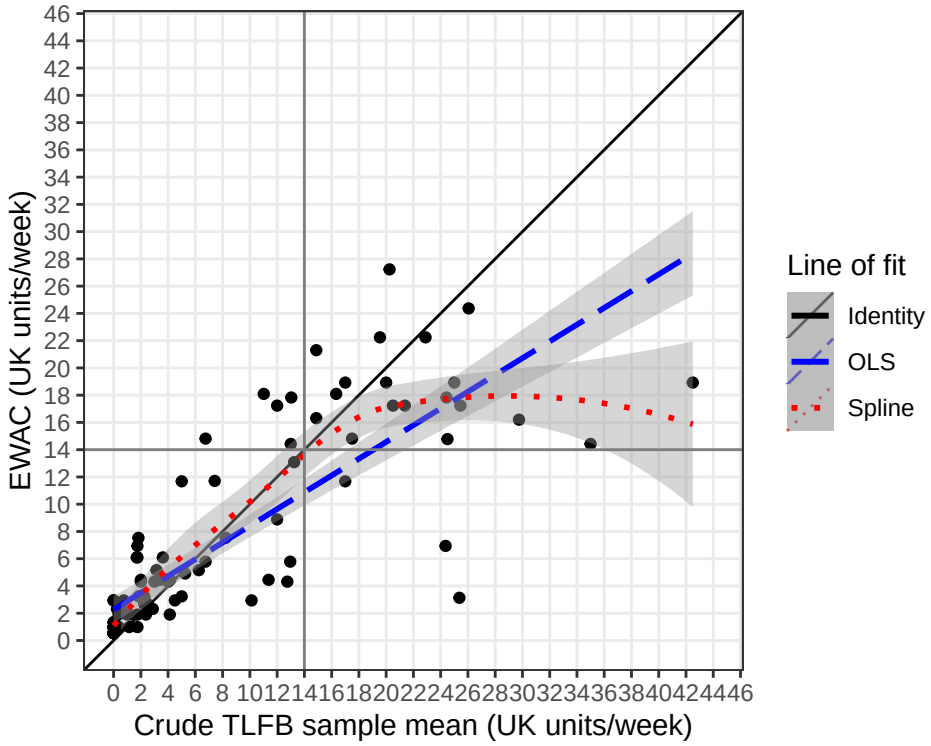


Figure 5: Plot of EWAC estimates against TLFB estimates in hospital participants with lines of fit

3.3 Study 3: Empirical distribution functions

Table 6 provides estimates of total alcohol consumption in adult resident of private households in England using four different estimators, and compares them with alcohol retail sales. The HSE schedules provide the highest estimates of alcohol consumption and coverage of sale statistics. The proposed EWAC estimates of total consumption are just 71% of a very reliable estimate, the HSE prospective diary.

Table 6: Summary statistics on alcohol consumption in England in residents aged 18 years and over (excluding abstainers)

Study	Mean (units/week)	Median	Variance	N	% of alcohol sold
HSE beverage-specific QF	14.0	7.3	474.6	6,545	72.6
HSE prospective diary	13.0	8.0	264.7	4,640	67.6
ATS GF	8.5	5.1	242.0	22,136	44.0
ATS EWAC	9.3	5.2	148.9	25,882	48.0
Retail sales	19.3	–	–	–	–

Figure 6 suggests that the EWAC, like the ATS GF it was estimated against, estimates a higher prevalence of low risk consumption ($\leq$ 14 units/week) and increased risk consumption than HSE. In contrast very high alcohol consumption ($\geq$ 50 units/week) is higher in HSE. This may be due to a combination of difference in sampling coverage, nonresponse bias, or measurement error in the alcohol schedules.

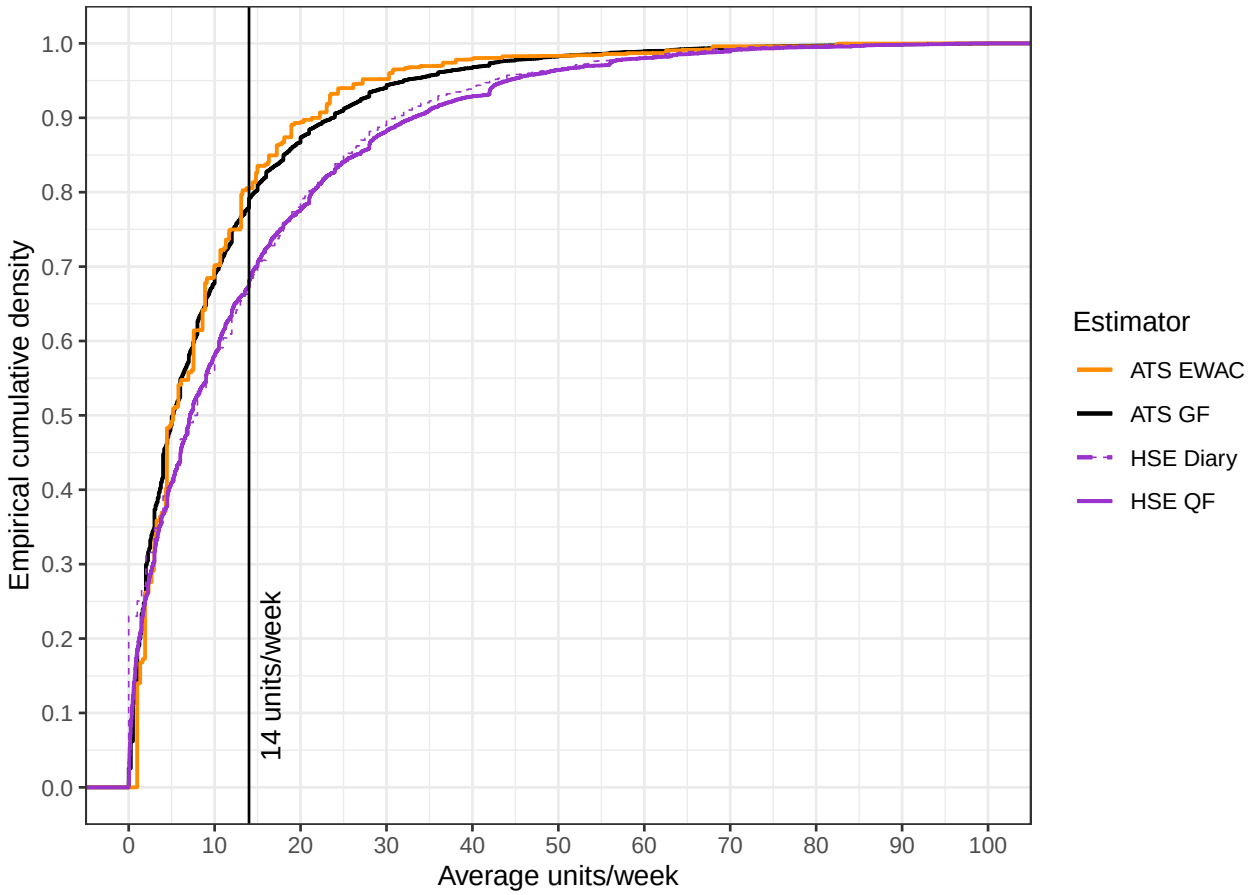


Figure 6: Empirical cumulative distribution function of weekly alcohol consumption in England according to four alcohol schedules in residents aged 18 years and over

4 Discussion

4.1 Main findings

This paper examined the predictive capabilities of the Extended AUDIT in assessing alcohol consumption in two English populations: community dwelling household residents, and hospital clinic outpatients. The Extended AUDIT is a variant of the AUDIT-C containing a choice of 6 response items to quantify drinking frequency, and 7 reponse items to quantify the average quantity consumed during any drinking day. The resulting EWAC estimates usual alcohol consumption with a mean precision of $\pm$ 11 units/week when compared to the GF alcohol schedule, or $\pm$ 5 units/week when compared to a 28-day TLFB in a sample of lower-risk alcohol users. Although bias is mostly consistent across subgroups examined (age/sex, education, smoking status, religion), there is strong evidence that EWAC overestimates alcohol consumption by 2-3 units/week in Black and Other ethnic groups. At the same time, average precision is better in Black, Asian and Other ethnic groups. This is likely due to a lower mean alcohol consumption than White British and Mixed ethnic groups. We also noted a weaker precision for both current and ex-smokers.

4.2 Strengths and limitations

To the best of our knowledge, this study was the first to (a) develop an EWAC using a well-accepted clinical alcohol use disorders identification tool such as the AUDIT; and (b) quantify its bias and precision with respect to a continuous measure of alcohol consumption. One study by [32] previously reported mean consumption by AUDIT-C score, yet without quantifying bias or precision of such a measure. Other studies have studied

the AUDIT-C's potential in estimating alcohol consumption, but only in relation to predicting consumption in excess of a predefined threshold [10]. Such studies achieved AUCs ranging .83-.96. Study 1 demonstrated the EWAC's superiority for predicting GF $\geq$ 14 units/week. While its AUC of 0.921 [95% CI: 0.917; 0.925] is comparable with many other studies, the specificity gain from .684 to .876 places the EWAC in line with the best-performing validations.

This study should provide strong confidence in the internal and external validity of findings in England on account of the large sample size of the ATS, the range of subgroup analyses reported, and replication of the analysis in a clinical population using the TLFB method (Study 2).

We note some limitations. Findings reported in this paper may not apply to other countries, or to small subpopulations in England with an atypical alcohol consumption, such as patients seeking care for conditions such as addition or alcohol liver disease. Alcohol survey data quality issues also raise questions of internal validity in Study 1.

The gold standard has a strong influence on estimates of precision (RMSD). Error in the gold standard will entirely propagate to the RMSD measure (unless this error correlates with error in the EWAC). This means some of the RMSD may in fact be due to error in the gold standard rather than the EWAC. In a telephone interview study on 119 participants in San Fransisco, US, [22] measured a Pearson's correlation coefficient of $r = .86$ and .87 in drinking frequencies and volumes measured by a GF interview followed by a 28 day prospective diary. In a telephone interview study of 773 participants across Switzerland, [20] measure a Kendall's rank correlation $\tau = 0.41$ between GF and a 7-day prospective diary. For reference, our study measured a correlation coefficient $r = .71$ and $\tau = .63$ between EWAC and GF. Our measures of RMSD are thus likely to be very conservative estimators of precision.

The ATS, in spite of a large sample size, may suffer from coverage issues. Like the HSE, ATS does not cover populations excluded from most sampling frames, such as residents of communal and carceral institutions, homeless people, or migrant populations. Independently from coverage, evidence suggests that survey nonrespondents may have higher alcohol consumption and harm, suggesting they may concentrate high levels of alcohol consumptions [33, 34, 35]. Furthermore, ATS potentially misses some (mostly low-risk) drinkers, even among respondents. [36] noted that the proportion of respondents classified as non-drinkers (based on AUDIT item 1) in ATS is 10% higher than in HSE, in which respondents are asked to confirm whether they never drink alcohol, or only drink 'very occasionally'. All these factors are likely to explain some of the discrepancy visible in Study 3, but also affect MD and RMSD estimates reported in Study 1.

4.3 Potential applications

While retaining the AUDIT's strengths (speed, accuracy, international standardisation), the Extended AUDIT improves the granularity of the information collected on alcohol consumption, and captures greater information from higher-risk drinkers, by remediating the right-truncation of the traditional AUDIT consumption items. The EWAC enhances the Extended AUDIT's health education and promotion qualities, by translating it into an understandable scale: alcohol consumption. This has the merit of focusing the recipient's attention to alcohol units or grams, and can facilitate uptake with brief interventions targetting skills in recognising the alcohol content of different drinks and drink sizes to reduce consumption.

This may be particularly relevant as part of efforts to provide earlier interventions (identification and brief advice) to increasing-risk drinkers, to prevent (instead of treat) alcohol use disorders [[37];]. Knowledge of alcohol beverage content is not universal. Many countries have not adopted a measure of standard drinks [38]. In England, two thirds of drinkers could assess the standard drink equivalent in wine or beer of one alcohol unit [39].

Contributions by [4] and [5] are an incentive to adopt a more clinical approach to alcohol use disorders prevention, treating alcohol consumption like any other risk factor: blood pressure or cholesterol are frequently monitored by general physicians in a non stigmatising way and can act as a trigger for health intervention. In a similar way, [4] and [5] argue that alcohol use disorders are best prevented if individuals know their consumption, and general practitioners can engage with patients effectively in managing risks, first with lifestyle interventions and, if relevant, pharmacological alternatives.

The proposed EWAC can fulfil the same alcohol use disorder screening functions as the AUDIT-C, while at the same time providing transparent information to patients in a less stigmatising way than terminology previously in use (hazardous and harmful drinking). Communication of an alcohol consumption estimate can motivate increasing risk drinkers to monitor their consumption, while at the same time encouraging low-risk drinkers to maintain this lifestyle. Its format is particularly suitable for digital intervention. Study 3 suggests that the EWAC is as reliable when the Extended AUDIT-C is self-administered as when it is administered by a researcher. The EWAC can be part of personalised feedback in a digital brief intervention, based on the overall consumption estimate as well as the intensity of alcohol consumption (AUDIT item 3).

5 Ethics

Study 1 and 3 were approved by the University of Southampton's Faculty of Medicine Ethics Committee (ERGO 44682). Study 2 was approved by the Health Research Authority National Research Ethics Service (IRAS 247458; REC 18/SC/0564).

6 Supplementary materials

6.1 Supplementary materials 1: EWAC coefficients (CSV file)

6.2 Supplementary materials 2: Bayesian model report (PDF file)

6.3 Supplementary materials 3: Subgroup analyses

Table 7: Coefficients of linear regression of the bias and error of EWAC compared with GF in all respondents (n = 21,874)

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu % Date and time: Sun, Feb 16, 2020 - 23:16:30 % Requires LaTeX packages: dcolumn

Table 8: Coefficients of linear regression of the bias and error of EWAC compared with GF in respondents with a hazardous/harmful alcohol use (AUDIT-C>=5 or AUDIT>=8; (n = 9,850)

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu % Date and time: Sun, Feb 16, 2020 - 23:16:30 % Requires LaTeX packages: dcolumn

Table 3:

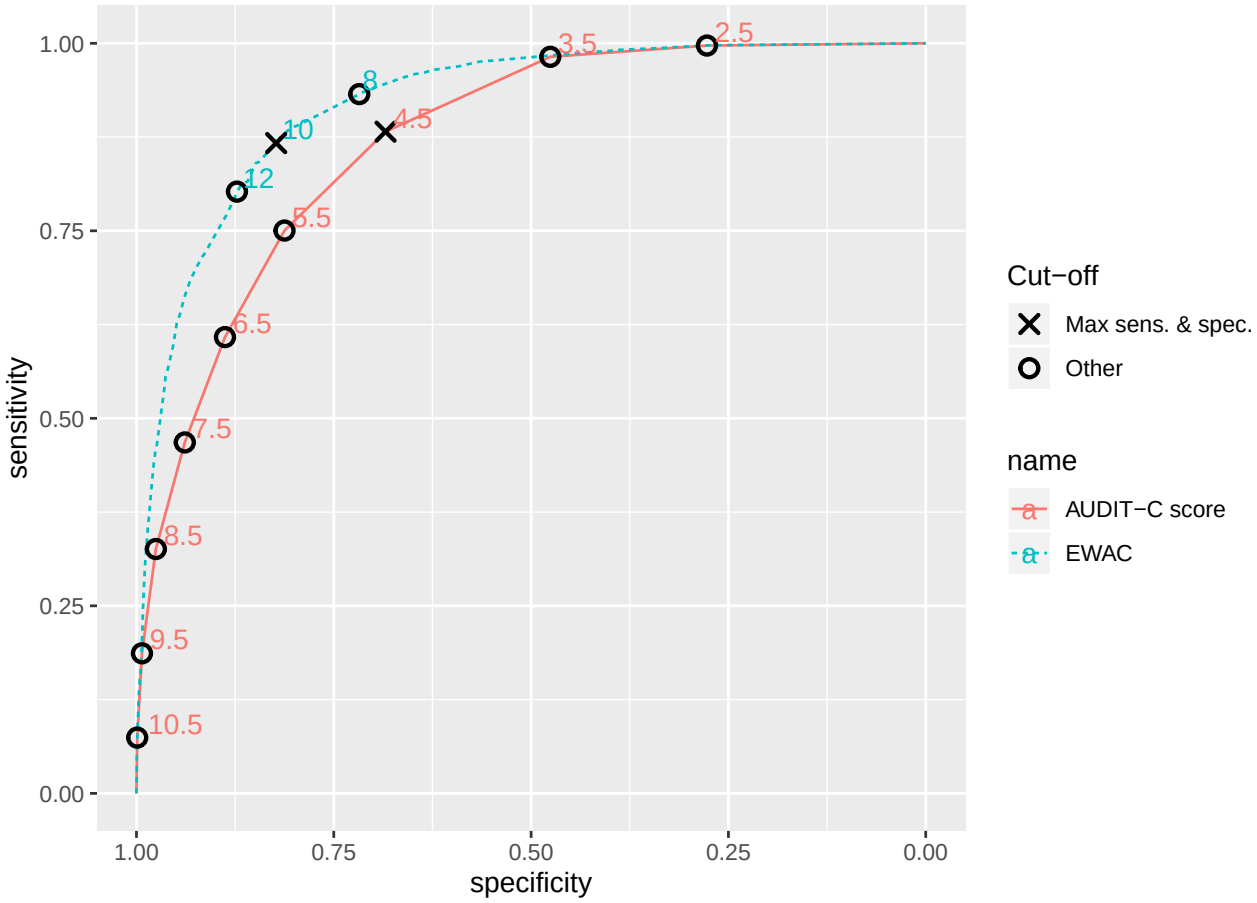
	(EWAC_QFV - GFMEANWEEKLY)	LOG((EWAC_QFV - GFMEANWEEKLY)^2
	1	2
Constant	0.7 (−0.1, 1.5)	0.7 (0.5, 0.9)***
sexMen	−1.2 (−2.0, −0.4)**	0.7 (0.5, 0.9)***
f_ageg16-24 years	0.7 (−0.1, 1.6)	0.6 (0.4, 0.8)***
f_ageg35-44 years	−0.6 (−1.4, 0.2)	0.03 (−0.2, 0.2)
f_ageg45-54 years	0.03 (−0.7, 0.8)	0.1 (−0.1, 0.3)
f_ageg55-64 years	0.5 (−0.3, 1.3)	0.1 (−0.1, 0.3)
f_ageg65-74 years	0.1 (−0.8, 0.9)	−0.2 (−0.4, 0.03)
f_ageg75+ years	−0.2 (−1.2, 0.8)	−0.3 (−0.5, −0.00)*
ethgrpWhite Other	1.1 (0.5, 1.7)***	−0.6 (−0.8, −0.4)***
ethgrpMixed	−0.5 (−1.8, 0.8)	−0.3 (−0.6, 0.1)
ethgrpAsian	0.9 (−0.2, 2.1)	−0.8 (−1.1, −0.5)***
ethgrpBlack	2.0 (0.9, 3.1)***	−0.7 (−1.0, −0.4)***
ethgrpOther	2.7 (0.8, 4.5)**	−1.1 (−1.6, −0.6)***
religionChristian	0.00 (−0.3, 0.3)	−0.2 (−0.3, −0.2)***
religionMuslim	1.0 (−1.8, 3.7)	−0.5 (−1.2, 0.2)
religionAny other religion	−0.5 (−1.3, 0.3)	−0.03 (−0.2, 0.2)
highqualNVQ <= 3	−0.01 (−0.6, 0.5)	0.01 (−0.1, 0.1)
highqualNVQ4+ (degree)	−0.1 (−0.7, 0.4)	0.1 (−0.02, 0.3)
highqualOther	−0.3 (−1.0, 0.4)	0.01 (−0.2, 0.2)
smokstatStopped>1y ago	−0.1 (−0.4, 0.3)	0.4 (0.3, 0.5)***
smokstatStopped in past year	−0.3 (−1.5, 0.9)	0.6 (0.3, 0.9)***
smokstatSmoker	−0.7 (−1.1, −0.2)**	0.9 (0.8, 1.0)***
sexMen:f_ageg16-24 years	−0.1 (−1.3, 1.0)	−0.1 (−0.4, 0.2)
sexMen:f_ageg35-44 years	0.04 (−1.1, 1.2)	0.00 (−0.3, 0.3)
sexMen:f_ageg45-54 years	−0.5 (−1.5, 0.6)	0.04 (−0.2, 0.3)
sexMen:f_ageg55-64 years	−0.1 (−1.2, 1.0)	0.2 (−0.1, 0.5)
sexMen:f_ageg65-74 years	1.3 (0.2, 2.4)*	0.3 (0.00, 0.6)*
sexMen:f_ageg75+ years	1.5 (0.3, 2.8)*	−0.2 (−0.5, 0.1)
Observations	21,874	21,874
R^2	0.01	0.05
Adjusted R^2	0.01	0.05
Residual Std. Error (df = 21846)	11.0	2.9
F Statistic (df = 27; 21846)	5.1***	40.0***

Notes: *P < .05
**P < .01
***P < .001

Table 4:

	(EWAC_QFV - GFMEANWEEKLY)	
	1	2
Constant	0.4 (−1.4, 2.2)	0.4 (−1.4, 2.2)
sexMen	−1.3 (−2.9, 0.3)	−1.3 (−2.9, 0.3)
f_ageg16-24 years	1.0 (−0.6, 2.6)	1.0 (−0.6, 2.6)
f_ageg35-44 years	−0.8 (−2.5, 0.9)	−0.8 (−2.5, 0.9)
f_ageg45-54 years	0.5 (−1.1, 2.2)	0.5 (−1.1, 2.2)
f_ageg55-64 years	2.2 (0.5, 3.9)*	2.2 (0.5, 3.9)*
f_ageg65-74 years	1.4 (−0.7, 3.4)	1.4 (−0.7, 3.4)
f_ageg75+ years	4.2 (0.9, 7.6)*	4.2 (0.9, 7.6)*
ethgrpWhite Other	1.6 (0.2, 3.0)*	1.6 (0.2, 3.0)*
ethgrpMixed	−1.5 (−4.0, 1.0)	−1.5 (−4.0, 1.0)
ethgrpAsian	2.4 (−0.5, 5.2)	2.4 (−0.5, 5.2)
ethgrpBlack	4.8 (2.1, 7.5)***	4.8 (2.1, 7.5)***
ethgrpOther	5.9 (1.6, 10.1)**	5.9 (1.6, 10.1)**
religionChristian	0.2 (−0.4, 0.8)	0.2 (−0.4, 0.8)
religionMuslim	4.6 (−4.7, 13.9)	4.6 (−4.7, 13.9)
religionAny other religion	−1.7 (−3.4, −0.01)*	−1.7 (−3.4, −0.01
highqualNVQ <= 3	0.3 (−0.9, 1.4)	0.3 (−0.9, 1.4)
highqualNVQ4+ (degree)	0.1 (−1.2, 1.3)	0.1 (−1.2, 1.3)
highqualOther	−0.8 (−2.4, 0.8)	−0.8 (−2.4, 0.8)
smokstatStopped>1y ago	−0.6 (−1.3, 0.2)	−0.6 (−1.3, 0.2)
smokstatStopped in past year	−0.5 (−2.6, 1.6)	−0.5 (−2.6, 1.6)
smokstatSmoker	−1.1 (−1.8, −0.4)**	−1.1 (−1.8, −0.4)
favdrinkCider	−1.1 (−2.5, 0.2)	−1.1 (−2.5, 0.2)
favdrinkMixed spirits	1.3 (−0.1, 2.6)	1.3 (−0.1, 2.6)
favdrinkOther	3.5 (−0.1, 7.0)	3.5 (−0.1, 7.0)
favdrinkSpirits alone	1.0 (−0.1, 2.1)	1.0 (−0.1, 2.1)
favdrinkWine	0.6 (−0.1, 1.4)	0.6 (−0.1, 1.4)
tryalclyc2Attempt to cut down in last 12 months	−0.5 (−1.2, 0.1)	−0.5 (−1.2, 0.1)
sexMen:f_ageg16-24 years	−0.1 (−2.2, 2.0)	−0.1 (−2.2, 2.0)
sexMen:f_ageg35-44 years	−0.1 (−2.3, 2.1)	−0.1 (−2.3, 2.1)
sexMen:f_ageg45-54 years	−0.9 (−3.0, 1.2)	−0.9 (−3.0, 1.2)
sexMen:f_ageg55-64 years	−1.1 (−3.3, 1.0)	−1.1 (−3.3, 1.0)
sexMen:f_ageg65-74 years	1.3 (−1.1, 3.7)	1.3 (−1.1, 3.7)
sexMen:f_ageg75+ years	−1.5 (−5.3, 2.3)	−1.5 (−5.3, 2.3)
Observations	9,850	9,850
R^2	0.02	0.02
Adjusted R^2	0.01	0.01
Residual Std. Error (df = 9816)	14.4	14.4
F Statistic (df = 33; 9816)	4.9***	4.9***
Notes:	*P < .05	
	**P < .01	
	***P < .001	

6.4 Supplementary materials 4: ROC curves



7 Appendix hospital study

Sample size was set to power three statistical hypotheses listed in [30]. An internal pilot conducted on 130 participants was used to obtain estimates of the variance in MD. The study was powered to detect a minimum MD between EWAC and TLFB of 1 unit. Using a two-sided paired t-test, 80% power and 95% confidence, the required sample size was 387 in total. In addition, the study was powered to detect a minimum RMSD greater than $\text{RMSD}_{H_0} = 2$ units using a one-sided, one-sample χ^2 test of variance. This involves testing the null hypothesis, $H_0 : \text{RMSD}^2 = 4$ versus the one-sided alternative hypothesis $H_a : \text{RMSD}^2 > 4$. Dixon & Massey [40] set the minimum detectable value of the variance as $\text{RMSD}^2_{H_a} = \text{RMSD}^2_{H_0} \cdot \chi^2_{n-1,1-\alpha}/\chi^2_{n-1,1-\beta}$. With $n = 130$, $\alpha = 0.05$ and $\beta = .2$, $\text{RMSD}_{H_a} = 2.37$. The test statistic $\chi^2 = (n-1)\text{RMSD}^2/4$ follows a χ^2 distribution with $n-1$ degrees of freedom. The normal approximation proposed by Dixon & Massey [40] to the sample size require is

$$n = 1/2 \left(\frac{z_{1-\alpha} - z_{\beta}}{\ln(\text{RMSD}_{H_a}) - \ln(\text{RMSD}_{H_0})} \right)^2$$

In our pilot study, the required sample size was 106, that is inferior to the target pilot sample size.

8 Tables

Table 1: Overview of alcohol schedule used in this paper.

Survey module

Schedule

ATS QF module

On how many days, if any, did you personally drink a drink containing alcohol in the last four weeks?

What was the maximum number of units you personally consumed on any one day when drinking an alcoholic drink or drinks in the last four weeks?

On how many days, if any, in the last four weeks did you personally drink... [prompting in turn '51-60 units?', '41-50 units?', ..., '1-2 units?

Health Survey for England

Thinking now about all kinds of drinks, how often have you had an alcoholic drink of any kind during the last 12 months? [8 items from 'Almost every day' to 'Not at all in the last 12 months']

Did you have an alcoholic drink in the seven days ending yesterday?

On how many days out of the last seven did you have an alcoholic drink?

Which day last week did you last have an alcoholic drink/have the most to drink?

Thinking about last [answer to previous question], what types of drink did you have that day? [list of 8 types of alcohol beverages]

[running through each type of beverage and recording number of units drunk]

References

[1] Shield K, Manthey J, Rylett M, Probst C, Wettlaufer A, Parry CDH, et al. National, regional, and global burdens of disease from 2000 to 2016 attributable to alcohol use: a comparative risk assessment study. The Lancet Public Health. 2020 jan;5(1):e51–e61. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2468266719302312.

[2] of Health NI, Excellence C. Alcohol-use disorders: prevention. Public health guideline [PH24]; 2010. Available from: www.nice.org.uk/guidance/ph24.

[3] National Institute of Health and Care Excellence. Alcohol-use disorders: diagnosis, assessment and management of harmful drinking and alcohol dependence. Clinical guideline [CG115]; 2011. Available from: https://www.nice.org.uk/guidance/cg115.

[4] Nutt DJ, Rehm J. Doing it by numbers: A simple approach to reducing the harms of alcohol. Journal of Psychopharmacology. 2014 jan;28(1):3–7. Available from: http://journals.sagepub.com/doi/10.1177/0269881113512038.

[5] Rehm J, Anderson P, Manthey J, Shield KD, Struzzo P, Wojnar M, et al. Alcohol Use Disorders in Primary Health Care: What Do We Know and Where Do We Go? Alcohol and Alcoholism. 2016 jul;51(4):422–427. Available from: https://academic.oup.com/alcalc/article-lookup/doi/10.1093/alcalc/agv127.

[6] Australian Government National Health and Medical Research Council. Australian Guidelines to Reduce Health Risks from Drinking Alcohol; 2009. Available from: https://www.nhmrc.gov.au/health-topics/alcohol-guidelines.

[7] UK Department of Health. UK Chief Medical Officers' Low Risk Drinking Guidelines. August 2016; 2016. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment{_}data/file/545937/UK{_}CMOs{_}{_}report.pdf.

[8] Wood AM, Kaptoge S, Butterworth AS, Willeit P, Warnakula S, Bolton T, et al. Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599912 current drinkers in 83 prospective studies. The Lancet. 2018;391(10129):1513–1523.

[9] Babor TF, Higgins-Biddle JC, Saunders JB, Monteiro MG. The Alcohol Use Disorders Identification Test. Guidelines for Use in Primary Care. Second Edition. Geneva: World Health Organisation, Department of Mental Health and Substance Dependence; 2001. Available from: http://www.who.int/substance{_}abuse/publications/audit/en/.

[10] de Meneses-Gaya C, Zuardi AW, Loureiro SR, Crippa JAS. Alcohol Use Disorders Identification Test (AUDIT): An updated systematic review of psychometric properties. Psychology & Neuroscience. 2009;2(1):83–97. Available from: http://doi.apa.org/getdoi.cfm?doi=10.3922/j.psns.2009.1.12.

[11] Kaner E, Bland M, Cassidy P, Coulton S, Dale V, Deluca P, et al. Effectiveness of screening and brief alcohol intervention in primary care (SIPS trial): pragmatic cluster randomised controlled trial. BMJ. 2013 jan;346(jan09 2):e8501–e8501. Available from: http://kar.kent.ac.uk/33045/http://www.bmj.com/cgi/doi/10.1136/bmj.e8501.

[12] Crane D, Garnett C, Michie S, West R, Brown J. A smartphone app to reduce excessive alcohol consumption: Identifying the effectiveness of intervention components in a factorial randomised control trial. Scientific Reports. 2018 dec;8(1):4384. Available from: http://www.nature.com/articles/s41598-018-22420-8.

[13] Beard E, Brown J, West R, Acton C, Brennan A, Drummond C, et al. Protocol for a national monthly survey of alcohol use in England with 6-month follow-up: 'The Alcohol Toolkit Study' Health behavior, health promotion and society. BMC Public Health. 2015;15(1). Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-84925003632{&}partnerID=40{&}md5=7cd590a6c2ad9d8d46227303618ca917.

[14] Lemmens P, Tan ES, Knibbe Ra. Measuring quantity and frequency of drinking in a general population survey: a comparison of five indices. Journal of studies on alcohol. 1992;53(5):476–486.

[15] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2017. Available from: https://www.R-project.org/.

[16] Wickham H. tidyverse: Easily Install and Load the 'Tidyverse'; 2017. R package version 1.2.1. Available from: https://CRAN.R-project.org/package=tidyverse.

[17] Stan Development Team. RStan: the R interface to Stan; 2018. R package version 2.18.2. Available from: http://mc-stan.org/.

[18] Dutey-Magni P. Concurrent validity of an Estimator of Weekly Alcohol Consumption (EWAC) based on the Extended AUDIT. Code repository; 2020. Available from: https://bitbucket.org/PeteDM/ewac_concurrent_validity/src.

[19] Greenfield TK. Ways of measuring drinking patterns and the difference they make: experience with graduated frequencies. Journal of Substance Abuse. 2000 sep;12(1-2):33–49. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0899328900000390.

[20] Heeb JL, Gmel G. Measuring alcohol consumption: A comparison of graduated frequency, quantity frequency, and weekly recall diary methods in a general population survey. Addictive Behaviors. 2005 mar;30(3):403–413. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0306460304001698.

[21] Rehm J. Measuring Quantity, Frequency, and Volume of Drinking. Alcoholism: Clinical and Experimental Research. 1998;22(Supplement S2):4s–14s. Available from: http://dx.doi.org/10.1111/j.1530-0277.1998.tb04368.x.

[22] Greenfield TK, Kerr WC, Bond J, Ye Y, Stockwell T. Improving Graduated Frequencies Alcohol Measures for Monitoring Consumption Patterns: Results from an Australian National Survey and a US Diary Validity Study. Contemporary Drug Problems. 2009;36(3-4):705–733. Available from: http://cdx.sagepub.com/lookup/doi/10.1177/009145090903600320.

[23] NatCen Social Research, Royal Free and University College Medical School. Health Survey for England, 2011 [computer file]. Colchester, Essex: UK Data Archive [distributor], April 2013. SN: 7260. Colchester, Essex; 2013.

[24] Boniface S, Kneale J, Shelton N. Drinking pattern is more strongly associated with under-reporting of alcohol consumption than socio-demographic factors: evidence from a mixed-methods study. BMC public health. 2014;14(1):1297. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25519144.

[25] Stockwell T, Zhao J, Greenfield T, Li J, Livingston M, Meng Y. Estimating under- and over-reporting of drinking in national surveys of alcohol consumption: identification of consistent biases across four English-speaking countries. Addiction. 2016 jul;111(7):1203–1213. Available from: http://doi.wiley.com/10.1111/add.13373.

[26] Grant KA, Tonigan JS, Miller WR. Comparison of three alcohol consumption measures: A concurrent validity study. Journal of Studies on Alcohol. 1995;56(2):168–172.

[27] Hoeppner BB, Stout RL, Jackson KM, Barnett NP. How good is fine-grained Timeline Follow-back data? Comparing 30-day TLFB and repeated 7-day TLFB alcohol consumption reports on the person and daily level. Addictive Behaviors. 2010;35(12):1138–1143.

[28] Vinson DC, Reidinger C, Wilcosky T. Factors affecting the validity of a Timeline Follow-Back interview. Journal of Studies on Alcohol. 2003 sep;64(5):733–740.

[29] Public Health England. Alcohol sales in England in 2014: Analysis to assess suitability for inclusion as an indicator in the Local Alcohol Profiles for England; 2017. Available from: https://webarchive.nationalarchives.gov.uk/20190501132530/https://fingertips.phe.org.uk/documents/Alcohol%20sales%20data%202014.pdf.

[30] Dutey-Magni P, Sinclair J, Brown J. Concurrent validity of an Estimator of Weekly Alcohol Consumption (EWAC) based on the Extended AUDIT. OSF; 2018. Available from: osf.io/7we4m.

[31] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics. 1988 sep;44(3):837.

[32] Rubinsky AD, Dawson DA, Williams EC, Kivlahan DR, Bradley KA. AUDIT-C Scores as a Scaled Marker of Mean Daily Drinking, Alcohol Use Disorder Severity, and Probability of Alcohol Dependence in a U.S. General Population Sample of Drinkers. Alcoholism: Clinical and Experimental Research. 2013 aug;37(8):1380–1390. Available from: http://doi.wiley.com/10.1111/acer.12092.

[33] Gorman E, Leyland AH, McCartney G, White IR, Katikireddi SV, Rutherford L, et al. Assessing the Representativeness of Population-Sampled Health Surveys Through Linkage to Administrative Data on Alcohol-Related Outcomes. American Journal of Epidemiology. 2014 nov;180(9):941–948. Available from: https://academic.oup.com/aje/article-lookup/doi/10.1093/aje/kwu207.

[34] Christensen AI, Ekholm O, Gray L, Glümer C, Juel K. What is wrong with non-respondents? Alcohol-, drug- and smoking-related mortality and morbidity in a 12-year follow-up study of respondents and non-respondents in the Danish Health and Morbidity Survey. Addiction. 2015 sep;110(9):1505–1512. Available from: http://doi.wiley.com/10.1111/add.12939.

[35] Boniface S, Scholes S, Shelton N, Connor J. Assessment of Non-Response Bias in Estimates of Alcohol Consumption: Applying the Continuum of Resistance Model in a General Population Survey in England. PLOS ONE. 2017 jan;12(1):e0170892. Available from: https://dx.plos.org/10.1371/journal.pone.0170892.

[36] de Vocht F, Brown J, Beard E, Angus C, Brennan A, Michie S, et al. Temporal patterns of alcohol consumption and attempts to reduce alcohol intake in England. BMC Public Health. 2016;16(1):917.

[37] Lavoie D. Alcohol identification and brief advice in England: A major plank in alcohol harm reduction policy. Drug and Alcohol Review. 2010;29(6):608–611.

[38] Kalinowski A, Humphreys K. Governmental standard drink definitions and low-risk alcohol consumption guidelines in 37 countries. Addiction. 2016;111(7):1293–1298. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/add.13341.

[39] UK Office for National Statistics. Drinking: adults' behaviour and knowledge in 2008. Opinions (Omnibus)

Survey Report No. 39; 2009. Available from: http://www.ias.org.uk/uploads/pdf/News%20stories/ons-report-jan09.pdf.

[40] Dixon WJ, Massey FJJ. Introduction to Statistical Analysis. 4th ed. New York: McGraw–Hill; 1983.