Prioritization of outcomes in efficacy and effectiveness alcohol brief intervention trials

Number of tables: 3

Word count: 3927

Prioritization of outcomes in efficacy and effectiveness alcohol brief intervention trials: International Multi-stakeholder e-Delphi Consensus Study to inform a core outcome set

Gillian W Shorter* BSc PhD[1,2]; Nick Heather BA MSc PhD[3]; Jeremy W Bray BA MA PhD[4]; Anne H Berman BSc MSc PhD[5]; Emma L Giles BSc PhD[2]; Amy J O'Donnell BSc PhD[6]; Carolina Barbosa PharmD MSc PHD[7] ; Mike Clarke BA DPhil[8]; Aisha Holloway BSc PhD[9]; Dorothy Newbury-Birch BSc PhD[2]

[1] Institute for Mental Health Sciences, School of Psychology, Ulster University, Coleraine, UK

[2] School of Health and Social Care, Teesside University, Middlesbrough, UK

[3] Faculty of Health & Life Sciences, Northumbria University, Newcastle upon Tyne, UK

[4] Bryan School of Business & Economics, University of North Carolina at Greensboro, Greensboro, NC, USA

[5] Department of Clinical Neuroscience, Karolinska Institutet, & Stockholm Health Care Services, Stockholm County Council, Sweden

[6] Institute of Health & Society, Newcastle University, Newcastle upon Tyne, UK

[7] Behavioral Health Economics Program, RTI International, Chicago, IL, USA

[8] Northern Ireland Methodology Hub, Queen's University of Belfast, Belfast, UK.

[9] School of Health in Social Sciences, University of Edinburgh, Edinburgh, UK

***Corresponding author**: Dr. Gillian W Shorter

Institute for Mental Health Sciences, School of Psychology, Ulster University, Cromore Road, Coleraine, Northern Ireland, BT52 1SA

Telephone: +44 (0) 2870 124025 (note no fax available)

Email: gw.shorter@ulster.ac.uk gillianwshorter@gmail.com

Grant support: Alcohol Research UK (Research Innovation Grant Number: R2016/04)

Abstract

Objective: Outcomes used in alcohol brief intervention (ABI) trials vary considerably. Achieving consensus about key outcomes can enhance evidence synthesis and improve healthcare guidelines. This was an international, e-Delphi study to prioritize outcomes for ABI trials as one step in a larger effort to develop an ABI core outcome set (COS).

Method: 150 registrants from 19 countries, and representing researchers, policymakers, and patients, participated in a two-round e-Delphi study. In Round 1, participants (n=137) rated 86 outcomes, derived from a review of the literature and a patient and public involvement panel, by importance. In Round 2, participants (n=114) received feedback on importance ratings for each outcome and a reminder of their personal rating before rating the outcomes for importance a second time. Seven additional outcomes suggested in Round 1 were added to the Round 2 questionnaire. We defined consensus *a priori* as 70% agreement across all stakeholder groups.

Results: Seven consumption outcomes met inclusion criteria: typical frequency, typical quantity, frequency of heavy drinking, alcohol-related problems, and weekly drinks, at risk drinking, and combined consumption measures. Others meeting the threshold were: alcohol-related injury; quality of life; readiness to change; and intervention fidelity.

Conclusions: This is the first international e-Delphi study to identify and prioritize outcomes for use in ABI trials. The use and reporting of outcomes in future ABI trials should improve evidence synthesis in systematic reviews and meta-analyses. Further work is required to refine these outcomes into a COS that includes guidance for measurement of outcomes.

Introduction

Alcohol brief interventions (ABIs) have emerged as the main approach to addressing hazardous and harmful alcohol use in a range of settings, including primary care, emergency departments, hospitals, online, criminal justice, workplaces, probation, and universities. According to NICE guidance PH24 (National Institute for Health and Clinical Excellence, 2010), ABIs are suitable for non-treatment-seeking alcohol users aged 16 or over who are currently experiencing, or are at risk of experiencing, problems from their alcohol use. ABIs are behavioral and/or motivational interventions designed to help drinkers reduce their alcohol consumption. They typically consist of a short, single session of feedback and tailored advice (brief advice), or longer, motivationally-based interventions that explore motivations for drinking and personal barriers to change (extended BI) (Cunningham et al., 2017). Essential components of ABI's are defined here as the assessment of personal alcohol use and tailored feedback provided directly to the drinker.

Systematic reviews of ABI trials do not always agree on the efficacy and effectiveness of ABIs to change alcohol use (e.g. Davoren et al., 2016; Kaner et al., 2018; Khadjesari et al., 2011; White et al., 2010). There are many possible reasons for this disagreement, such as changes in the population being studied over time, changes in baseline drinking, variability in ABI content and reporting, and inclusion and exclusion criteria variations, among other issues. An avoidable source of disagreement in the literature, however, arises from the wide variation in outcomes used and the difficulty in combining diverse outcomes in meta-analyses (Cumming, 2013; Kaner, et al., 2018). This can be variation in 'what' outcome is measured or, for a given outcome, variation in 'how' the outcome is measured.

Given the increasing role of systematic reviews and meta-analyses in determining health policy and given the potential for outcome heterogeneity to compromise these reviews

and analyses, there is a growing effort across a wide range of disciplines and disease categories to standardize trial outcomes (Williamson & Clarke, 2012). The importance of standardizing trial outcome measurement is recognized by the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT; Chan et al., 2013) and Consolidated Standards of Reporting Trials (CONSORT; Moher et al., 2010) statements; both statements recommend the use of a well-designed core outcome set (COS). A formal process for defining a COS has been established by the Core Outcome Measures in Effectiveness Trials (COMET) Initiative (Williamson et al., 2012; Williamson et al., 2017), and numerous studies using this process have been undertaken (Gargon et al., 2017). Given the current lack of a COS for ABI trials and the increasing importance of ABIs in alcohol policies worldwide, an ABI COS is urgently needed.

The selection and application of a COS is relevant to all ABI stakeholders, including beneficiaries of ABIs (service users), practitioners, and policymakers (Williamson et al. 2012). A COS ensures outcomes meaningful to service users are routinely considered in clinical trials and policymakers' perspectives are reflected in trial outcomes. Without a COS, the selection of trial outcomes remains at the sole discretion of the involved researchers whose decisions about which outcomes to include may be impacted by implicit biases and cause unnecessary heterogeneity in the outcomes used across trials. The systematic review which informed this work assessed what outcomes are used and how they were measured in all ABI trials since 2000 across all settings. Briefly, in 405 eligible trials (out of 33,134 studies screened), 2,641 outcomes were reported, measured in approximately 1,560 different ways (Shorter et al., under review). As every researcher has the opportunity to select from a range of outcomes, better standardization of the minimum requirement to measure change will maximize the potential of ABI research to influence decision-making, as it has in other research areas such as eczema or rheumatoid arthritis (Boers, 1994; Schmitt et al., 2011).

To achieve improved standardization of the outcomes used in ABI trials, the Outcome Reporting in Brief Intervention Trials: Alcohol (ORBITAL) project (Shorter et al., 2017) is working to establish a COS for ABIs using COMET procedures. Endorsed by the International Network on Brief Interventions for Alcohol and Other Drugs (INEBRIA) and with oversight by the INEBRIA Research Measurement Standardization Special Interest Group (RMS-SIG), ORBITAL undertook three, inter-related efforts to establish an ABI COS (Shorter et al, 2017). The first was a comprehensive systematic review to determine what outcomes are reported in ABI trials (Shorter et al., under review). The second, and the focus of this paper, was an international, multi-perspective e-Delphi consensus study to prioritize outcomes for use as a minimum set of reported outcomes in all ABI efficacy and effectiveness trials. The third step in the COMET process is to recommend a final set of specific measures using criteria recommended by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative (Mokkink et al., 2016). Although ORBITAL is developing the COS as a minimum data standard for ABI trials, trials can use other measures alongside the COS as appropriate. Based on the COMET methodology (Williamson et al., 2017), ORBITAL places importance on involving a wide range of stakeholders in the development of the COS, and in particular the client group considered to benefit most from an efficacious or effective intervention. As reported in this paper, we used an e-Delphi approach to understand what outcomes are priorities for ABI stakeholders in addition to prioritizing outcome domains for use in ABI trials.

Method

We conducted an international, multi-stakeholder e-Delphi consensus study by generating a list of relevant outcomes which participants were asked to rate in two successive rounds. The e-Delphi approach is an iterative consensus technique which presents a series of sequential questionnaires asking individuals to rank outcomes in terms of priority for

inclusion in a COS for ABI efficacy or effectiveness trials. Ethical approval was granted by the School of Health and Social Care Ethics Committee at Teesside University (reference: 018/17).

The e-Delphi is an online implementation of the Delphi approach for consensus building (Hasson & Keeney, 2011). The e-Delphi approach solicits the opinions of thought-leaders and experts on a particular topic in successive rounds, with each round providing input into the next (Sinha et al., 2011). All Delphi studies use at least two rounds, but some use three or more. E-Delphi panelists are informed of the results of prior rounds and allowed to revise their opinion based on those results. The goal is to achieve some pre-defined threshold of consensus. Key to the e-Delphi approach is the anonymity of panelists. By ensuring panel members remain anonymous throughout the process, panelists are free to revise their opinion without fear of reputational harm or to refuse to revise their opinion without pressure from the group to do so (Hasson, Keeney & McKenna, 2000).

Participants

There are no accepted guidelines for panel size to achieve stable consensus in an e-Delphi study. As such, we were guided by practicality, scope, and time available (Blackwood et al., 2015). Consistent with the purposive or criterion sampling approach used by many Delphi studies (Hasson et al, 2000), our sampling strategy focused on identifying electronic forums used by the relevant stakeholders and then allowing the sample to evolve organically as stakeholders shared the invitation to participate. Our sampling approach is best described as a purposive, snow-ball sampling approach. Participants were recruited in the following ways: emails to relevant mailing lists of researchers, practitioners, and policy-makers in the field, such as the INEBRIA google group; emails to corresponding authors of ABI trials identified by the systematic review; a tweet circulated on the @teamalphatees at Teesside

University; and emails forwarded by recruited participants to additional contacts with relevant expertise.

Participants were recruited between July 4th and August 1st 2017. Those recruited included trial investigators, INEBRIA RMS-SIG members, executive leadership of scientific organizations, Cochrane Review Group on Drugs and Alcohol members, NICE alcohol use disorder prevention PH24 membership group, trialists, statisticians, COS developers, service users/patient and public representatives, practitioners, groups involved in developing ABI clinical guidance, funders, and research ethics committee members (Shorter et al., 2017). Participants often held multiple roles. Consistent with COMET methodology, we included researchers, policymakers, and service users/patients in our sample of experts to ensure a broad representation of opinions. The patient perspective is particularly important to the COMET methodology since user input into this kind of study adds the lived experience of the alcohol consumer. Outcomes perceived to be relevant by stakeholders further removed from the user experience can appear less relevant to users (Henihan et al., 2015; Henihan et al., 2016). Also, users can suggest outcomes not immediately apparent as important to researchers. Recruitment text and round instructions are available from the corresponding author. To minimize attrition, recruitment text stressed the importance of completing both rounds.

Given the organic and evolving nature of our recruitment procedure, it is impossible to say how many individuals received an invitation to participate during the window of recruitment. For example, the INEBRIA Google group has 653 members, but not all members actively monitor the group. Among those members that do monitor the group, we cannot determine, nor could we have monitored, how many members forwarded the invitation to colleagues. Similarly, we have no way of tracking how many of the 458 @teamalphatees twitter followers saw the e-Delphi invitation. Furthermore, the anonymous

nature of the e-Delphi made it impossible to track the acceptance rate of the approximately 250 invitation emails we sent in any systematic way. Thus, we cannot provide a response rate in a traditional sense. The relevance of a traditional response rate is unclear, however, given the purposive, anonymous, snow-ball sampling approach we used. Therefore, rather than focus on a response rate per se, we monitored the number of panelists in each of the three basic stakeholder types: researcher, policymaker, and patient. We did not attempt to "balance" the participants across types but rather tried to ensure a sufficient number of each type.

The Delphi questionnaire and rounds

The e-Delphi used a bespoke online e-management system 'DelphiManager', maintained by the COMET initiative to facilitate core outcome set development. In both rounds, participants scored each outcome using the Grading of Recommendations Assessment, Development, and Evaluations (GRADE) scale of 1-9, with 1–3 labelled 'not important for inclusion', 4–6 labelled 'important but not critical' and 7–9 labelled 'critical for inclusion' (Guyatt et al., 2011). Outcomes were derived from the first 100 papers in a systematic review of existing ABI effectiveness and efficacy trials (Shorter et al, under review; Registered at PROSPERO, CRD42016047185; Shorter et al., 2016). These papers were not randomly selected but did represent a range of ABI settings and the full spectrum of years from 2000-2016. Current or former hazardous drinkers (n=9) formed a patient and public involvement panel, some from an established service user representative group (Belfast Experts by Experience), and others known to the lead author as drinking hazardously or above (and not researchers, clinicians, or members of other professional groups related to drinking or other addictive behaviors). The hazardous drinking individuals on this panel (n=5) were recruited through personal invitation from the lead author, and were verified as hazardous drinkers by an AUDIT score of eight or more. The patient and public involvement

panel added additional outcomes to the questionnaire at Round 1. Every outcome was given a descriptor. The outcomes were discussed and refined for clarity by the patient and public involvement panel and the authors.

In Round 1, there were 86 outcomes presented to participants. Participants could add additional outcomes and comment on the reason for their outcome ranking. Suggested outcomes from Round 1 were reviewed and coded to determine their novelty (i.e. that they were not covered by existing outcomes in the questionnaire). The additional outcomes and decisions made can be seen in supplementary material A. Round 2 included the 86 original outcomes, the seven additional outcomes, the individual's personal ranking, and rankings grouped by stakeholder group (researchers, healthcare and other professionals, and service users/representatives). Round 1 and Round 2 both used the same GRADE ranking system. All those who registered in Round 1 were invited to take part, with Round 2 closing on September 12th 2017. Consensus was defined *a priori* (Shorter, et al., 2017) as 70% or more of the respondents scoring an outcome from seven to nine and fewer than 15% scoring it one to three (Blackwood et al., 2015; Eleftheriadou et al., 2015). This would illustrate an outcome agreed critically important by the majority and little or no importance by a small minority (Shorter, et al., 2017). Although there is no formal guidance for the reporting of e-Delphi studies, we followed recommendations by Sinha et al. (2011). Participants received no financial incentive to participate.

Results

There were 150 total registrants. Overall, 137 took part in at least one question in Round 1 (including five partial completions) and 114 took part in at least one question in Round 2 (including 10 partial completions) – referred to as participants. In total, 107 took part in at least one question in both rounds, 30 completed Round 1 only, seven completed

Round 2 only, and seven registered but did not complete either round. A single person's response contributed between 1.1%- 0.7% (Round 1), and 1.4%- 0.9% (Round 2) to a percentage total (variability range includes missing data or 'prefer not to answer'). As noted in a recent systematic review (Boulkedid et al., 2011), few Delphi studies report response rates for all rounds, so it is difficult to determine if our rate of attrition from Round 1 to Round 2 is typical. This same review found the median number of invited participants in Delphi studies was only 17. Thus, we conclude that our sample size is more than sufficient and since our sample is intentionally purposive, not representative, we also conclude that any attrition from Round 1 to Round 2 is not problematic.

Details of participants/registrants are given in Table 1. The largest proportions of respondents were researchers, female, and from the UK or USA. Because ABI "patients" are most often hazardous drinkers who are not treatment seeking, and often do not consider themselves to be alcohol patients, this group was the most problematic to identify and recruit, and consequently had the lowest representation across participant types. In total, participants were from 19 countries (several noted 'other' but without stating country name). The majority had been involved in at least one ABI trial (70.7%) and around a quarter had been involved in four or more. Most participants had no experience of reviews of ABIs or of developing measurement instruments (59.3% and 60.7% respectively). In addition, the majority had no experience with core outcome set development (71.3%). Of those with previous experience, most had been involved in developing one core outcome set. The majority of ABI trial experience was in a healthcare setting: 38.0% had experience in alcohol or drug treatment settings; 36.7% in primary care; outpatient and inpatient care both had 33.3% each; and 31.3% in emergency care settings.

The ranking of consumption measures is given in Table 2. Based on Round 1 ranking, four met the 70% threshold. On review by participants in Round 2, seven met this criterion.

These were: typical frequency; frequency of heavy drinking; number of drinks in a week; hazardous or harmful drinking; alcohol-related problems; combined consumption measure; and typical quantity. There was least change in views on alcohol-related problems, with an increase of 0.8% in those ranking this outcome 'critical for inclusion' between rounds. By contrast, the largest increase in those ranking 'critical for inclusion' was in the typical quantity outcome which increased by 15%.

Rankings of the remaining domains are given in Table 3. Biomarkers were typically under-ranked, with a higher proportion selecting 'unable to score' than in any other domain. However, of those that were ranked, the highest ranked were levels of Phosphatidylethanol and Alanine aminotransferase, but none met either threshold for scores in the 'critical for inclusion' range or 'not important for inclusion' range. In the resource use and economic factors domain, none met the 70% threshold for those in the 'critical for inclusion' range. However, four met the lowered 60% threshold. These were: alcohol-related injury; use; alcohol or drug treatment; emergency healthcare; and hospitalization. In the life impact domain, the highest ranked outcome was quality of life. This outcome was ranked 79.4% in the 'critical for inclusion' range at the end of Round 2, reflecting an increase of 16% from the corresponding range in round 1. Only one of the health domain outcomes met the lowered criterion of 60% in the 'critical for inclusion' range: psychological or mental health (64.7% in the 'critical for inclusion' range at Round 2). Only one item from the psychological factors domain met the 70% threshold of 'critical for inclusion' range. This was 'interest in making changes around alcohol use'. Finally, only one item in the intervention factors domain was ranked as 'critical for inclusion'. This related to whether the intervention was delivered as planned. The ranking for this item was 81.4% in the 'critical for inclusion' range in Round 2.

Discussion

Given that ABIs are a key component of alcohol policies worldwide, it is vital that policy makers, service commissioners, and practitioners are able to access and synthesize robust, consistent evidence to inform their implementation (Babor et al., 2007; Bernstein et al., 2010). A key factor currently impeding existing evidence synthesis efforts is a lack of standardized outcomes used in ABI trials (Shorter, et al., 2017). As seen in other fields, standardization of outcomes will improve the ability of others to synthesize and evaluate the literature. Thus, the COMET Initiative has developed a formal, multi-phase methodology (Williamson et al., 2012; Williamson et al., 2017) that researchers can use to establish a core outcome set (COS).

As part of the larger, multi-phase ORBITAL project endorsed by INEBRIA, this study is one step in establishing a COS for ABI trials using the COMET methodology. ORBITAL aims to simplify and inform future ABI trial decision-making (Daykin et al., 2016; Daykin et al., 2017) and move beyond individual trial researcher preference as the primary vehicle by which outcomes are chosen to one of consensus between stakeholders (Williamson et al., 2017). This study presents the results of the ORBITAL e-Delphi study and is the first attempt to seek international, multi-stakeholder perspectives on which outcomes should be prioritized for ABI trials.

The results of our e-Delphi study suggest that considerable standardization of outcomes used in the ABI trials is possible. A systematic review conducted as part of the larger ORBITAL effort (Shorter et al., under review) found that 2,641 outcomes, measured in approximately 1,560 different ways, were reported in ABI trials, suggesting enormous variability in the outcomes that the ABI research community prioritize. Yet our e-Delphi study found only nine outcomes met our *a priori* consensus threshold, seven of which were related to alcohol consumption. Relaxing our *a priori* threshold resulted in an additional five outcomes, four of which were related to healthcare use. Thus, our e-Delphi study suggests

that much of the variation in the outcomes used in the ABI literature is driven by idiosyncratic decisions by individual researchers regarding the specific outcomes for any given trial rather than by a fundamental diversity of relevant outcome domains. If this conclusion is correct, then the ORBITAL effort to develop an ABI COS will greatly improve the ability of ABI researchers to provide consistent, policy-relevant evidence across studies on the outcomes they view as most important.

The validity of this conclusion, however, depends on the composition of our e-Delphi panel. As recommended by current best practice guidelines, we included a diverse set of panelists in our e-Delphi (Blackwood, et al., 2015). Participants were from a range of countries (19 countries across six continents) and stakeholder groups (researchers, policymakers, and service users/patients) in order to capture a broad range of perspectives. Most were from the UK or the USA, however, and participants from South America and Asia were under-represented. We must therefore be cautious about the cultural relevance of prioritized outcomes in these locations (Hula et al., 2014). Furthermore, most panelists were researchers, which may have over-represented the consensus views of ABI researchers compared to other vital perspectives such as those of healthcare professionals, policymakers, and patient or public representatives. Diverse perspectives are likely to result in wider acceptance of the prioritized outcomes deemed critical to include in ABI studies, although we note priorities may differ in different participant groups and ABI settings (Hula et al., 2014).

For example, despite a wide range of critical outcomes identified by the panel, no critical outcomes were identified in the biomarkers domain. We can only speculate as to why no biomarker measure made it to the critical measure threshold. It may be that biomarker measures were less well understood by our online Delphi participants. However, it is also important to note they are less commonly reported in ABI trials (Kypri, 2007). This may be because biomarkers are generally considered more relevant to dependence, have poor

sensitivity and/or specificity, or are inconvenient to use in comparison to self-report (Allen & Litten, 2003; Babor et al., 2000). Despite the lack of biomarker measures, our Delphi study identified outcomes across six domains, broadening the types of outcomes typically considered by any given ABI trial, while at the same time offering the possibility of standardizing outcomes across studies. This broadening highlights the importance of selecting outcome measures based on a consensus of the field rather than simply relying on what has been measured in prior research (Sinha et al., 2011).

Although our online Delphi study is the most rigorous attempt to identify the appropriate outcome measures for ABI trials thus far, it is subject to some limitations. There is ambiguity as to what constitutes consensus (Sinha et al., 2011) and so our *a priori* choice of 70% agreement is subject to possible criticism. Although we attempted to balance perspectives within our Delphi panel, difficulties in recruiting some participant types, particularly policymakers and patients, may have skewed the overall panel recommendations towards a researcher perspective. Similarly, the predominance of English-speaking countries among our panelists, especially the UK and the USA, may also have influenced our results and suggests caution with regard to the generalizability of our findings to non-English speaking and to low- or middle-income countries. Finally, given the nature of recruitment into our Delphi panel, it is not possible to determine the true response rate to our Round 1 invitation. This, combined with attrition between Rounds 1 and 2, may limit the validity of our results. Our use of anonymous voting and the diverse composition of our panel, however, adds to what is known about outcome priorities in the ABI field.

This study is the first attempt to identify outcomes using consensus methods for consideration in a core outcome set in ABI trials. It prioritizes outcomes that are most important to a range of key stakeholders in the field and will help guide researchers in choosing outcomes in future trials. The items prioritized here will be useful to improve

evidence synthesis in future systematic reviews in the field. However, the prioritization of these outcomes is a dynamic rather than fixed process. More research is needed to: a) further prioritize these outcomes into a core outcome set for all trials of ABIs ; b) replicate this priority list over time and in under-represented groups; c) identify the best measures to represent these outcomes; and d) to determine if the adoption of these recommended outcomes improves standards in the field. The ORBITAL project, with oversight from the INEBRIA RMS-SIG, is pursuing these next steps to fulfill its charge of developing a consensus-based ABI COS to help drive the future of ABI research.

Acknowledgements

We would like to thank and acknowledge the funding from Alcohol Research UK (Research Innovation Grant Number: R2016/04) and the INEBRIA Research Measurement Standardization Special Interest Group for feedback and useful commentary at the INEBRIA 2016 conference workshop on the topic. We are grateful to patient and public representatives who gave their views on the proposed work.

References

Allen, J. P., & Litten, R. Z. (2003). Recommendations on Use of Biomarkers in Alcoholism Treatment Trials. *Alcoholism: Clinical and Experimental Research, 27*(10), 1667-1670.

Babor, T. F., McRee, B. G., Kassebaum, P. A., Grimaldi, P. L., Ahmed, K., & Bray, J. (2007). Screening, Brief Intervention, and Referral to Treatment (SBIRT): toward a public health approach to the management of substance abuse. *Substance Abuse, 28*(3), 7-30.

Babor, T. F., Steinberg, K., Anton, R., & Del Boca, F. (2000). Talk is cheap: measuring drinking outcomes in clinical trials. *Journal of studies on alcohol, 61*(1), 55-63.

Bernstein, J. A., Bernstein, E., & Heeren, T. C. (2010). Mechanisms of change in control group drinking in clinical trials of brief alcohol intervention: Implications for bias toward the null. *Drug and Alcohol Review, 29*(5), 498-507.

Blackwood, B., Ringrow, S., Clarke, M., Marshall, J., Rose, L., Williamson, P., & McAuley, D. (2015). Core Outcomes in Ventilation Trials (COVenT): protocol for a core outcome set using a Delphi survey with a nested randomised trial and observational cohort study. *Trials, 16*, 368-374.

Boers, M. (1994). World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. *J Rheumatol Suppl, 41*.

Boulkedid, R., Abdoul, H., Loustau, M., Sibony, O., & Alberti, C. (2011). Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PloS one*, *6*(6), e20476.

Chan, A. W., Tetzlaff, J. M., Altman, D. G., Laupacis, A., Gotzsche, P. C., Krleza-Jeric, K., . . . Moher, D. (2013). SPIRIT 2013 statement: defining standard protocol items for clinical trials. Annals of Internal Medicine, 158(3), 200-207.

Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

Cunningham, J. A., Shorter, G. W., Murphy, M., Kushnir, V., Rehm, J., & Hendershot, C. S. (2017). Randomized Controlled Trial of a Brief Versus Extended Internet Intervention for Problem Drinkers. International Journal of Behavioral Medicine, 24, 760-767

Davoren, M. P., Demant, J., Shiely, F., & Perry, I. J. (2016). Alcohol consumption among university students in Ireland and the United Kingdom from 2002 to 2014: a systematic review. *BMC Public Health, 16*, 173. doi:10.1186/s12889-016-2843-1

Daykin, A., Selman, L. E., Cramer, H., McCann, S., Shorter, G. W., Sydes, M. R., . . . Shaw, A. J. T. (2017). 'We all want to succeed, but we've also got to be realistic about what is happening': an ethnographic study of relationships in trial oversight and their impact. 18(1), 612.

Daykin, A., Selman, L. E., Cramer, H., McCann, S., Shorter, G. W., Sydes, M. R., . . . Shaw, A. J. T. (2016). What are the roles and valued attributes of a Trial Steering Committee? Ethnographic study of eight clinical trials facing challenges. 17(1), 307.

Eleftheriadou, V., Thomas, K., van Geel, N., Hamzavi, I., Lim, H., Suzuki, T., . . . Vitiligo Global Issues Consensus, G. (2015). Developing core outcome set for vitiligo clinical trials: international e-Delphi consensus. *Pigment Cell Melanoma Res, 28*(3), 363-369.

Gargon, E., Williamson, P. R., Altman, D. G., Blazeby, J. M., Tunis, S., & Clarke, M. (2017). The COMET Initiative database: progress and activities update (2015). *Trials*, *18*(1), 54.

Guyatt, G. H., Oxman, A. D., Kunz, R., Atkins, D., Brozek, J., Vist, G., . . . Schunemann, H. J. (2011). GRADE guidelines: 2. Framing the question and deciding on important outcomes. *Journal of Clinical Epidemiology, 64*(4), 395-400.

Hasson, F., & Keeney, S. (2011). Enhancing rigour in the Delphi technique research. Technological Forecasting and Social Change, 78(9), 1695-1704.

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of advanced nursing*, *32*(4), 1008-1015.

Henihan, A. M., Klimas, J., Bury, G., O'Toole, T., Rieckman, T., Shorter, G., & Cullen, W. (2015). Methodological challenges and issues of recruiting for mental health and substance use disorders trials in primary care. Addiction science & clinical practice, 10(1), A21.

Henihan, A. M., McCombe, G., Klimas, J., Swan, D., Leahy, D., Anderson, R., . . . Cullen, W. (2016). Feasibility of alcohol screening among patients receiving opioid treatment in primary care. BMC Family Practice, 17(1), 153.

Hula, W. D., Fergadiotis, G., & Doyle, P. J. (2014). A core outcome set for aphasia treatment research: Obstacles, risks, and benefits. *Aphasiology, 28*(11), 1396-1399.

Kaner, E. F. S., Beyer, F. R., Muirhead, C., Campbell, F., Pienaar, E. D., Bertholet, N., . . . Burnand, B. (2018). Effectiveness of brief alcohol interventions in primary care populations. *Cochrane Database of Systematic Reviews*(2), CD004148.

Khadjesari, Z., Murray, E., Hewitt, C., Hartley, S., & Godfrey, C. (2011). Can stand-alone computer-based interventions reduce alcohol consumption? A systematic review. *Addiction, 106*(2), 267-282.

Kypri, K. (2007). Methodological issues in alcohol screening and brief intervention research. *Substance Abuse, 28*(3), 31-42.

Moher, D., Schulz, K. F., & Altman, D. G. (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *The Lancet, 357*(9263), 1191-1194.

Mokkink, L. B., Prinsen, C. A., Bouter, L. M., de Vet, H. C., & Terwee, C. B. (2016). The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Brazilian journal of physical therapy*, *20*(2), 105-113.

National Institute for Health and Clinical Excellence. (2010). NICE Public Health (PH) Guideline 24: Alcohol use disorders – preventing harmful drinking. Retrieved from http://www.nice.org.uk/PH24

Prinsen, C. A., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., . . . Terwee, C. B. (2016). How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. Trials, 17(1), 449.

Schmitt, J., Langan, S., Stamm, T., & Williams, H. C. (2011). Core outcome domains for controlled trials and clinical recordkeeping in eczema: international multiperspective Delphi consensus process. *J Invest Dermatol, 131*(3), 623-630.

Shorter, G. W., Newbury- Birch, D., Heather, N., Giles, E. L., Holloway, A., Bray, J. W., . . . O'Donnell, A. J. (2016). Systematic review to identify and appraise outcome measures and domains used in trials evaluating alcohol screening and brief interventions: the Outcome Reporting in Brief Intervention Trials (ORBIT) project review. Retrieved from https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=47185

Shorter, G. W., Heather, N., Bray, J. W., Giles, E. L., Holloway, A., Barbosa, C., . . . Newbury-Birch, D. (2017). The 'Outcome Reporting in Brief Intervention Trials: Alcohol' (ORBITAL) framework: protocol to determine a core outcome set for

efficacy and effectiveness trials of alcohol screening and brief intervention. *Trials, 18*(1), 611.

Shorter, G. W., Bray, J. W., Giles, E. L., Heather, N., Berman, A. H., O'Donnell, A. J., Holloway, A., Barbosa, C., Scott, S. J., Stockdale, K. J., Clarke, M., Newbury-Birch, D. (Under review). The variability of outcomes used in efficacy and effectiveness trials of alcohol brief interventions: A systematic review. *Addiction*.

Sinha, I., Smyth, R., & Williamson, P. (2011). Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. *PLoS Med, 8*(1), e1000393.

White, A., Kavanagh, D., Stallman, H., Klein, B., Kay-Lambkin, F., Proudfoot, J., . . . Young, R. (2010). Online Alcohol Interventions: A Systematic Review. *Journal of Medical Internet Research, 12*(5), e62. doi:10.2196/jmir.1479

Williamson, P. R., Altman, D. G., Bagley, H., Barnes, K. L., Blazeby, J. M., Brookes, S. T., . . . Young, B. (2017). The COMET Handbook: version 1.0. *Trials, 18*(Suppl 3), 280. doi:10.1186/s13063-017-1978-4

Williamson, P. R., Altman, D. G., Blazeby, J., Clarke, M., & Gargon, E. (2012). Driving up the quality and relevance of research through the use of agreed core outcomes. *Journal of health services research & policy, 17*(1), 1-2.

Williamson, P. R., & Clarke, M. (2012). The COMET (Core Outcome Measures in Effectiveness Trials) Initiative: Its Role in Improving Cochrane Reviews. *Cochrane Database Syst Rev*(5), ED000041.

Tables

Table 1: Characteristics of registrants, and the participants in e-Delphi survey rounds 1 and 2

	Registrants n=150	**Round 1 N=137**	**Round 2 N=114**
Stakeholder group			
Researchers	91 (60.7%)	87 (63.5%)	72 (63.2%)
Healthcare and other professionals (including policy makers)	50 (33.3%)	42 (30.7%)	34 (29.8%)
Patient/Public involvement panel	9 (6.0%)	8 (5.8%)	8 (7.0%)
Gender			
Male	63 (42.0%)	61 (44.5%)	49 (43.0%)
Female	85 (56.7%)	75 (54.7%)	64 (56.1%)
Trans*	1 (0.7%)	1 (0.7%)	1 (0.9%)
Rather not say	1 (0.7%)	0 (0.0%)	0 (0.0%)
Country			
Argentina	1 (0.7%)	1 (0.7%)	1 (0.9%)
Australia	3 (2.0%)	3 (2.2%)	3 (2.6%)
Brazil	1 (0.7%)	1 (0.7%)	1 (0.9%)
Canada	2 (1.3%)	2 (1.5%)	2 (1.8%)
Denmark	1 (0.7%)	1 (0.7%)	1 (0.9%)
France	1 (0.7%)	1 (0.7%)	1 (0.9%)
Germany	3 (2.0%)	2 (1.5%)	2 (1.8%)
Ireland	1 (0.7%)	1 (0.7%)	1 (0.9%)
Italy	2 (1.3%)	2 (1.5%)	1 (0.9%)
Mexico	1 (0.7%)	0 (0.0%)	0 (0.0%)
Netherlands	4 (2.7%)	4 (2.9%)	2 (1.8%)
Other (not specified)	3 (2.0%)	3 (2.2%)	3 (2.7%)
Portugal	1 (0.7%)	1 (0.7%)	1 (0.9%)
Spain	1 (0.7%)	1 (0.7%)	1 (0.9%)
Sweden	3 (2.0%)	3 (2.2%)	2 (1.8%)
Switzerland	7 (4.7%)	7 (5.1%)	6 (5.3%)
Thailand	2 (1.3%)	2 (1.5%)	1 (0.9%)
Uganda	1 (0.7%)	1 (0.7%)	1 (0.9%)
United Kingdom	68 (45.3%)	59 (43.1%)	48 (42.1%)
United States of America	44 (29.3%)	42 (30.7%)	36 (31.6%)
# alcohol brief intervention trials you have been involved in			
0	45 (29.3%)	37 (27.0%)	30 (26.3%)
1	22 (14.7%)	18 (13.1%)	15 (13.2%)
2	30 (20.0%)	30 (21.9%)	27 (23.7%)
3	15 (10.0%)	14 (10.2%)	12 (10.5%)
4+	39 (26.0%)	38 (27.7%)	30 (26.3%)
# systematic reviews which include alcohol brief intervention trials you have been involved in			
0	89 (59.3%)	80 (58.4%)	67 (58.8%)
1	26 (17.23)	23 (16.8%)	20 (17.5%)
2	18 (12.0%)	17 (12.4%)	16 (14.0%)
3	9 (6.0%)	9 (6.6%)	7 (6.1%)
4+	8 (5.3%)	8 (5.8%)	4 (3.5%)
# measurement instruments developed which could be used as an outcome in an alcohol brief intervention trial			
0	92 (60.7%)	82 (59.9%)	67 (58.8%)
1	34 (22.7%)	30 (21.9%)	26 (22.8%)

2	12 (8.0%)	12 (8.8%)	10 (8.8%)
3	6 (4.0%)	6 (4.4%)	5 (4.4%)
4+	7 (4.7%)	7 (5.1%)	6 (5.3%)
# core outcome sets you have been involved in developing (including as a Delphi participant)			
0	107 (71.3%)	97 (70.8%)	83 (72.8%)
1	20 (13.3%)	20 (14.6%)	16 (14.0%)
2	9 (6.0%)	8 (5.8%)	5 (4.4%)
3	3 (2.0%)	3 (2.2%)	3 (2.6%)
4+	11 (7.3%)	9 (6.6%)	7 (6.1%)
In which setting do you have alcohol brief intervention experience			
Alcohol or Drug Treatment	57 (38.0%)	51 (37.2%)	42 (36.8%)
Criminal Justice (Prison, Probation, or other setting)	22 (14.7%)	21 (15.3%)	17 (14.9%)
Emergency care (e.g. Trauma Center, Emergency Room)	47 (31.3%)	45 (32.8%)	35 (30.7%)
Inpatient hospital care (not alcohol/drug treatment/emergency care)	50 (33.3%)	47 (34.3%)	37 (32.5%)
Outpatient hospital care (e.g. sexual health clinic)	50 (33.3%)	46 (33.6%)	37 (32.5%)
Mobile based brief interventions	28 (18.7%)	27 (19.7%)	23 (20.2%)
Online web based brief interventions	42 (28.0%)	40 (29.2%)	35 (30.7%)
Pharmacy or Drug Store brief interventions	8 (5.3%)	8 (5.8%)	4 (3.5%)
Primary Care (General Practice or Family Physician)	55 (36.7%)	52 (38.0%)	42 (36.8%)
Schools	21 (14.0%)	20 (14.6%)	16 (14.0%)
Universities and Colleges	43 (28.7%)	41 (29.9%)	33 (28.9%)
Veterans/military	2 (1.3%)	2 (1.5%)	2 (1.8%)
Workplaces (includes employee assistance programs/job centers)	24 (16.0%)	22 (16.1%)	14 (12.3%)
Licensed Premises	5 (3.3%)	5 (3.6%)	1 (0.9%)
General Population, community, or convenience samples	8 (5.3%)	8 (5.8%)	7 (6.1%)
Social Services	2 (1.3%)	2 (1.5%)	1 (0.9%)

Table 2: Consumption domain rankings from e-Delphi questionnaires in Rounds 1 and 2

	Round 1				Round 2			
	N	Scored 1-3 n (%)	Scored 4-6 n (%)	Scored 7-9 n (%)	n	Scored 1-3 n (%)	Scored 4-6 n (%)	Scored 7-9 n (%)
Consumption								
How often (frequency) a person drinks heavily or large number of drinks on occasion	136	3 (2.2%)	20 (14.7%)	113(83.1%)	105	1 (1.0%)	6 (5.7%)	98 (93.3%)
Total number of standard drinks consumed in a week	136	2 (1.5%)	28 (20.6%)	106(77.9%)	107	2 (1.9%)	8 (7.5%)	97 (90.7%)
How often (frequency) a person drinks alcohol	136	3 (2.2%)	32 (23.5%)	101(74.3%)	105	2 (1.9%)	12 (11.4%)	91 (86.7%)
Drinking at a level which puts you at risk of harm (hazardous or harmful drinking)	136	4 (2.9%)	39 (28.7%)	93 (68.4%)	104	1 (1.0%)	18 (17.3%)	85 (81.7%)
Alcohol related problems or consequences due to alcohol use	134	2 (1.5%)	28 (20.9%)	108(80.6%)	113	0 (0.0%)	21 (18.6%)	92 (81.4%)
Combined consumption measure which takes into consideration one or more consumption measures together (e.g. frequency, quantity, or frequency of heavy drinking together in one measure)	136	6 (4.4%)	38 (27.9%)	92 (67.6%)	104	4 (3.8%)	16 (15.4%)	84 (80.8%)
Typical number of drinks consumed in a drinking occasion	136	5 (3.7%)	55 (40.4%)	76 (55.9%)	102	2 (1.9%)	28 (27.2%)	73 (70.9%)
Abuse symptomatology (severity of the symptoms of alcohol abuse; excessive use)	135	10 (7.4%)	80 (59.3%)	65 (48.1%)	109	5 (4.6%)	49 (45.0%)	55 (50.5%)
Days abstinent: number of days in a period of time in which a person does not drink/abstains from alcohol	133	15 (11.3%)	64 (48.1%)	54 (40.6%)	108	12 (11.1%)	51 (47.2%)	45 (41.7%)
Dependence symptomatology (Severity of the symptoms of alcohol dependence; physical or psychological need to drink alcohol)	134	16 (11.9%)	102(76.1%)	48 (35.8%)	109	6 (5.5%)	65 (59.6%)	38 (34.9%)
Drinking above the government guidelines for low risk drinking in a given country	136	24 (17.6%)	61 (44.9%)	51 (37.5%)	102	18 (17.6%)	51 (50.0%)	33 (32.4%)
Number of drinks consumed in a month or other period	136	31 (22.8%)	63 (46.3%)	42 (30.9%)	104	23 (22.1%)	51 (49.0%)	30 (28.8%)
Largest number of drinks on occasion	133	21 (15.8%)	69 (51.9%)	43 (31.6%)	104	7 (6.7%)	68 (65.4%)	29 (27.9%)
The use of alcohol with another drug (e.g. tobacco or an illegal drug) at the same time or in the same time period	135	24 (17.8%)	76 (56.3%)	35 (25.9%)	104	15 (14.4%)	62 (61.4%)	26 (23.8%)
How often (frequency) a person drinks enough to feel drunk/intoxicated	136	22 (16.2%)	67 (49.3%)	47 (34.6%)	104	16 (15.4%)	67 (64.4%)	21 (20.2%)
Drinks consumed in the heaviest week of drinking in a given time period	136	18 (13.2%)	79 (58.1%)	39 (28.7%)	104	12 (11.5%)	76 (73.1%)	16 (15.4%)
Blood Alcohol Consumption (Levels of alcohol in the blood; may be measured using a breathalyzer; or calculated based on reports of alcohol consumption)	136	53 (39.0%)	60 (44.1%)	23 (16.9%)	108	42 (38.9%)	53 (49.1%)	13 (12.0%)
Matching goals set before drinking about how much alcohol you plan to drink or how long you plan to drink for	134	39 (29.1%)	71 (53.0%)	24 (17.9%)	104	23 (22.1%)	73 (70.2%)	8 (7.7%)
The time spent drinking alcohol	136	51 (37.5%)	68 (50.0%)	17 (12.5%)	103	50 (48.5%)	49 (47.6%)	4 (3.9%)
The type of drink consumed	136	68 (50.0%)	53 (39.0%)	15 (11.0%)	102	67 (65.7%)	31 (30.4%)	4 (3.9%)
If your drinking matches the report of someone who was there at the same time	135	73 (54.1%)	59 (43.7%)	3 (2.2%)	104	74 (71.2%)	29 (27.9%)	1 (1.0%)

How much alcohol a family member or partner drinks; or other people who are similar to the participant	135	73 (54.1%)	53 (39.3%)	9 (6.7%)	104	67 (64.4%)	37 (35.6%)	0 (0.0%)

Table 3: Rankings from e-Delphi rounds 1 and 2 in the biomarkers, economic factors/resource use, health, psychological factors, intervention factors, and life impact domains

	Round 1				Round 2			
	n	Scored 1-3 N (%)	Scored 4-6 N (%)	Scored 7-9 N (%)	N	Scored 1-3 N (%)	Scored 4-6 N (%)	Scored 7-9 N (%)
Biomarkers								
Phosphatidylethanol (PETH) suggested by three individuals and scored 8; 7; and 4	N/A – not included in Round 1				76	20 (26.3%)	39 (51.3%)	17 (22.4%)
Alanine aminotransferase	91	25 (27.5%)	51 (56.0%)	15 (16.5%)	81	20 (24.7%)	48 (59.3%)	13 (16.0%)
Gamma-glutamyltransferase	93	28 (30.1%)	48 (51.6%)	17 (18.3%)	79	23 (29.1%)	44 (55.7%)	12 (15.2%)
Analyzing hair for ethyl-glucuronide suggested by two participants scoring 4, and 7.	N/A – not included in Round 1				79	32 (40.5%)	37 (46.8%)	10 (12.7%)
Aspartate aminotransferase	89	26 (29.2%)	52 (58.4%)	11 (12.4%)	78	23 (29.5%)	46 (59.0%)	9 (11.5%)
Mean corpuscular volume (MCV)	92	34 (37.0%)	46 (50.0%)	12 (13.0%)	78	30 (38.5%)	40 (51.3%)	8 (10.3%)
Levels of whole blood-associated acetaldehyde suggested by one participant and scored 6	N/A – not included in Round 1				74	24 (32.4%)	43 (58.1%)	7 (9.5%)
Carbohydrate-deficient transferrin	92	25 (27.2%)	48 (52.2%)	19 (20.7%)	79	20 (25.3%)	44 (55.7%)	15 (9.0%)
Economic Factors/Resource Use								
Alcohol related injury (physical injury as a result of alcohol use)	130	3 (2.3%)	44 (33.8%)	83 (63.8%)	102	5 (4.9%)	26 (25.5%)	71 (69.6%)
Use of drug/alcohol treatment in a healthcare setting or by a healthcare professional	130	11 (8.5%)	49 (37.7%)	70 (53.8%)	102	6 (5.9%)	32 (31.4%)	64 (62.7%)
Use of emergency healthcare services	130	12 (9.2%)	48 (36.9%)	70 (53.8%)	102	6 (5.9%)	34 (33.3%)	62 (60.8%)
Hospitalizations (inpatient healthcare services in a ward other than the emergency room)	129	8 (6.2%)	51 (39.5%)	70 (54.3%)	101	5 (5.0%)	35 (34.7%)	61 (60.4%)
Alcohol related driving offences/impaired driving (including drink driving or accidents)	131	9 (6.9%)	48 (36.6%)	74 (56.5%)	102	6 (5.9%)	40 (39.2%)	56 (54.9%)
Use of primary healthcare services (e.g. general practice/primary care/family physician)	130	15 (11.5%)	55 (42.3%)	60 (46.2%)	102	10 (9.8%)	37 (36.3%)	55 (53.9%)
Seeking help for alcohol or drugs not from a healthcare provider	128	19 (14.8%)	61 (47.7%)	48 (37.5%)	100	10 (10.0%)	44 (44.0%)	46 (46.0%)
General healthcare use (an overall measure of the use of healthcare services)	130	20 (15.4%)	63 (48.5%)	47 (36.2%)	101	9 (8.9%)	50 (49.5%)	42 (41.6%)
Alcohol related offences (may relate to the setting e.g. alcohol related violence or university rule violations)	131	10 (7.6%)	71 (54.2%)	50 (38.2%)	102	8 (7.8%)	56 (54.9%)	38 (37.3%)
General accident costs (not just alcohol related accidents)	129	31 (24.0%)	55 (43.4%)	42 (32.6%)	101	25 (24.8%)	53 (52.5%)	23 (22.8%)
General criminal justice costs (not those directly related to an alcohol offence)	129	36 (27.9%)	56 (43.4%)	38 (29.0%)	101	30 (29.7%)	50 (49.5%)	21 (20.8%)
Prescribed medication use (medication with a prescription from a Doctor)	129	34 (26.4%)	63 (48.8%)	32 (24.8%)	100	22 (22.0%)	59 (59.0%)	19 (19.0%)
Social service use (e.g. child protection; government sponsored unemployment support)	126	21 (16.7%)	76 (60.3%)	29 (23.0%)	100	21 (21.0%)	60 (60.0%)	19 (19.0%)
Use of pharmacies or drug store advice	129	46 (35.7%)	65 (50.4%)	18 (14.0%)	99	42 (42.4%)	50 (50.5%)	7 (7.1%)
Over the counter medication use	126	49 (38.9%)	64 (50.8%)	13 (10.3%)	99	39 (39.4%)	55 (55.6%)	5 (5.1%)
Life impact								
Quality of life (the standard of health/comfort/happiness experienced by an individual)	131	6 (4.6%)	42 (32.1%)	83 (63.4%)	102	2 (2.0%)	19 (18.6%)	81 (79.4%)
Alcohol causing harm to other people - recommended by one participant and scored 7	N/A – not included in Round 1				101	10 (9.9%)	55 (54.5%)	36 (35.6%)

Workplace or college/university productivity (such as the ability to work; or ability to meet deadlines or targets)	131	21 (16.0%)	72 (55.0%)	38 (29.0%)	102	14 (13.7%)	66 (64.7%)	22 (21.6%)
Ability to participate in society; or quality of relationships (e.g. within a family)	131	16 (12.2%)	74 (56.5%)	41 (31.3%)	102	8 (7.8%)	73 (71.6%)	21 (20.6%)
Satisfaction with social roles and activities	131	18 (13.7%)	76 (58.0%)	37 (28.2%)	102	13 (12.7%)	75 (73.5%)	14 (13.7%)
Improvement in finances (money available to spend on other things)- suggested by one participant and scored 6	N/A – not included in Round 1				102	28 (27.5%)	68 (66.7%)	6 (5.9%)
Health								
Psychological/mental health (unpleasant feelings which impact ability to live life)	131	4 (3.1%)	52 (39.7%)	76 (55.9%)	102	5 (4.9%)	31 (30.4%)	66 (64.7%)
Overall health or how healthy the person feels	131	7 (5.3%)	60 (45.8%)	64 (48.9%)	102	4 (3.9%)	44 (43.1%)	54 (52.9%)
Risk of alcohol withdrawal symptoms (like delirium tremens)	126	21 (16.7%)	70 (55.1%)	35 (27.8%)	100	13 (13.0%)	43 (43.0%)	44 (44.0%)
Mortality or death related to alcohol use - Recommended by one person and scored 8	N/A – not included in Round 1				93	18 (19.4%)	35 (37.6%)	40 (43.0%)
Physical health (ability to carry out physical activities from basic self-care to running)	130	10 (7.7%)	63 (48.5%)	57 (43.8%)	102	7 (6.9%)	57 (55.9%)	38 (37.3%)
Severity of the symptoms of depression or low mood	130	8 (6.2%)	64 (49.2%)	58 (44.6%)	102	6 (5.5%)	51 (59.6%)	38 (34.9%)
Suicidal ideas or beliefs	129	20 (15.5%)	72 (55.8%)	37 (28.7%)	101	15 (14.9%)	62 (61.4%)	24 (23.8%)
Severity of the symptoms of anxiety (feeling worried)	130	16 (12.3%)	73 (56.2%)	41 (31.5%)	102	13 (12.7%)	71 (69.6%)	8 (17.6%)
How often a person experiences a hangover (a range of unpleasant symptoms experienced after drinking alcohol which may include tiredness; thirst; nausea or vomiting; trouble sleeping; low mood; headache; anxiety or other aspects)	129	30 (23.3%)	75 (58.1%)	24 (18.6%)	102	19 (18.6%)	66 (64.7%)	17 (16.7%)
Problems sleeping (either too much or too little)	130	26 (20.0%)	77 (59.2%)	27 (20.8%)	101	21 (20.8%)	65 (64.4%)	15 (14.9%)
Number of medical conditions someone has (as diagnosed by a doctor)	129	29 (22.5%)	75 (58.1%)	25 (19.4%)	102	24 (23.5%)	64 (62.7%)	14 (13.7%)
Factors relating to heart health (such as blood pressure)	129	33 (25.6%)	73 (56.6%)	23 (17.8%)	101	29 (28.7%)	62 (61.4%)	10 (9.9%)
Factors relating to obesity (such as body mass index; body fat percentage)	129	34 (26.4%)	76 (58.9%)	19 (14.7%)	101	29 (28.7%)	66 (65.3%)	10 (9.9%)
Post-traumatic stress disorder symptoms (anxiety caused by events including upsetting memories; or sleep problems; or avoiding reminders of the event)	129	37 (28.7%)	74 (57.4%)	18 (14.0%)	102	28 (27.7%)	63 (62.4%)	10 (9.9%)
Quality of working partnership with healthcare provider	129	45 (34.9%)	63 (48.8%)	21 (16.3%)	101	34 (33.7%)	58 (57.4%)	9 (8.9%)
Problems with stomach or digestion (including abdominal pain/swelling/vomiting/nausea)	129	32 (24.8%)	78 (60.5%)	19 (14.7%)	101	28 (27.7%)	64 (63.4%)	9 (8.9%)
Problems with sex life	129	32 (24.8%)	83 (64.3%)	14 (10.9%)	101	33 (32.7%)	62 (61.4%)	6 (5.9%)
Psychological Factors								
Interest in making changes around alcohol use (motivation/readiness to change)	131	8 (6.1%)	51 (38.9%)	72 (55.0%)	102	2 (2.0%)	27 (26.5%)	73 (71.6%)
If alcohol is used to cope with stress; anxiety; or life events	130	12 (9.2%)	63 (48.5%)	55 (42.3%)	101	7 (6.9%)	50 (49.5%)	44 (43.6%)
If the participant believes their alcohol use affects their health	130	13 (10.0%)	61 (46.9%)	56 (43.1%)	102	13 (12.7%)	46 (45.1%)	43 (42.2%)
Engaging in protective behavioral strategies	128	6 (4.7%)	70 (54.7%)	52 (40.6%)	101	12 (11.9%)	49 (48.5%)	40 (39.6%)
Cravings or a powerful desire for alcohol	130	22 (16.9%)	55 (42.3%)	53 (40.8%)	102	11 (10.8%)	51 (50.0%)	40 (39.2%)
Self-efficacy or belief in ability to succeed/achieve goals	131	15 (11.5%)	57 (43.5%)	59 (45.0%)	102	9 (8.8%)	55 (53.9%)	38 (37.3%)

Ability to refuse alcohol (sometimes called drinking refusal self-efficacy or how able someone is to refuse alcohol in places it may be usually consumed)	130	19 (14.6%)	57 (43.8%)	54 (41.5%)	102	8 (7.8%)	59 (57.8%)	35 (34.3%)
Outcome expectancies; the belief that drinking leads to specific positive or negative outcomes or what a person expects to happen as a result of a given action	129	5 (3.9%)	61 (46.6%)	97 (75.2%)	102	14 (13.7%)	56 (54.9%)	32 (31.4%)
Alcohol's effect on the ability to reach goals (called goal striving)	129	21 (16.3%)	73 (56.6%)	35 (27.1%)	101	14 (13.9%)	58 (57.4%)	29 (28.7%)
Engaging in other risky behaviors (e.g. putting yourself in a dangerous place/situation)	130	22 (16.9%)	71 (54.6%)	37 (28.5%)	101	15 (14.9%)	62 (61.4%)	24 (23.8%)
Attitudes to alcohol consumption in pregnancy	130	24 (18.5%)	56 (43.1%)	50 (38.5%)	101	21 (20.8%)	57 (56.4%)	23 (22.8%)
Feeling supported (perhaps by family and friends)	130	31 (23.8%)	67 (51.5%)	32 (24.6%)	101	24 (23.8%)	60 (59.4%)	17 (16.8%)
How confident an individual feels/their self esteem	131	20 (15.3%)	71 (54.2%)	40 (30.5%)	102	17 (16.7%)	68 (66.7%)	17 (16.7%)
How positively or negatively alcohol is viewed by the participant	131	19 (14.5%)	78 (59.5%)	34 (26.0%)	102	16 (15.7%)	70 (68.6%)	16 (15.7%)
Empathy (ability to understand and share feelings of another)	126	54 (42.9%)	59 (46.8%)	13 (10.3%)	100	60 (60.0%)	35 (35.0%)	15 (15.0%)
Engaging in healthy behaviors such as exercise or healthy eating	130	17 (13.1%)	76 (58.5%)	37 (28.5%)	102	14 (13.7%)	74 (72.5%)	14 (13.7%)
How stressed out someone feels	130	23 (17.7%)	84 (64.6%)	23 (17.7%)	102	18 (17.6%)	70 (68.6%)	14 (13.7%)
Feeling alone or isolated	130	34 (26.2%)	70 (53.8%)	26 (20.0%)	101	27 (26.7%)	64 (63.4%)	10 (9.9%)
Aggression or anger (either feelings or actions)	130	28 (21.5%)	80 (61.5%)	22 (16.9%)	102	23 (22.5%)	69 (67.6%)	10 (9.8%)
Acting impulsively (acting without thinking/considering longer-term costs & benefits)	131	38 (29.0%)	78 (59.5%)	15 (11.5%)	102	44 (43.1%)	55 (53.9%)	3 (2.9%)
How positively other people's alcohol use is viewed	130	39 (30.2%)	75 (57.7%)	16 (12.3%)	102	39 (38.2%)	61 (59.8%)	2 (2.0%)
Seeking new and exciting experiences (interest in experiences which are new/exciting)	131	47 (35.9%)	74 (56.5%)	10 (7.6%)	102	59 (57.8%)	41 (40.2%)	2 (2.0%)
How stressed out someone feels	130	23 (17.7%)	84 (64.6%)	23 (17.7%)	102	18 (17.6%)	70 (68.6%)	14 (13.7%)
Intervention factors								
If the intervention was delivered as planned; or the participant used the intervention.	129	5 (3.9%)	27 (20.9%)	97 (75.2%)	102	4 (3.9%)	15 (14.7%)	83 (81.4%)
Satisfaction with intervention (the view of intervention from the person taking part)	131	2 (1.5%)	61 (46.6%)	68 (51.9%)	102	2 (2.0%)	41 (40.2%)	59 (57.8%)
Clinician satisfaction with intervention: suggested by one participant, rated 7	N/A – not included in Round 1				102	19 (18.6%)	65 (63.7%)	18 (17.6%)

Supplementary Material A: Additional outcomes suggested for Round 2 questionnaire with explanation of whether outcome added, covered by existing item, or not relevant

Suggested outcome with ranking in brackets	Matched existing Round 1 outcome if relevant	New outcome if relevant (with descriptor)	Narrative
Those added to the Round 2 questionnaire			
Mortality (8)		Mortality or Death	Not currently present
PETh (8)		Levels of Phosphatidylethanol in the blood	Not currently present
PETh (4)		(Phosphatidylethanol can be detected in the blood for up to three weeks and can indicate regular drinking, or heavy drinking.)	
EtG - Ethyl glucuronide on head hair (7)		Analyzing hair for ethyl-glucuronide (ethyl-glucuronide is present in hair up to 3/4 days after alcohol is consumed- it shows if someone has had an alcoholic drink recently)	Not currently present
Biomarkers: Hair analysis for ethyl-glucuronide(4)			
Clinician Satisfaction (7)		Clinician satisfaction (Clinician satisfaction with the alcohol brief intervention)	Not currently present
broader measures of harm to others e.g. domestic violence for economic outcomes (7)		Alcohol causing harm to other people often called 'harm to others' (The general impact that a person's alcohol has on other people than the drinker)	This was partially covered by other issues (such as alcohol related offences, or role and relationship factors but was considered a distinct composite outcome)
Improved social aspects – finance (6)		Improvement in finances (Changes in the amount of money a person has (either more or less) as a result of change in how much is being spent on alcohol)	Not currently present
WBAA (6)		Levels of whole blood-associated acetaldehyde (whole blood-associated acetaldehyde tests can detect heavy alcohol consumption through the presence of acetaldehyde (a by-product of alcohol consumption) for up to three weeks following use)	Not currently present

Suggested outcomes covered by existing outcomes

Concomitant use of Tobacco; Marijuana; Other Drug (8)	The use of alcohol with another drug	Was covered by existing outcome
Self-reported general health (8)	Overall health	Was covered by existing outcome
Patient adherence to the BI (8)	If the intervention was delivered as planned, or if the participant used the intervention	Was covered by existing outcome
Followed-up by attending/seeking further professional help. (7)	Use of drug or alcohol treatment services provided in a healthcare setting or by a healthcare professional	Was covered by existing outcome; timing is a measurement issue to be covered by guidance on how to measure an outcome if selected for core outcome set
Likelihood of seeking to change alcohol use in the future (6)	Interest in making changes around alcohol use	Was covered by existing outcome; timing is a measurement issue to be covered by guidance on how to measure an outcome if selected for core outcome set
Other sources of help (current) (6)	Seeking help for alcohol or drugs not from a healthcare provider	Was covered by existing outcome; timing is a measurement issue to be covered by guidance on how to measure an outcome if selected for core outcome set
Improved social aspects - employability or education (general) (6)	Workplace or college productivity	Was covered by existing outcome
Number of arrests (6);	General criminal justice costs	Was covered by existing outcome
Number of nights incarcerated (6)	General criminal justice costs	Was covered by existing outcome
Number of court appearances (6)	General criminal justice costs	Was covered by existing outcome

Number of days on community supervision (probation/parole) (societal cost driver/distinct from other criminal justice outcomes)(6)	General criminal justice costs	Was covered by existing outcome
How the person evaluates the impact of alcohol in their life (5)	Alcohol's effect on the ability to reach life goals	Was covered by existing outcome
Previous / history use of alcohol treatment services (5)	Use of drug or alcohol treatment services provided in a healthcare setting or by a healthcare professional	Was covered by existing outcome
DSM-V criteria (5)	Dependence/abuse symptomatology	As the core outcome set is directed to non-treatment seekers, this is partially covered by other outcomes but may not be relevant to the population
Not included in the questionnaire at Round 2		
Clinician decision to consider alcohol use in treatment of existing medical conditions. (6)		Not an indicator of BI effectiveness or efficacy
Clinician decision to consider alcohol use in prescribing medications. (6)		Not an indicator of BI effectiveness or efficacy
Number of brief interventions offered in the past (5)		Not an indicator of BI effectiveness or efficacy
Setting in which the brief intervention was offered (6)		Not an indicator of BI effectiveness or efficacy
Was the brief intervention facilitated or un-facilitated (6)		Not an indicator of BI effectiveness or efficacy
Was it an blended intervention (mix of face to face and on line) (6)		Not an indicator of BI effectiveness or efficacy

Improvement in housing (6)	Unsure of relevance to ABI studies, and was ranked below 7-9 “critical for inclusion” cut off.