

# BMJ Open A systematic review of alcohol screening and assessment measures for young people: a study protocol

Paul Toner,<sup>1</sup> Jan R Böhnke,<sup>2</sup> Jim McCambridge<sup>1</sup>

**To cite:** Toner P, Böhnke JR, McCambridge J. A systematic review of alcohol screening and assessment measures for young people: a study protocol. *BMJ Open* 2017;0:e016406. doi:10.1136/bmjopen-2017-016406

► Prepublication history and additional material are available. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-016406>).

Received 11 February 2017  
Revised 3 April 2017  
Accepted 10 May 2017

## ABSTRACT

**Introduction** Alcohol consumption creates a significant public health burden, and young people who drink alcohol place themselves at risk of harm. Expert guidance and reviews have highlighted the pressing need for reliable and valid, age-appropriate alcohol screening and assessment measures for young people. The proposed systematic review will evaluate existing alcohol screening and assessment measures for young people aged 24 and under.

**Methods and analysis** Six electronic databases will be searched for published and grey literature. In addition, reverse and forward citation searching and consultation with experts will be performed. Three sets of search terms will be combined, including alcohol use/problems, young people and validation studies. The titles and abstracts of reports from the searches will be screened, and potentially relevant full-text reports will be retrieved and independently assessed for inclusion by two reviewers based on prespecified criteria. Discrete validation studies within included reports will then be assessed for eligibility. There will be an a priori basic quality threshold for predictive validity, internal and test–retest for studies to warrant full data extraction. Studies above the quality threshold will be assessed for quality using the modified consensus-based standards for the selection of health measurement instruments checklist and a quality assessment tool for diagnostic accuracy studies.

**Dissemination** This review will highlight the best performing measures both for screening and assessment based on their psychometric properties and the quality of the validation studies supporting their use. Providing clear guidance on which existing measures perform best to screen and assess alcohol use and problems in young people will inform policy, practice and decision-making, and clarify the need for further research.

**Trial registration number** International Prospective Register of Systematic Reviews, CRD42016053330.

## INTRODUCTION

The National Institute for Health and Care Excellence public health guidance 24 (2010) emphasises the need for research to identify which screening tool should be considered as the ‘gold standard’ for assessing the drinking behaviour of those under the age of 18 (p43).<sup>1</sup> The proposed systematic review will contribute to the scientific evidence base

## Strengths and limitations of this study

- No existing reviews of validation studies of alcohol screening and assessment measures with youth populations.
- Rigorous systematic review design using advanced psychometric principles.
- Brings together and appraises both alcohol screening and assessment literatures.
- Most in-depth analyses focus on the best performing screening and assessment measures.
- English-language studies only were included (although evaluated instruments could be in other languages).

by evaluating existing alcohol screening and assessment measures for young people aged 24 and under. Within this broader age range, inferences can be drawn about validation data for a range of subgroups defined by age and other characteristics. Included measures will be categorised based on key characteristics, such as whether they are used for screening and/or assessment purposes. Review findings will inform decision-making about either the adaptation and testing of an existing alcohol measure(s) or the development and testing of a new measure(s) in a parent study. Having an assessment measure(s) that correctly identifies young people at risk from alcohol-related harm provides researchers, practitioners and policymakers with an instrument(s) capable of measuring prevalence and patterns of risk and which can inform appropriate interventions.

## Target condition

Alcohol consumption creates a significant public health burden. Epidemiological evidence indicates that the majority of alcohol-related problems are experienced by heavy episodic drinkers who are low-level consumers in comparison with those who drink at consistently high levels.<sup>2</sup> In the UK, while the proportion of young people who drink alcohol is declining, for the majority who do drink, heavy episodic drinking is



CrossMark

<sup>1</sup>Department of Health Sciences, University of York, York, UK  
<sup>2</sup>Dundee Centre for Health and Related Research, School of Nursing and Health Sciences, University of Dundee, Dundee, Scotland, UK

## Correspondence to

Dr Paul Toner;  
[paul.toner@york.ac.uk](mailto:paul.toner@york.ac.uk)

normative.<sup>3 4</sup> Contemporary evidence indicates that it is the extent of alcohol involvement in adolescence rather than age of first use that confers risk for both acute and longer term adverse consequences.<sup>5 6</sup> Fifteen-year-olds to sixteen-year-olds in the UK have one of the highest rates of underage drinking and drunkenness in Western Europe.<sup>7</sup> Young people who drink are also at increased risk of using other substances and for other risk-taking behaviours.<sup>8</sup>

There are currently no reviews of validation studies of alcohol screening and assessment measures in youth populations. This review will bring together evidence from both screening and assessment literatures to comprehensively assess which instruments work best to identify young people who are drinking hazardously and/or harmfully. Hazardous drinking refers to consumption that is risky, whereas harmful drinking is that which is problematic. Providing rigorous review evidence will enable decision-makers to make informed choices about which alcohol measures are most appropriate for youth populations. They will also have precise information, for example, on optimum cut-points for the best performing instruments by age, gender and settings in which the instruments were validated to make evidence informed decisions about implementation. A study that is strongly informed by psychometric principles will also be informative about the strength of existing evidence and the need for further research, including about the desirability of new measures.

Practitioners will have guidance on which existing instruments are most appropriate to screen for, and assess, a continuum of alcohol risk and harm, which has been described as unhealthy alcohol use.<sup>9</sup> This provides a foundation for decision-making about interventions.

### Index test(s)

Systematic review evidence shows that alcohol screening questionnaires perform better than alcohol markers or breath alcohol concentration in all age groups.<sup>10</sup> In the USA and elsewhere, there are a number of instruments containing alcohol-related items developed specifically for adolescents, such as Car Relax Alone Friends Family Trouble (CRAFFT)<sup>11</sup> or Problem Oriented Screening Instrument for Teenagers.<sup>12</sup> However, these measures are not substance-specific and use a composite score for both alcohol and other drugs. As a result, they may not be optimal for assessing drinking. For example, when the Alcohol Use Disorder Identification Test (AUDIT)<sup>13</sup> was compared with CRAFFT, AUDIT demonstrated higher sensitivity and specificity at an optimal cut-point than CRAFFT in identifying young people at risk of alcohol-related problems.<sup>14</sup> For this review, index tests (ie, those that are evaluated) are defined as screening or assessment measures for alcohol use or problems only.

### RATIONALE

This systematic review will evaluate the validity of available instruments for screening and assessing alcohol

consumption and related problems in young people aged 24 and under by summarising and interrogating available evidence. This review will highlight the best performing measures both for screening and assessment based on their psychometric properties and the quality of the validation studies supporting their use. Providing clear guidance on which existing measures perform best to screen and assess alcohol use and problems in young people will make a significant contribution to policy and practice.

### OBJECTIVES

The objective is to summarise and psychometrically evaluate validation studies comparing the accuracy of an alcohol measure with a previously validated questionnaire or diagnostic interview for identifying hazardous and/or harmful drinking in young people.

### TARGET CONDITIONS

This will be current alcohol use, including hazardous drinking defined as exceeding a validated screening score or recommended limits on consumption (eg, USA — National Institute on Alcohol Abuse and Alcoholism), and harmful drinking defined as a pattern of drinking causing damage to physical (eg, injuries, poisoning) or psychological (eg, anxiety, depression) health, including alcohol dependence, or causing social consequences (eg, educational problems) that are supported by epidemiological or other empirical evidence.

### METHODS AND ANALYSIS

#### Design

We will adhere to the National Institute for Health Research Centre for Reviews and Dissemination (2009) guidelines in conducting the review<sup>15</sup> and the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols<sup>16</sup> guidelines in reporting the review protocol, which will be published on International Prospective Register of Systematic Reviews.

#### Search methods for identification of reports

A comprehensive search will be conducted in order to identify relevant reports. Help will be sought from information specialists in the design of the search strategy. We will search for reports written in the English language only as resources are not available to translate foreign language reports.

#### Electronic searches

Searches will be run in the following databases: Medical Literature Analysis and Retrieval System Online (MEDLINE; Ovid 1946–), Excerpta Medica Database (Embase; Ovid 1974–), Psychological Information Database (PsycINFO; Ovid 1806–) and Social Sciences Citation Index (SSCI; Web of Science 1956–).

The Health Management Information Consortium Database (Ovid 1979–) and the University of Washington

Alcohol and Drug Abuse Institute Library Search – Substance Use Screening and Assessment Instruments Database will be searched for grey literature.

The search strategies will be designed using medical subject headings and free text words adapted for each database. Three sets of search terms will be combined. The first set will encompass different terms for alcohol use and alcohol problems including substance use to identify alcohol measures of interest. The second set of terms will comprise terms for young people. The third set will include terms relating to validation studies. The SSCI will be used for reverse and forward citation searching. A draft MEDLINE search strategy is appended (see online supplementary appendix I).

### Searching other resources

We will also look for additional reports by screening the reference lists of retrieved articles and reviews and contacting authors of retrieved reports, and experts in the field for reports that may not have been identified through the searches.

### Selection of studies

Two reviewers will separately screen the titles and abstracts of reports retrieved by the searches using EndNote X7. Reports identified as potentially relevant will be obtained as full-text articles, which we will assess for inclusion using a checklist based on our prespecified selection criteria. Discrete validation studies within included reports (many such reports evaluate more than one measure, see Data extraction section) will also be assessed for inclusion using the same criteria. Two reviewers will independently screen reports and studies for inclusion; where eligibility is unclear, this will be resolved by discussion, or if necessary consultation with a third reviewer.

### Selection criteria

#### Types of studies

We will include any type of validation study published in the English language from 1980 (when the classification of alcohol changed in the Diagnostic and Statistical Manual of Mental Disorders) onwards that aims to validate an alcohol use or problems screening or assessment measure (index test) in comparison with a previously validated alcohol measure (reference test).

#### Participants

Eighty per cent or above of study participants should be aged 24 and under. If only mean age and SD are stated, the expected proportion of those aged 24 and under will be calculated based on the corresponding normal distribution. Where only mean or median age are reported, this has to be under 21.0 years for the report to be included. Studies undertaken in student samples without the sample age defined will be eligible for inclusion, unless there are specific reasons to be concerned that below 80% of the participants are aged 24 and under.

### Index tests

Screening or assessment measures assessing only alcohol use or problems.

### Reference tests (comparators)

The reference tests are previously validated questionnaires or diagnostic interviews assessing alcohol use or problems. Where alcohol is assessed alongside other drugs, the study will only be selected if the reference test provides an alcohol-only result against which the index test is compared in the validation study.

The following will not be considered as a valid comparator: 'clinician judgement', alcohol biomarkers, alcohol diagnoses that are a composite of information contained within medical records, substance use measures that do not report a validated assessment of alcohol use or problems and alcohol questions that have not been previously validated.

### Outcomes

The direct report of predictive (including concurrent) validity of the index test against a comparator. The following are the data required: standardised regression coefficient or odds ratio (OR) or correlations or area under the curve (AUC) or % sensitivity; % specificity or % positive predictive values (PPV); % negative predictive values (NPV) or likelihood ratio.

## DATA COLLECTION AND ANALYSIS

### Data extraction

One reviewer will record all relevant data from included studies using a data extraction form. A second reviewer will verify the data extracted from included studies. Any discrepancies will be resolved by discussion or if required consultation with a third reviewer.

Included full papers or reports may contain multiple validation studies, defined for the purposes of this review as comparisons of index and reference tests. A single record for preliminary data extraction will be created for each validation study. The data extraction process will be as follows: (1) The eligibility criteria previously used to include reports in the review will be applied to each of the validation studies within the included reports. If, for example, the study aim was not to validate the index test or there is no information present on predictive validity (as previously defined) for a comparison of index and reference test, further data extraction will not take place. This is because this particular validation study is not eligible for inclusion in this review, even though the report within which it is published contains another validation study which is eligible for inclusion. (2) If a validation study is included, as defined in step 1, then basic quality threshold data will be extracted (see below). If the index test fails to make any of the a priori quality thresholds on predictive validity, internal or test-retest reliability, this study will be recorded as included in the review at step 2, and no further data extraction will take place. (3) If the quality thresholds (see below) are met,

then full data extraction and quality assessment will take place in step 3.

If a validation study reports only data on subscales of a questionnaire, data will be extracted as described in steps 2 and 3, thus treating the subscale as the index test. The same applies to studies only reporting validation data for specific subpopulations (eg, age categories, gender), where each will be treated as a separate validation study.

Data collection in relation to the index test(s): (1) predictive validity: cut-off scores (thresholds on each questionnaire), standardised regression coefficient, OR, correlations, AUC, % sensitivity, % specificity, % PPV, % NPV, and likelihood ratio; (2) internal validity: item-to-total correlations and percentage of explained variance by proposed factor model; (3) reliability: (adjusted) Cronbach's alpha, Guttman's lambda, omega, Pearson correlation, intraclass correlation coefficient and kappa coefficient; and (4) information on acceptability/feasibility.

Descriptive details on the index test(s) will also be recorded on a Microsoft Excel 2010 spreadsheet: instrument name and acronym; original English-language version or a translation; measurement construct, that is, alcohol use/problems or both; purpose, that is, screening/assessment or both; dichotomous/continuous scoring or both; administration mode, that is, interviewer-assisted or self-completion; and recall timeframe. Other information will include reference test(s) used to validate index test(s), study authors, country, year of publication, study setting, sample size, % female, ethnicity and mean age with standard deviation (SD).

### Quality assessments

First, there will be an a priori basic quality threshold for studies to warrant full data extraction as follows: the index test must achieve a predictive validity of above 0.7 (eg, standardised regression coefficient) or 0.8 AUC, % sensitivity or an internal consistency above 0.8 (adjusted Cronbach's alpha for 10 items), or a test-retest of above 0.7 (eg, kappa coefficient). These quality thresholds were aggregated from several strands of literature. Some are standards adopted by testing communities,<sup>17 18</sup> others are seen, for example, as compromises between the increased reliability of longer instruments versus the need for short assessments (eg<sup>19</sup>). In relation to the validity coefficients, these have been shown to be realistic thresholds when assessing similar constructs with the same method.<sup>20</sup> Studies that compare a short version of a parent instrument as index and reference test respectively will be excluded from quantitative synthesis due to the potential for overestimation of validity.

Second, included studies above the quality threshold will also be assessed for quality using a modified consensus-based standards for the selection of health measurement instruments (COSMIN) checklist and a quality assessment tool for diagnostic accuracy studies (QUADAS-2).

## STATISTICAL ANALYSIS AND DATA SYNTHESIS

### Descriptive methods

All included studies will be documented in a table providing their authors, country, and population parameters, for example, age, gender distribution, setting and so on. In addition, a summary table of included studies identifying the number of validation studies per index test will be presented. Summary reliability and validity data with associated study quality indicators (COSMIN/QUADAS-2) will be presented in a third table. This third table will rank-order identified instruments based on their psychometric properties and the extent and quality of validation studies supporting their use. This table will be discussed qualitatively to identify those instruments where there is evidence of minimal psychometric quality based on our defined criteria for screening and assessment, that is, whether the index test is used to make a dichotomous decision (screening) or to provide a continuous score as a measure of alcohol consumption and/or related problems (assessment).

### META-ANALYSIS

Relevant data checked and agreed by two reviewers will be exported from Microsoft Excel 2010 to Stata V.14 for quantitative synthesis. We will calculate cut-off specific summary estimates of sensitivity and specificity for each test by alcohol category, using a cut-off score recommended by empirical studies. These will be presented graphically on forest plots showing paired sensitivities and specificities with 95% confidence intervals (CIs).

The goal of presenting these meta-analytic estimates is to provide a range of estimates across a variety of identified instruments as a benchmark for further research on instrument development. This is a departure from the standard meta-analytic goal of providing a single summary.

If at least five studies are available for analysis, summary estimates of test accuracy will be produced and a summary receiver operating characteristic (ROC) curve will be presented.<sup>21</sup> Subgroup analyses will be conducted for each subgroup where at least five studies are available, and Q and I<sup>2</sup> statistics will be calculated to gauge the degree of heterogeneity within subgroups. The potential for a meta-regression will be evaluated once all data are coded. Decision-making will also take account of the nature of the outcome data (dichotomous or continuous), and we will decide once the data set is complete whether and how to combine screening and assessment measures.

The same approach will be taken for a separate meta-analysis, which will aggregate the available reliability estimates.

### Investigation of sources of heterogeneity

We will investigate the following study characteristics as potential sources of heterogeneity on sensitivity and specificity by conducting subgroup analyses: year of publication, sample size, percentage female, mean age, country, ethnicity, index tests, reference tests, population (ie, clinical, community) and setting (ie, health, school).

## DISSEMINATION

### What the review adds

Depending on the extent of data available, this review will contribute the following:

1. To research plans for a parent study developing alcohol screening and/or assessment measures for young people aged 15–17 years old in the UK. This will include having the best instruments for screening and assessment ranked by their psychometric properties in included validation studies. In addition, candidate items for new measures will be identified from high factor/component loadings and/or item-to-total correlations. The information generated from the review can contribute more broadly to research agendas on alcohol screening and assessment in young people aged under 25.
2. ROC summary plots combining the best performing instruments to identify a benchmark against which future studies validating and/or developing instruments in this population can compare their performance.
3. ROC summary plots for selected individual instruments that will enable investigators and practitioners to select the measure that performs best for their domain of interest. For example, the instrument has demonstrated in multiple studies at an optimal cut-point that it is 100% sensitive in identifying young people at risk of an alcohol use disorder.

### DISSEMINATION PLANS

Dissemination activities will take place in the academic, practice and policy arenas. The review findings will be submitted for publication in one (or more) peer-reviewed journal(s) and will be presented at conferences such as the Society for the Study of Addiction (SSA) Annual Conference and the International Network on Brief Interventions for Alcohol and Other Drugs (INEBRIA) Annual Conference. Data sets generated will be made available by the corresponding author on reasonable request.

### REVIEW STATUS

At the time of submission, data extraction has commenced and the review is due for completion in June 2017.

**Contributors** PT conceived the study, is guarantor and codesigned the review, and will lead on data collection, analysis and interpretation. He led on preparing the manuscript. JRB codesigned the review and will provide advanced psychometric expertise on data collection, analysis and interpretation. He was involved in drafting the manuscript for important intellectual content. JM codesigned the review and will provide expert guidance and support on data collection, analysis and interpretation. He critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

**Competing interests** None declared.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

## REFERENCES

1. *Alcohol-use disorders: preventing the development of hazardous and harmful drinking*. NICE public health guidance 24, 2010.
2. Danielsson AK, Wennberg P, Hibell B, et al. Alcohol use, heavy episodic drinking and subsequent problems among adolescents in 23 European countries: does the prevention paradox apply? *Addiction* 2012;107:71–80.
3. Bellis MA, Hughes K. Getting drunk safely? Night-life policy in the UK and its public health consequences. *Drug Alcohol Rev* 2011;30:536–45.
4. Townshend TG. Youth, alcohol and place-based leisure behaviours: a study of two locations in England. *Soc Sci Med* 2013;91:153–61.
5. McCambridge J, McAlaney J, Rowe R. Adult consequences of late adolescent alcohol consumption: a systematic review of cohort studies. *PLoS Med* 2011;8:e1000413.
6. Maimaris W, McCambridge J. Age of first drinking and adult alcohol problems: systematic review of prospective cohort studies. *J Epidemiol Community Health* 2014;68:268–74.
7. Hibell B, Guttormsson U, Ahlström S, et al. The 2011 ESPAD report. *Substance use among students in 36 European countries*. Stockholm, Sweden: The Swedish Council for Information on Alcohol and Other Drugs, 2012.
8. MacArthur GJ, Smith MC, Melotti R, et al. Patterns of alcohol use and multiple risk behaviour by gender during early and late adolescence: the ALSPAC cohort. *J Public Health* 2012;34 Suppl 1:20–30.
9. Saitz R. Unhealthy alcohol use. *N Engl J Med* 2005;352:596–607.
10. Jackson R, Johnson M, Campbell F, et al. UK: SchARR Public Health Collaborating Centre, University of Sheffield, Screening and brief interventions for prevention and early identification of alcohol use disorders in adults and young people. 2009.
11. Stewart SH, Connors GJ. Screening for alcohol problems: what makes a test effective? *Alcohol Res Health* 2004;28:5–16.
12. Rahdert ER. *The Adolescent Assessment/Referral System Manual*. Rockville, US: Department of Health and Human Services, Alcohol, Drug Abuse, and Mental Health Administration, 1991.
13. Babor TF, De La Fuente JR, Saunders J, et al. *AUDIT: the Alcohol Use Disorders Identification Test: guidelines for use in primary health care*. Geneva, Switzerland: WHO, 1989.
14. Cook RL, Chung T, Kelly TM, et al. Alcohol screening in young persons attending a sexually transmitted disease clinic. comparison of AUDIT, CRAFFT, and CAGE instruments. *J Gen Intern Med* 2005;20:1–6.
15. *Systematic reviews: CRD's guidance for undertaking reviews in health care*. UK: Centre for Reviews and Dissemination, University of York, 2009.
16. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;349:g7647.
17. Fayers PM, Machin D. *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes*. 3rd edn. Chichester, UK: Wiley, 2016.
18. Nunnally JC, Bernstein IH. *Psychometric Theory*. 3rd edn. New York, USA: McGraw-Hill, 1994.
19. Böhnke JR, Lutz W. Using item and test information to optimize targeted assessments of psychological distress. *Assessment* 2014;21:679–93.
20. Meyer GJ, Finn SE, Eyde LD, et al. Psychological testing and psychological assessment. A review of evidence and issues. *Am Psychol* 2001;56:128–65.
21. Harbord RM, Deeks JJ, Egger M, et al. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007;8:239–51.