The Variability of Outcomes Used in Efficacy and Effectiveness Trials of Alcohol Brief Interventions: A Systematic Review

Shorter, G. W., Bray, J. W., Giles, E. L., O'Donnell, A. J., Berman, A. H., Holloway, A., Heather, N., Barbosa, C., Stockdale, K. J., Scott, S. J., Clarke, M., & Newbury-Birch, D. (2019). The Variability of Outcomes Used in Efficacy and Effectiveness Trials of Alcohol Brief Interventions: A Systematic Review. *Journal of Studies on Alcohol and Drugs*, *80*(3), 286-298. https://doi.org/10.15288/jsad.2019.80.286

Published in:
Journal of Studies on Alcohol and Drugs

Document Version:
Publisher's PDF, also known as Version of record

The Variability of Outcomes Used in Efficacy and Effectiveness Trials of Alcohol Brief Interventions: A Systematic Review

GILLIAN W. SHORTER, B.SC., PH.D.,[a,b,*] JEREMY W. BRAY, B.A., M.A., PH.D.,[c] EMMA L. GILES, B.SC., PH.D.,[b] AMY J. O'DONNELL, B.SC., PH.D.,[d] ANNE H. BERMAN, B.SC., M.SC., PH.D.,[e,f] AISHA HOLLOWAY, B.SC., PH.D.,[g] NICK HEATHER, B.A., M.SC., PH.D.,[h] CAROLINA BARBOSA, PHARM.D., M.SC. PH.D.,[i] KELLY J. STOCKDALE, B.A., M.A., PH.D.,[j] STEPHANIE J. SCOTT, B.SC., M.A., PH.D.,[k] MIKE CLARKE, B.A., D.PHIL.,[l] & DOROTHY NEWBURY-BIRCH, B.A., PH.D.[k]

[a]*Institute of Mental Health Sciences, School of Psychology, Ulster University, Coleraine, Northern Ireland*

[b]*School of Health and Social Care, Teesside University, Middlesbrough, England*

[c]*Department of Economics, Bryan School of Business and Economics, University of North Carolina at Greensboro, Greensboro, North Carolina*

[d]*Institute of Health and Society, Newcastle University, Newcastle upon Tyne, England*

[e]*Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden*

[f]*Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden*

[g]*Nursing Studies, School of Health in Social Science, University of Edinburgh, Edinburgh, Scotland*

[h]*Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, England*

[i]*Behavioral Health Economics Program, RTI International, Chicago, Illinois*

[j]*School of Psychological and Social Sciences, York St. John University, York, England*

[k]*School of Social Sciences, Humanities and Law, Teesside University, Middlesbrough, England*

[l]*Northern Ireland Methodology Hub, Queen's University Belfast, Belfast, Northern Ireland*

ABSTRACT. Objective: The purpose of this study was to characterize recent alcohol brief intervention (ABI) efficacy and effectiveness trials, summarize outcomes, and show how variability in outcomes and reporting compromises the evidence base. **Method:** A systematic review and narrative synthesis of articles from 10 databases were undertaken (January 2000–November 2017); study selection represented recent, readily available publications. The National Institute of Care Excellence (NICE) Public Health Guideline 24 (Alcohol use disorders: prevention) informed ABI definitions. The review was conducted using Centre for Reviews and Dissemination (CRD) guidance and pre-registered on PROSPERO (CRD42016047185). Seven a priori specified domains were used to classify outcomes: biomarkers, alcohol-related outcomes, economic factors/resource use, health measures, life impact, intervention factors, and psychological/behavioral factors. **Results:** The search identified 405 trials from 401 eligible papers. In 405 trials, 2,641 separate outcomes were measured in approximately 1,560 different ways. The most common outcomes used were the number of drinks consumed in a week and frequency of heavy episodic drinking. Biomarkers were least frequently used. The most common primary outcome was weekly drinks. By trial type, the most frequent outcome in efficacy and effectiveness trials was frequency of heavy drinking. **Conclusions:** Consumption outcomes predominated; however, no single outcome was found in all trials. This comprehensive outcome map and methodological detail on ABI effectiveness and efficacy trials can aid decision making in future trials. There was a diversity of instruments, time points, and outcome descriptions in methods and results sections. Compliance with reporting guidance would support data synthesis and improve trial quality. This review establishes the need for a core outcome set (COS)/minimum data standard and supports the Outcome Reporting in Brief Interventions: Alcohol initiative (ORBITAL) to improve standards in the ABI field through a COS for effectiveness and efficacy randomized trials. (*J. Stud. Alcohol Drugs, 80,* 286–298, 2019)

ALCOHOL BRIEF INTERVENTIONS (ABIs) are key strategies to address problematic alcohol use worldwide (Coffield et al., 2001; National Institute for Health and Care Excellence [NICE], 2010; U.S. Preventive Services Task Force, 2004; World Health Organization [WHO], 2016). Many systematic reviews (Ballesteros et al., 2003, 2004a, 2004b; Beich et al., 2003; Bertholet et al., 2005; Kaner et al., 2018; O'Donnell et al., 2014) suggest that ABIs in primary health care are effective, but these reviews report substantial outcome heterogeneity, limiting the strength of conclusions. Commentators have urged caution in making broad clinical practice recommendations as a result (Bernstein et al., 2009, 2010; Field et al., 2010; Heather, 2016; McCambridge & Saitz, 2017; Saitz, 2010; Saitz et al., 2006).

An avoidable problem is the diversity in definition and measurement of outcomes used. This reduces the ability to synthesize information. For example, in a recent and comprehensive review (Kaner et al., 2018), authors excluded 22 of 69 otherwise eligible studies because of outcome reporting issues. Differing outcomes across studies weaken meta-analyses of the efficacy and effectiveness of ABIs and contribute to research waste because not all articles can

Received: January 23, 2019. Revision: April 11, 2019.

This research was supported by Alcohol Research UK (Research Innovation Grant Number: R2016/04). The funding body had no role in the design of the study and the writing of this article.

*Correspondence may be sent to Gillian W. Shorter at the Institute for Mental Health Sciences, School of Psychology, Ulster University, Cromore Road, Coleraine, BT52 1SA, United Kingdom, or via email at: gw.shorter@ulster.ac.uk.

be used for the evidence base (Glasziou, 2014). Given the number of reviews mentioning outcome heterogeneity across all populations in which we now use ABIs, it is no longer appropriate to dismiss this heterogeneity as a limitation when it can and should be addressed.

To address outcome heterogeneity in ABI trials, future ABI studies should use a coherent, consistent set of outcomes, known as a core outcome set (COS). A COS is a feature of a mature research base, and many health care areas have developed, or are developing, a COS to support advances in their field (COMET Initiative, 2017). A COS reduces selective and inconsistent reporting in research trials, improves the quality of treatment guidance for a condition, and increases the number of studies synthesized in systematic reviews. Both the Consolidated Standards of Reporting Trials (CONSORT; Moher et al., 2010) and the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) statements recommend COS use, and a formal process for COS development is established by the Core Outcome Measures in Effectiveness Trials (COMET) Initiative (Williamson et al., 2012, 2017). A COS is a minimum reporting standard and does not restrict the measurement of additional outcomes. A comprehensive map of outcomes can support decision making on other outcomes to be measured alongside the COS; reducing a potential source of conflict in trial planning (Daykin et al., 2016, 2017).

Recognizing the benefits an ABI COS could provide, the International Network on Brief Interventions for Alcohol and Other Drugs (INEBRIA) Research Measurement Standardization-Special Interest Group (IRMS-SIG) established the Outcome Reporting in Brief Intervention Trials: Alcohol (ORBITAL) project to derive a COS using COMET guidelines. This systematic review is a component of ABI COS development and follows the ORBITAL protocol (Shorter et al., 2017). Most ABI systematic reviews have aimed to establish efficacy, effectiveness, and/or cost-effectiveness, and their included studies meet a restrictive set of eligibility criteria, including a pre-specified outcome of interest. No study to date has compiled all outcomes used across ABI studies. This article fills this gap through a definitive catalogue of outcomes used in recent ABI trial literature. A catalogue is needed to (a) map outcomes used to show efficacy and effectiveness in peer-reviewed, published ABI trials; (b) show the variability in outcome type and measurement; (c) highlight methodological issues in the ABI field around outcomes and reporting; (d) inform COS development, including identifying outcomes for a Delphi prioritization exercise (see Shorter et al., 2019); and (e) support ABI trial protocol decision making on outcomes by trial area.

Method

The review protocol was registered in advance on PROSPERO (CRD42016047185) (Shorter et al., 2016b). Medline (OVID), EMBASE, PsycINFO (OVID), Health Management Information Consortium (HMIC), Cumulative Index to Nursing and Allied Health Literature (CINAHL), Allied and Complementary Medicine Database (AMED), Cochrane Library, ERIC (EBSCO), Web of Science, Google Scholar, Clinicaltrials.gov, and WHO International Clinical Trials Registry Platform (ICTRP) databases were searched to identify trials published between January 2000 and November 2017 in peer-reviewed journals. This date range provided a balance between capturing an extensive evidence base and reflecting the current state of ABI research. We focused on peer-reviewed publications as these are readily available in the public domain and reflect the work of the COS target audience (policy, practice, and research). Core search concepts related to three domains: alcohol use, brief interventions, and randomized trials. Terms were coupled with relevant Medical Subject Headings (MeSH)/thesaurus terms, truncated as appropriate, and variant spellings were used to identify useful records (Supplementary Material A contains the OVID search, which was adapted for other databases). (Supplemental material appears as an online-only addendum to the article on the journal's website.)

Eligible studies were individual or cluster randomized trials focused on efficacy or effectiveness of ABIs designed to reduce alcohol consumption published in peer-reviewed journals. Trials that did not analyze outcomes by randomized arm were excluded (e.g., subsample analysis only). Articles with the same trial registration number were included if they assessed different outcomes in each. Specific search parameters are described below.

Population

The population was current drinkers (at least one drink in the past year) who were aged 16 years or older. Trials of drinkers aged 15 years or below were excluded, as were trials including individuals seeking treatment for alcohol problems, following related UK NICE guidance (NICE, 2010).

Intervention

ABIs were defined as those suitable for drinkers not seeking treatment for an alcohol problem but who are identified by screening as having, or being at risk of, problems from their alcohol use (NICE, 2010). This definition covers brief advice and extended brief interventions, delivered once or more frequently. An ABI should assess an individual's alcohol use and provide feedback on their alcohol assessment. Trials including a multi-component intervention arm or where one or more intervention components addressed health behaviors that were not related to alcohol (e.g., smoking cessation) were included if alcohol intervention components and outcomes could be clearly distinguished.

Comparator

Comparators could be any active or control intervention.

Outcomes

All outcomes analyzed by randomized arm were extracted, including detail of how the outcome was defined and measured if possible. This was used to estimate the variability in outcome measurement, that is, to what extent an outcome in one article was exactly the same as in another article (what the outcome represents, how it was measured and scored, and time period referred to). Other extracted information included number and nature of sample randomized (sex, age, and population); trial details, including region, number of trial arms, trial arm composition, trial type (efficacy/effectiveness/not reported); and details of follow-up timing. These were summarized either as number (%) or mean (standard deviation) of trials included, with indication of missing data in the total number. Broad indicators of trial reporting quality were as follows: stating "trial" in the title or including a participant flow chart in line with early CONSORT guidance (Begg et al., 1996). Where study information was not provided this was stated. Effectiveness and efficacy ABI reviews often contact authors for missing data; we did not do so because our aim was to highlight where reporting falls short to improve standards in the field as do similar high-quality methodological reviews (Harman et al., 2017; Riddle et al., 2008; Thornley & Adams, 1998).

A map of outcomes was created by G.W.S. This was then refined by others (N.H., D.N.B., J.W.B., A.H.B., C.B., and E.L.G.). This outcome map listed outcomes under seven domains: alcohol-related outcomes, biomarkers, health measures, economic factors/social impacts, psychological/behavioral factors, life impact, and intervention factors. A range of sources influenced the outcome map. A presentation at the COMET V meeting in Amsterdam informed the first draft (attended by G.W.S. in September 2016), since published in Dodd et al. (2018). However, given that the ABI topic area is not directly concerned with physical pathology, many clinical factors in this taxonomy were irrelevant (e.g., musculoskeletal outcomes), whereas other outcomes were not specific enough (e.g., emotional functioning/well-being). Other sources included the Outcome Measurement Sets for Clinical Trials (OMERACT) filter (Boers et al., 2014); this was helpful to derive such core areas as death, life impact, resource use, or pathophysiological manifestations, but was too broad to capture outcomes relevant to ABIs. The Patient-Reported Outcomes Measurement Information System (PROMIS; Cella et al., 2007) provided elaboration to describe some outcomes in ABI trials (anxiety, depression, or sleep disturbance), but there were classification limitations, and some outcomes (e.g., PROMIS alcohol use questionnaire) were absent from ABI articles. We drew upon health economic reviews to inform the economic outcomes domain (Barbosa et al., 2010, 2015; Bray et al., 2011). Outcome data extracted from ABI trials were used to refine the taxonomy further.

Search results were downloaded to EndNote Version X7 (Clarivate, Philadelphia, PA) and de-duplicated. G.W.S. screened all titles and abstracts of articles and excluded those that did not meet the inclusion criteria. D.N.B. checked 28% of these for accuracy; discrepancies were resolved by discussion. All full-text versions of potentially eligible articles were reviewed by G.W.S., and all double-screened by one of E.L.G., D.N.B., J.W.B., A.J.O.D., A.H.B., and A.H.; discrepancies were resolved by discussion. Extraction forms were piloted by G.W.S. and K.J.S. All data were extracted by G.W.S., and all extracted data cross-checked for accuracy by at least one of E.L.G., D.N.B., S.J.S., K.J.S., J.W.B., A.J.O.D., and A.H.B. Data were presented from all trials and split by population (primary care, emergency department, university/college, general population [i.e., a general adult sample not selected as having specific characteristics], other health care, and other populations [including workplaces or job-related populations, $n = 14$; veteran populations, $n = 19$; community sample of persons with an intellectual disability, $n = 1$; homeless population, $n = 1$; criminal justice populations, $n = 14$; licensed premises, $n = 1$; sports clubs, $n = 1$; and young people ages 16 years and older, $n = 6$]). A PRISMA checklist is in Supplementary Material B.

Results

Searches identified 33,134 articles (after de-duplication) for eligibility screening by title and abstract. Exclusion at title and abstract stage reflected unambiguous violation of the above PICO (population, intervention, comparator, outcomes) based on topic area (i.e., not alcohol) or a known alcohol treatment sample (e.g., Project MATCH). Any unclear matches were referred to full-text assessment for closer inspection; 1,612 articles were retrieved for full-text evaluation against PICO criteria, and 401 were eligible (Figure 1). The 401 included articles covered 405 individual trials (some articles reported two trials), representing 182,272 randomized participants in total (see Supplementary Material C for included articles).

The mean trial size was 450 individuals (range: 12–7,935). Typically, higher numbers were randomized in primary care, emergency department, and general population samples compared with the remaining populations (Table 1). There were slightly more males than females on average (mean [*M*] % male = 56.2, *SD* = 28.1); highest in the "other" population. Most trials took place in North America (60.7%); this was particularly evident in the university/college population, with 81.1% of trials from this region. Two-arm trial designs predominated, with trials in the emergency department and university/college populations more likely to have over two

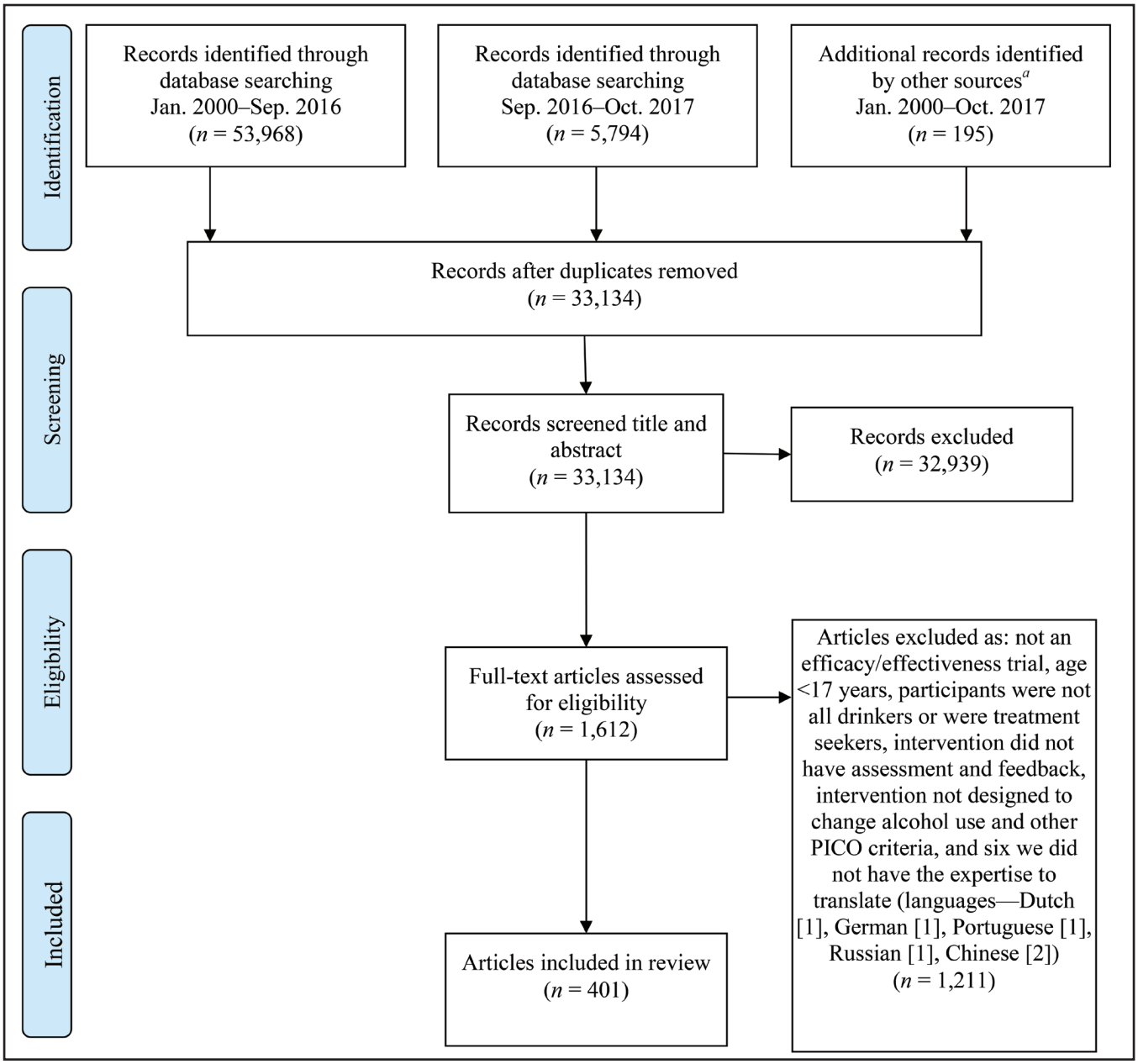

FIGURE 1. PRISMA Flow Diagram for the systematic review of outcomes in efficacy and effectiveness trials of alcohol brief interventions. PICO = population, intervention, comparator, outcomes. [a]Other sources were readily available peer-reviewed articles that were cited by or were citing those articles included in the review from the database screening that met the inclusion criteria.

arms. Around 83% of trials had a non-ABI control group. Other populations were most likely to have a non-ABI control (91.2%), and general population was least likely (75.0%). More trials were efficacy trials (52.8%) compared with effectiveness trials (42.0%). Twenty-one trials did not state their type. Only university/college populations had more effectiveness trials (56.1%) than efficacy trials (40.2%).

Just over half of the trials indicated they were a trial in the article's title (52.1%); 63.7% included a flow chart of participants through the trial. University/college populations were least likely to report these elements, with general population trials more likely to state that they were a trial (67.2%), and other health care populations more likely to include a flow chart (78.2%). Broadly similar percentages had two or three data collection waves (42% and 38%, respectively). Longer-term follow-up of 2 or more years was more likely in primary care (n = 7; 14%). Short-term follow-up was more likely in university/college samples. Overall, trialists most often selected 3-month intervals for follow-up (e.g., 3 or 6 months). Over time, there was a general increase in the

TABLE 1. Features of readily available peer-reviewed randomized trials of alcohol brief interventions (published January 1, 2000–October 31, 2017)

Variable	Overall (*n* = 405)	Primary care (*n* = 50)	Emergency department (*n* = 47)	University/ college (*n* = 132)	General population (*n* = 64)	Other health care (*n* = 55)	Other[a] (*n* = 57)
Trial size							
M (*SD*)	450.1 (730.1)	540.0 (967.6)	628.3 (692.9)	414.1 (662.3)	549.5 (1074.3)	277.3 (246.5)	362.4 (404.7)
Range of those randomized	12–7,935	29–6,897	45–4,476	18–5,227	29–7,935	40–975	12–1,449
Sex of those randomized							
Males, *M* % (*SD*)	56.2 (28.1)	65.4 (27.0)	67.6 (17.2)	45.9 (17.6)	56.1 (28.4)	46.4 (39.1)	76.1 (27.3)
Not reported/split by arm, *n*	65	9	17	15	6	8	10
Age of those randomized							
Age, in years, *M* (*SD*)	31.6 (13.3)	48.2 (14.8)	30.9 (6.5)	20.4 (2.3)	39.1 (10.4)	37.3 (10.5)	34.1 (14.1)
Not reported/split by arm/ grouped age, *n*	137	23	24	36	19	16	19
Trial region, *n* (%)							
North America	246 (60.7%)	23 (46.0%)	32 (68.1%)	107 (81.1%)	32 (50.0%)	22 (40.0%)	30 (52.6%)
Australia & New Zealand	20 (4.9%)	1 (2.0%)	1 (2.1%)	6 (4.5%)	4 (6.3%)	3 (5.5%)	5 (8.8%)
Europe	110 (27.2%)	18 (36.0%)	13 (27.7%)	16 (12.1%)	25 (39.1%)	19 (34.5%)	19 (33.3%)
South America	4 (1.0%)	0	1 (2.1%)	2 (1.5%)	0	0	1 (1.8%)
Africa	10 (2.5%)	4 (8.0%)	0	1 (0.8%)	0	5 (9.1%)	0
Asia	15 (3.7%)	4 (8.0%)	0	0	3 (4.7%)	6 (10.9%)	2 (3.5%)
No. of trial arms, *n* (%)							
With 2 arms	293 (72.3%)	39 (78.0%)	31 (66.0%)	85 (64.4%)	46 (71.9%)	45 (81.8%)	47 (82.5%)
With 3 arms	82 (20.2%)	9 (18.0%)	14 (29.8%)	28 (21.2%)	11 (17.2%)	10 (18.2%)	10 (17.5%)
With ≥4 arms	30 (7.5%)	2 (4.0%)	2 (4.3%)	19 (14.4%)	7 (10.9%)	0	0
Trial arm composition, *n* (%)							
With ≥1 arms not ABI	335 (82.7%)	39 (78.0%)	40 (85.1%)	108 (81.8%)	48 (75.0%)	48 (87.3%)	52 (91.2%)
With 1 ABI	249 (61.5%)	28 (56.0%)	31 (66.0%)	68 (51.5%)	36 (56.3%)	43 (78.2%)	43 (75.4%)
With ≥2 ABI	156 (38.5%)	22 (44.0%)	16 (34.0%)	64 (48.5%)	28 (43.7%)	12 (21.8%)	14 (24.6%)
Trial type, *n* (%)							
Efficacy	214 (52.8%)	34 (68.0%)	31 (66.0%)	53 (40.2%)	31 (48.4%)	33 (60.0%)	32 (56.1%)
Effectiveness	170 (42.0%)	16 (32.0%)	15 (31.9%)	74 (56.1%)	27 (42.2%)	17 (30.9%)	21 (36.8%)
Unclear	21 (5.2%)	0	1 (2.1%)	5 (3.8%)	6 (9.4%)	5 (9.1%)	4 (7.0%)
Reporting quality, *n* (%)							
Included "trial" in title	211 (52.1%)	22 (44.0)	23 (48.9%)	53 (40.2%)	43 (67.2%)	33 (60.0%)	37 (64.9%)
Included flowchart	258 (63.7%)	35 (70.0)	35 (74.5%)	63 (47.7%)	44 (68.8%)	43 (78.2%)	38 (66.7%)
Data collection waves, *n* (%)							
With 2 waves	168 (41.8%)	20 (40.8%)	16 (34.0%)	63 (48.1%)	27 (42.2%)	19 (34.5%)	23 (41.1%)
With 3 waves	153 (38.1%)	18 (36.7%)	24 (51.1%)	44 (33.6%)	24 (37.5%)	29 (52.7%)	14 (25.0%)
With ≥4 waves	81 (20.1%)	11 (22.4%)	7 (14.9%)	24 (18.3%)	13 (20.3%)	7 (12.8%)	19 (33.9%)
Follow up wave timing, *n* (%)							
0–2 weeks	10 (2.5%)	0	0	8 (6.1%)	0	1 (1.8%)	1 (1.8%)
>2 weeks–1 month	91 (22.5%)	5 (10.0%)	2 (4.3%)	54 (40.9%)	13 (20.1%)	8 (14.5%)	9 (15.8%)
>1 month–2 months	56 (13.8%)	3 (6.0%)	0	27 (20.5%)	9 (14.1%)	7 (12.7%)	10 (17.5%)
>2 month–3 months	168 (41.5%)	19 (38.0%)	25 (53.2%)	44 (33.3%)	30 (46.9%)	20 (36.4%)	30 (52.6%)
>3 month–4 months	14 (3.5%)	1 (2.0%)	1 (2.1%)	5 (3.8%)	1 (1.6%)	2 (3.6%)	4 (7.0%)
>4 month–6 months	202 (49.9%)	30 (60.0%)	23 (48.9%)	51 (38.6%)	36 (56.3%)	33 (60.0%)	29 (50.9%)
>6 month–9 months	32 (7.9%)	3 (6.0%)	3 (6.4%)	10 (7.6%)	3 (4.7%)	8 (14.5%)	5 (8.8%)
>9 month–12 months	120 (29.6%)	24 (48.0%)	30 (63.8%)	20 (15.2%)	14 (21.9%)	20 (36.4%)	12 (21.1%)
>12 month–18 months	14 (3.5%)	0	2 (4.3%)	2 (1.5%)	0	4 (7.3%)	6 (10.5%)
>18 month–24 months	17 (4.2%)	6 (12.0%)	0	3 (2.3%)	4 (6.3%)	3 (5.5%)	1 (1.8%)
>24 months	13 (1.7%)	7 (14.0%)	1 (2.1%)	3 (2.3%)	2 (3.1%)	0	0
No. and type of outcomes, *M* (*SD*)							
No. of primary outcomes (*n* = 285)	2.4 (2.0)	2.7 (3.0)	2.3 (1.3)	2.9 (2.4)	2.3 (1.3)	1.9 (1.3)	2.0 (1.6)
No. of secondary outcomes (*n* = 285)	4.1 (6.0)	5.3 (6.8)	5.1 (6.4)	2.6 (3.6)	4.3 (8.7)	4.5 (4.1)	3.9 (5.0)
No. not specified (*n* = 120)	6.4 (4.5)	11.2 (8.8)	5.9 (2.8)	6.0 (4.1)	4.8 (1.7)	5.6 (4.1)	8.1 (5.6)
No. of outcomes per trial (*n* = 405)	6.5 (5.8)	8.4 (7.3)	7.0 (5.8)	5.7 (3.9)	6.3 (8.0)	6.1 (4.1)	6.5 (5.0)
Trials with at least one outcome from each of the following domains, *n* (%)							
Alcohol-related outcomes	388 (95.8%)	46 (92.0%)	45 (95.7%)	132 (100%)	63 (98.4%)	52 (94.5%)	50 (87.7%)
Health	80 (19.8%)	16 (32.0%)	6 (12.8%)	10 (7.6%)	16 (25.0%)	19 (34.5%)	13 (22.8%)
Economic factors/social impacts	87 (21.5%)	20 (40.0%)	23 (48.9%)	13 (9.8%)	5 (7.8%)	8 (14.5%)	18 (31.6%)
Psychological/behavioral factors	114 (28.1%)	5 (10.0%)	11 (23.4%)	47 (35.6%)	14 (21.9%)	21 (38.2%)	16 (28.1%)
Life impact	57 (14.1%)	9 (18.0%)	8 (17.0%)	13 (9.8%)	11 (17.2%)	6 (10.9%)	10 (17.5%)
Biomarkers	13 (3.2%)	6 (12.0%)	0	0	0	2 (3.6%)	5 (8.8%)
Intervention factors	38 (9.4%)	1 (2.0%)	3 (6.4%)	14 (10.6%)	14 (21.9%)	2 (3.6%)	4 (7.0%)

Notes: No. = number. [a]Other includes workplaces or job related populations (*n* = 14), veteran populations (*n* = 19), community sample of persons with an intellectual disability (*n* = 1), homeless population (*n* = 1), criminal justice populations (*n* = 14), licensed premises (*n* = 1), sports clubs (*n* = 1), and young people (*n* = 6).

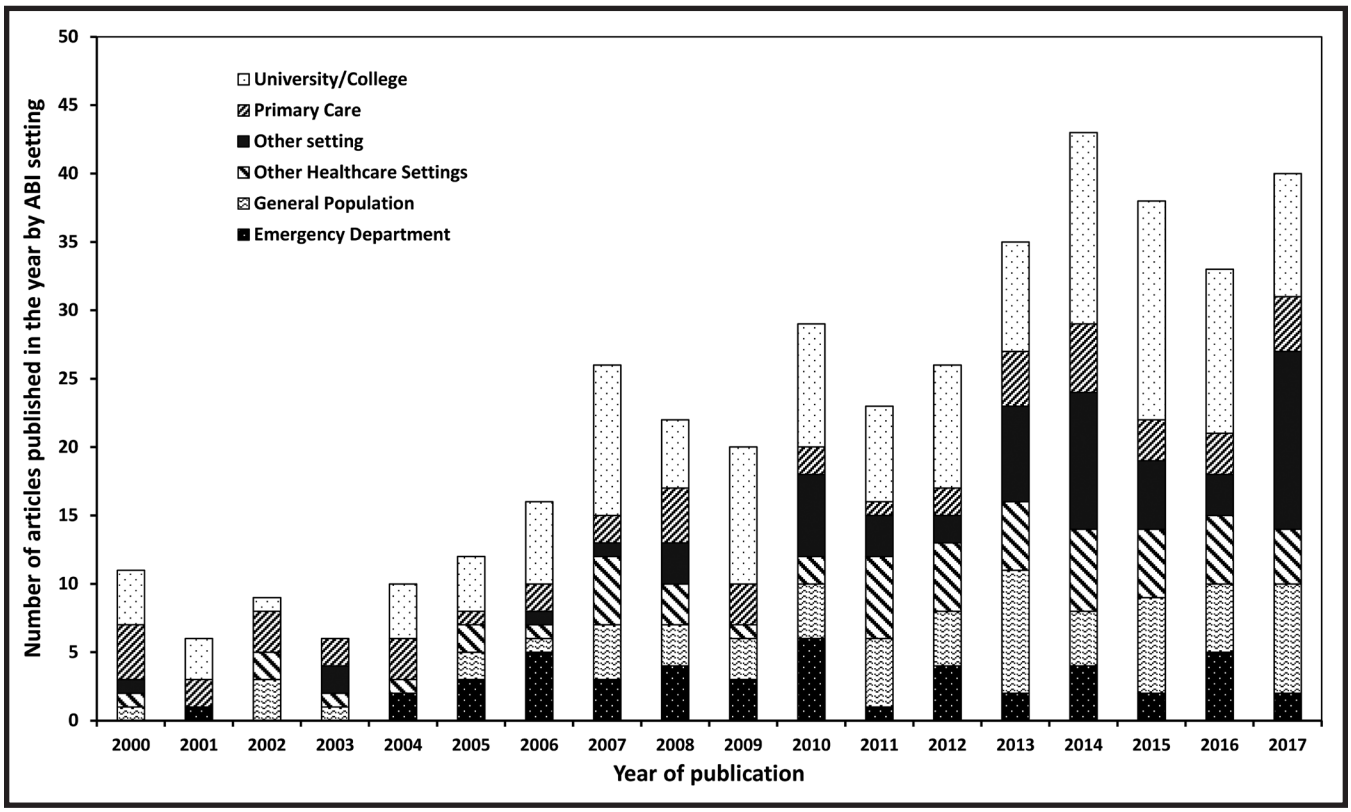

FIGURE 2. Number of alcohol brief intervention articles published per year, by population. ABI = alcohol brief intervention.

number of ABI trials published per year. The largest number were published in 2014. The number of trials per year is given in Figure 2.

Outcomes

Overall, 2,641 outcomes were extracted from 405 trials. Only 285 trials stated if their outcomes were primary or secondary. The mean number of outcomes per trial was 6.5 (ranging from 1 to 56); highest in primary care, and lowest in university/college samples. On average, there were two primary and four secondary outcomes reported in the included trials. Most trials had at least one alcohol-related outcome measure. The highest percentage of trials with at least one health outcome was in the primary care or other health care population, least likely in the university/college population. Economic factors or social impacts were most likely in the primary care or emergency department population. Psychological factors were found in around 28% of trials, most commonly in other health care populations. Life impact outcomes were present in 56 trials. Less than 10% of trials looked at intervention factors, and university/college samples were more likely to have one outcome of this type. Biomarkers were infrequently used: Only 13 trials had measured at least one biomarker, more likely in primary care or other populations.

Alcohol-related outcomes

Alcohol-related outcomes include those connected to the amount or pattern of alcohol consumption, those related to the comorbid use of other substances, and those reflecting substance use disorder symptomology. It is broader than just alcohol consumption measures, but we have kept the term *alcohol-related outcomes* for ease of exposition and to maintain consistency with our protocol (Shorter et al., 2017) and Delphi study (Shorter et al., 2019). In the 405 trials, there were 1,456 alcohol-related outcomes measured in 744 different ways (Table 2). The most commonly reported alcohol-related outcome variables were frequency of heavy drinking (n = 213), weekly drinks (n = 205), alcohol-related problems or consequences (n = 190), typical quantity (n = 137), typical frequency (n = 117), and hazardous or harmful drinking (n = 111). Many of the infrequently measured outcomes were also the most diversely measured. An exception included at-risk drinking (which measures risk derived from publicly available recommendations such as weekly or single episode limits). By population, primary care trials were most likely to report weekly drinks, frequency of heavy drinking, and at-risk drinking. This was somewhat similar to the general population, emergency department, and university/college populations, which often measured weekly drinks, frequency of heavy drinking, and alcohol-related problems or

Table 2. Frequency and variability in reporting of alcohol related outcome variables overall and by population

Outcome	*n* times measured	*n* ways measured	Ratio of variability[a]	Trial type[b] efficacy/ effectiveness/ not specified	Outcome type[c] primary/ secondary/ not specified	Primary care (*n* = 50)	Emergency department (*n* = 47)	University/ college (*n* = 132)	General population (*n* = 63)	Other health care (*n* = 55)	Other (*n* = 57)
Frequency of heavy drinking	213	128	0.6	105/103/5	90/65/58	30	32	72	28	24	27
No. of drinks consumed in a week	205	63	0.3	99/97/9	121/27/57	28	19	85	37	18	18
Alcohol-related problems or consequences	190	73	0.4	84/98/8	47/62/81	9	28	97	26	8	22
Typical quantity	137	63	0.5	81/49/7	53/38/46	17	16	45	22	19	18
Typical frequency	117	73	0.6	58/50/9	47/25/45	5	12	43	16	22	19
Hazardous or harmful drinking	111	27	0.2	83/20/8	34/49/28	21	15	14	15	25	21
Blood alcohol concentration	76	39	0.5	41/33/2	28/18/30	–	1	62	7	3	3
At-risk drinking	72	63	0.9	48/23/1	34/29/9	9	8	9	15	6	5
Largest number of drinks on occasion	57	36	0.6	22/31/4	23/17/17	–	10	30	9	5	3
Days abstinent	44	28	0.6	28/13/3	16/22/6	8	4	–	7	14	11
Combined consumption measure	42	17	0.4	21/9/2	20/9/13	5	3	10	12	7	5
Tobacco	29	17	0.6	5/24/0	7/17/5	2	–	13	9	3	2
Cannabis/marijuana use	26	19	0.7	7/18/1	6/13/7	2	2	16	2	2	2
No. of drinks in a month	22	12	0.5	15/7/0	13/3/6	5	1	12	3	1	–
Dependence symptomatology	19	10	0.5	14/4/1	3/10/6	3	5	3	3	3	2
Polydrug use (alcohol +)	17	12	0.7	9/7/1	6/8/3	2	1	2	4	6	2
Frequency of intoxication	15	9	0.6	9/5/1	1/4/10	2	3	6	–	2	2
Other substance use	13	11	0.8	10/3/0	2/7/4	5	–	2	–	2	4
Problems with other substances	13	13	1	4/8/1	4/4/5	–	3	7	1	2	–
Drinks on a specific occasion	9	8	0.9	5/4/0	2/1/6	–	1	7	–	–	1
No. of drinks in other period	8	6	0.8	8/0/0	6/0/2	1	2	1	–	3	1
No. of drinks consumed in 2 weeks	6	4	0.7	4/2/0	0/2/4	–	–	4	–	2	–
Other consumption measure[d]	5	5	1	1/3/1	0/4/1	–	–	1	2	1	1
Abuse symptomatology	4	4	1	2/2/0	1/1/2	2	–	–	–	1	1
Drinking game participation	4	2	0.5	0/4/0	0/0/4	–	–	4	–	–	–
Preloading alcohol	2	2	1	2/0/0	2/0/0	–	–	1	1	–	–
Totals	1,456	744	0.5			176	166	546	219	179	170

Notes: No. = number. [a]Ratio of variability is the approximate number of ways an outcome is measured divided by the number of outcomes; a higher number suggests greater variability. [b]Refers to the number of times an outcome appeared in an effectiveness or efficacy trial, or a trial not specified as either. [c]Refers to the number of times that an outcome appeared as first, as second, or not specified as either. [d]Includes the following measures (times measured) drinking non-beverage alcohol (1), average time spent drinking (1), substance use successfully verified by a significant other (1), drinking the number of drinks planned to consume that night/meeting personal drinking goal (1), whether the participant thought their drinking decreased, increased, or stayed the same (1).

consequences. Most trials that measured blood alcohol concentration were in the university/college population. Other health care populations often measured typical and heavy drinking frequencies, and hazardous and harmful drinking. Frequency of heavy drinking was the most commonly reported outcome in both efficacy and effectiveness trials, with the number of drinks consumed in a week the most frequent primary outcome.

Other outcomes

In total, 32 biomarker outcomes were reported across the 405 trials (Table 3). Of these, the most commonly reported was gamma-glutamyltransferase (GGT). Biomarkers were only found in primary care, other health care, and other populations. The most frequent biomarker in efficacy trials was GGT; in effectiveness trials it was carbohydrate-deficient transferrin (CDT). GGT and CDT tied as the most common primary outcome.

In the economic factors/social impacts domain, the most commonly reported outcomes were driving-related offenses and hospitalizations. This domain has some overlap with measures of alcohol-related consequences in the alcohol-related outcome domain, but measures here are intended to assess social costs and impacts, not to assess the possibility of a diagnosable alcohol disorder. In primary care, the most commonly reported economic factors/social impacts outcomes were driving-related offenses, hospitalizations, other criminal justice use, or other health care use. For ABI trials set in the emergency department, the most common were seeking alcohol treatment, driving-related offenses, and emergency health care use. In other health care populations, the most commonly assessed economic variable was that of provider intervention costs. In other populations, given the composition of this group, other criminal justice use was most common. General population or university/college ABI trials did not often measure economic factors/social impacts outcomes. The intervention cost to the provider was the most common economic factors/social impacts measure for efficacy trials. Driving-related offenses was the most common measure in effectiveness trials, and the most reported primary outcome.

Table 3. Frequency and variability in reporting of non-consumption variables overall and by population

Domain and outcome	*n* times measured	*n* ways measured	Ratio of variability[a]	Trial type[b] efficacy/ effectiveness/ not specified	Outcome type[c] primary/ secondary/ not specified	Primary care (*n* = 50)	Emergency department (*n* = 47)	University/ college (*n* = 132)	General population (*n* = 63)	Other healthcare (*n* = 55)	Other (*n* = 57)
Biomarkers											
Gamma-glutamyltransferase	10	7	0.7	9/1/0	3/4/3	5	–	–	–	–	1
Carbohydrate-deficient transferrin	7	4	0.6	4/3/0	3/4/0	4	–	–	–	1	2
Mean corpuscular volume	6	4	0.7	5/1/0	2/3/1	3	–	–	–	1	4
Alanine aminotransferase	4	3	0.8	3/1/0	2/2/0	3	–	–	–	–	1
Aspartate aminotransferase	4	3	0.8	3/1/0	2/0/2	3	–	–	–	1	2
Ethyl Glucuronide/ethyl sulfate	1	1	1	0/1/0	0/1/0	–	–	–	–	–	1
Totals	32	22	0.7	–	–	18	–	–	–	3	11
Economic factors/social impacts											
Driving-related offenses	60	36	0.6	17/41/2	17/36/7	22	29	3	–	–	6
Hospitalizations	36	11	0.3	26/9/1	5/29/2	17	8	–	2	4	5
Use of/seeking alcohol treatment	35	18	0.5	24/9/2	5/25/5	1	15	1	5	4	9
Other criminal justice use	35	26	0.7	17/18/0	10/19/6	17	6	1	–	1	10
Emergency healthcare use	30	8	0.3	18/11/1	7/21/2	12	10	–	1	2	5
General or other healthcare use	29	23	0.8	23/5/1	2/23/4	16	2	1	1	4	5
Intervention cost provider	28	5	0.2	28/0/0	0/28/0	1	–	–	1	21	5
General practitioner/primary care use	24	13	0.5	23/1/0	0/24/0	9	9	–	3	–	3
Alcohol-related injuries	18	10	0.6	11/6/1	2/6/10	2	9	–	–	1	6
Outpatient healthcare	14	8	0.6	12/2/0	0/12/2	6	3	–	1	–	4
Social care use	13	10	0.8	13/0/0	1/13/0	9	3	–	–	1	–
Over the counter/prescribed medication use	11	5	0.5	10/1/0	1/8/2	8	–	–	2	1	–
Alcohol-related offences	11	8	0.7	1/10/0	6/3/2	8	–	3	–	–	–
Use of self-help for alcohol	9	8	0.9	6/2/1	0/7/2	1	1	2	3	1	1
Quality-adjusted life years	9	1	0.1	9/0/0	0/9/0	6	–	–	–	1	2
Other service use	7	7	1	7/0/0	0/7/0	1	4	–	–	–	2
Intervention cost client	6	5	0.8	6/0/0	0/6/0	2	–	–	–	–	4
Injuries (general)	4	3	0.8	2/2/0	1/2/1	2	2	–	–	–	–
Incremental cost-effectiveness ratio	3	1	0.3	3/0/0	0/3/0	2	1	–	–	–	–
Intervention cost overall/not specified	3	2	0.7	3/0/0	0/3/0	3	–	–	–	–	–
Other health economic measures	3	3	1	0/2/1	0/1/2	–	–	3	–	–	–
Productivity losses	2	1	0.5	2/0/0	0/2/0	2	–	–	–	–	–
Societal perspectives	2	1	0.5	2/0/0	0/2/0	2	–	–	–	–	–
Substance-free reinforcement	2	1	0.5	0/2/0	0/2/0	–	–	2	–	–	–
Totals	401	220	0.5	–	–	149	102	23	19	41	67
Health											
Alcohol-exposed pregnancy factors	31	16	0.5	12/19/0	14/13/4	–	–	4	7	18	2
Psychological health	26	15	0.6	20/4/2	0/23/3	7	3	1	2	6	7
Sexual violence or coercion	25	21	0.8	17/8/0	4/16/5	2	4	5	–	6	8
Severity of depression symptoms	24	13	0.5	13/10/1	6/17/1	4	–	4	7	4	5
Physical health	13	6	0.5	13/0/0	0/11/2	7	2	–	1	1	2
General health	10	9	0.9	9/1/0	1/5/4	3	–	1	2	3	1
Cardiac outcomes	8	2	0.3	8/0/0	0/8/0	6	–	–	–	–	2

Table continued

Health outcomes most commonly reported were alcohol-exposed pregnancy factors, psychological health measures, sexual violence or coercion, and severity of depression symptoms. In primary care, cardiac factors, psychological health, and physical health were most commonly reported. In general population samples, alcohol-exposed pregnancy factors or severity of depression were more commonly reported. Sleep disruption was only measured in university/college ABI trials. Other health care populations most commonly reported alcohol-exposed pregnancy factors. The most frequent efficacy outcome in this domain was psychological health; the most common outcome in effectiveness trials was alcohol-exposed pregnancy factors. The most commonly reported outcome from the intervention factors domain was intervention satisfaction; true for both effectiveness and efficacy trials. ABI trials in university/college and general population samples were more likely to ask participants about this outcome.

In the domain of psychological and behavioral factors, the most commonly reported outcomes across all trials were drinking refusal self-efficacy, alcohol outcome expectancies, risky behaviors, and readiness to change. ABI trials in the primary care and emergency department populations were least likely to measure these outcomes. By contrast, university/college samples were particularly likely to measure the perception of others' drinking, for example, the typical quan-

TABLE 3. *Continued*

Domain and outcome	*n* times measured	*n* ways measured	Ratio of variability[a]	Trial type[b] efficacy/ effectiveness/ not specified	Outcome type[c] primary/ secondary/ not specified	Primary care (*n* = 50)	Emergency department (*n* = 47)	University/ college (*n* = 132)	General population (*n* = 63)	Other healthcare (*n* = 55)	Other (*n* = 57)
Other health factors	7	7	1	7/0/0	0/5/2	2	2	–	–	3	–
Severity of anxiety symptoms	6	6	1	2/4/0	1/5/0	–	–	1	2	2	1
Severity of PTSD symptoms	6	4	0.7	3/3/0	1/5/0	–	–	1	–	–	5
Weight/obesity	6	4	0.7	6/0/0	0/6/0	2	–	–	4	–	–
Sleep disturbance	6	6	1	6/0/0	0/6/0	–	–	6	–	–	–
Mortality/death	5	5	1	0/5/0	0/0/5	–	–	–	5	–	–
Suicidality	3	2	0.7	2/0/1	0/3/0	2	1	–	–	–	–
Totals	176	116	0.7	–	–	35	12	23	30	43	33
Intervention factors											
Intervention satisfaction	73	54	0.7	29/27/17	1/47/25	3	2	31	26	6	5
Intervention delivered/used as expected	15	13	0.9	12/3/0	1/9/5	–	4	4	6	1	–
Perceived change in alcohol use	9	9	1.0	3/6/0	0/6/3	–	–	3	5	–	1
Other intervention factors	2	2	1.0	1/1/0	0/2/0	–	–	–	2	–	–
Totals	99	78	0.8	–	–	3	6	38	39	7	6
Psychological/behavioral factors											
Readiness to change	80	50	0.6	60/17/3	4/49/27	4	14	18	17	12	15
Risky behaviors	49	43	0.9	36/13/0	18/22/9	9	3	6	–	27	4
Drinking refusal self-efficacy	43	31	0.7	32/8/3	1/25/17	1	2	8	17	2	13
Alcohol outcome expectancies	42	42	1	34/8/0	2/28/12	–	2	11	15	2	12
Perception of others' drinking	40	39	1	4/36/0	1/10/29	–	–	37	1	–	2
Protective behavioral strategy use	17	15	0.9	2/15/0	9/4/4	–	–	17	–	–	–
Anger or aggression	14	14	1	13/1/0	1/1/12	1	–	–	–	–	13
Other psychological factors	14	13	0.9	9/5/0	0/6/8	1	–	–	–	–	13
Sexual factors	11	8	0.7	10/1/0	2/6/3	3	–	–	–	7	1
Knowledge of alcohol	10	10	1	9/1/0	1/9/0	–	–	–	8	1	1
Negative/positive views of alcohol	9	8	0.9	9/0/0	0/6/3	–	1	2	6	–	–
Alcohol demand curve measures	7	6	0.9	0/7/0	0/3/4	–	–	7	–	–	–
Others' concern about drinking	7	4	0.6	5/1/1	0/4/3	1	–	2	1	1	2
Drinking to cope	6	6	1	4/1/1	1/1/4	–	–	4	–	1	1
Alcohol-induced memory loss	5	4	0.8	4/0/1	0/3/2	–	–	1	1	1	2
Readiness to receive help	5	5	1	4/1/0	1/2/2	–	1	1	1	2	–
Guilt after drinking	4	3	0.8	3/0/1	0/2/2	–	–	1	1	2	–
Drinking in the morning	3	2	0.7	2/0/1	0/2/1	–	–	–	–	1	2
Impulsivity	2	2	1	1/1/0	0/1/1	–	–	1	–	–	1
Goals and goal striving	2	2	1	1/0/1	0/1/1	–	–	1	1	–	–
Totals	363	301	0.8	–	–	20	23	110	68	58	84
Life impact											
Role-functioning/relationship factors	66	44	0.7	26/26/14	5/44/17	4	12	16	5	14	15
Quality of life	48	27	0.6	44/1/3	1/44/3	9	5	–	22	3	9
Totals	114	71	0.6	–	–	13	17	16	27	17	24

Notes: PTSD = posttraumatic stress disorder. [a]Ratio of variability is the approximate number of ways an outcome is measured divided by the number of outcomes; a higher number suggests greater variability. [b]Refers to the number of times an outcome appeared in an effectiveness or efficacy trial, or a trial not specified as either. [c]Refers to the number of times that an outcome appeared as first, as second, or not specified as either.

tity drunk by students at their institution. For general population samples, drinking refusal self-efficacy and readiness to change were most common. In other health care populations, risky behaviors were the most commonly reported; these include aspects such as sex without effective contraception. Last, in other populations, anger and aggression, drinking refusal self-efficacy, other psychological factors, and readiness to change were the most commonly reported outcomes. The most frequent measure in efficacy trials was readiness to change; for effectiveness trials, it was the perception of others' drinking. The most common primary outcome for both was engagement in risky behaviors. Life impact measures were most commonly role-functioning or relationship factors or quality of life. The former was most common in effectiveness trials (and as a primary outcome); the latter the most common for efficacy trials.

Discussion

This review is the first to go beyond stating outcome heterogeneity as a weakness in ABI systematic reviews; it quantifies the heterogeneity and inconsistency in outcomes reported in effectiveness and efficacy trials of ABIs. Overall, there were 2,641 outcomes measured in approximately 1,560 different ways, truly a "Tower of Babel." The estimated 1,560 different ways authors measured outcomes may be a

conservative guess of the true variability, given the lack of precision on how outcomes were measured. The variation in the outcomes used and reported across ABI trials reflects similar reviews conducted in different research areas (Harman et al., 2017). For the ABI field, the substantial heterogeneity represents an important challenge. Meta-analyses will continue to be compromised, as they cannot draw on all evidence to decide whether ABIs work as intended. Just over half (53%) measured the most common consumption measure frequency of heavy drinking; this creates a considerable conflict between the drive to include all studies meeting criteria in high quality systematic reviews and the ability to include all studies in the meta-analysis.

Determining efficacy or effectiveness depends on outcomes measured, and therefore all ABI trial papers should contain sufficient detail on outcome measurement. One way this may affect meta-analyses is through the combination of an outcome (e.g., weekly drinks), which hides considerable variability. For example, "weekly drinks" may refer to an average week, a typical week, or the last week. It may refer to a typical week in the past month, 28 days, 90 days, 6 months, or since last measurement. The definition of *drink* may be specified or left to the respondent. Weekly drinks may be reported directly or calculated based on other information in a range of different questionnaires. We can calculate some differences to be equivalent, but some measure genuine differences and their combination compromises the validity of estimates. At a minimum, trials should report (a) what the outcome is, (b) the question or questionnaire used to measure and how this is used (e.g., scale score, or the binary above and below a cut-off point), (c) measure of aggregation (e.g., mean value or mean individual difference), and (d) time point (e.g., 1, 4, and 8 weeks postintervention).

Some trials did not specify whether their outcomes were primary or secondary outcomes. This could be because the trial was a pilot study and specification may not be required (Eldridge et al., 2016), or it might be stated in a trial registry. However, excluding this from reporting is problematic (Begg et al., 1996; Moher et al., 2010). In addition, although one might expect trials to have only one primary outcome, we found, of those who specified, the average was two primary outcomes. This was an underestimate of the total average because some articles only reported secondary outcomes; their primary outcome(s) were in other articles with the same trial registration number. The correct interpretation of secondary outcomes is "through" the primary analysis on the premise that, if the primary outcome is positive, then secondary outcomes can help to understand how the ABI worked. The secondary designation may also be useful for outcomes more distal on the causal pathway that reduced drinking would be expected to change. If the primary outcome is neutral, the secondary outcomes are hypothesis generating. If the primary outcome is negative, the secondary outcomes provide insight into how the treatment caused harm (Freemantle, 2001). If change is shown in some primary outcomes but not others, interpretation can become difficult and it may be a challenge to state that the ABI brought about change. To improve the aggregation of trials into the evidence base, outcomes (from a COS or otherwise) should be detailed, identified as primary or secondary with a clear statistical analysis plan, well reported in results sections that include point estimates and variability around estimates, and follow reporting guidance.

Alcohol-related outcomes, particularly consumption outcomes, were the predominant outcomes measured in ABI trials. Although some have called for an increase in biomarkers in ABI trials (Kypri, 2007), this call has not been heeded; most outcomes were self-reported. ABI effectiveness or efficacy meta-analyses rely on the outcomes reported without validating them against objective measures (Moyer et al., 2002), exacerbating the problem of outcome heterogeneity in ABI trials. Our review provides the first systematic and quantifiable evidence to support previous calls for standard definitions of ABI outcomes to compare across studies (Bernstein et al., 2010).

Despite efforts to identify literature from across the globe, most trials were from North American or European countries. This may reflect the predominance of publishing or funding opportunities available to those researchers, be evidence of the high levels of hazardous and harmful use of alcohol in these countries (Rehm et al., 2009), or be a consequence of the pre-specified databases searched. We attempted to minimize English language bias and improve the quality of the review by including studies reported in languages other than English (Moher et al., 2003). The searching was largely conducted in English, and our ability to extract data from articles in languages other than English was limited, as shown in the CONSORT flowchart (Figure 1). Although focusing on peer-reviewed literature may have also limited the number of non-English articles included, it is in keeping with our intention to focus on those articles that are likely to be most accessible and influential for many decision makers. Our searches of the gray literature, which constitute a separate part of our PROSPERO-registered systematic review not reported here, will be one opportunity to explore how improving access to a wider range of literature from low-resource settings, or from reports in languages other than English, may influence the evidence base. This limitation is likely to have shown additional heterogeneity in findings, as the number of valid trials increased.

There was also a predominance of efficacy trials in the included studies, and attention should turn toward effectiveness trials within the different populations. Efficacious interventions may not be effective in routine practice (McCambridge & Saitz, 2017). Some trials did not specify their trial approach as either efficacy or effectiveness, although this may be a consequence of challenges of specification across the efficacy-to-effectiveness continuum (Heather,

2014). Short-term follow-up was common, as reported by other systematic reviews (Moyer et al., 2002). This is perhaps expected given that effect sizes tend to be larger at early follow-up, and there are concerns about the longitudinal effects of ABI (Donoghue et al., 2014). The predominant follow-up interval was around 3 months between data collection points. With about 20% of studies having four or more follow-up points, there is a balance between minimizing loss to follow-up, timely collection of only important information, and respondent burden (Lin et al., 2012).

By synthesizing outcome selection, this review offers the opportunity to consider outcome choice and the implications for the ABI field; other health care areas have noted the importance of design in attrition (Kilburn et al., 2014). Some have considered respondent burden (Cunningham et al., 1999; Kypri, 2007); as the number of outcomes reported was 56 in one trial, this may need careful consideration. Decision making around which outcomes to use for particular trials can be assisted by this outcome map, broken down by research area, effectiveness/efficacy, and primary/secondary/other outcomes. The structure of this outcome map was the process of discussion between co-authors, and we recognize that other structures of categorization may also exist.

This review highlights the importance of a COS for efficacy or effectiveness ABI randomized trials. This review will contribute to the efforts to establish a COS using high-quality, established methodologies (Williamson et al., 2017) that will improve and standardize reporting in the ABI field. This review also informs a preliminary list of outcomes for a related e-Delphi prioritization exercise (Shorter et al., 2019) and for discussion at the consensus meeting as outlined in the ORBITAL protocol (Shorter et al., 2017). We aimed to better understand how the extent of variability and reporting of outcomes compromises the evidence base and have conclusively shown that this variability is considerable, and reporting is incomplete. The ability of users of ABI research to compare and understand findings is restricted because we do not know what exactly was measured and how, nor can we confidently compare seemingly alike outcomes. We did not seek to improve the completeness of the data by contacting the original authors, but used the incompleteness (contrary to usual systematic review practice) as a tool to highlight shortcomings in the field. We must improve issues of reporting and methodological quality to advance the field; the ABI evidence base cannot move from middle age to more established without it (Babor et al., 2007).

Acknowledgments

The authors thank the INEBRIA Measurement Standardization SIG and participants at the INEBRIA 2016 conference workshop on the topic for their useful commentary/feedback.

Registration: PROSPERO review registration (CRD42016047185) on September 20, 2016, before review start (Shorter et al., 2016b). This work is part of and informs the Outcome Reporting in Brief Intervention Trials: Alcohol project (ORBITAL), which aims to create a core outcome set or minimum data standard for alcohol brief intervention trials. This ORBITAL project arose from collaborations in the International Network on Brief Interventions for Alcohol and Other Drugs (INEBRIA) Research Measurement Standardization Special Interest Group. This was registered at the start of the project at the COMET Initiative (Shorter et al., 2016a) and the protocol has been published (Shorter et al., 2017).

References

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,* 7–30. doi:10.1300/J465v28n03_03

Ballesteros, J., Ariño, J., González-Pinto, A., & Querejetad, I. (2003). Eficacia del consejo médico para la reducción del consumo excesivo de alcohol. Metaanálisis de estudios españoles en atención primaria [Effectiveness of medical advice for reducing excessive alcohol consumption. Meta-analysis of Spanish studies in primary care]. *Gaceta Sanitaria, 17,* 116–122. doi:10.1016/S0213-9111(03)71708-7

Ballesteros, J., Duffy, J. C., Querejeta, I., Ariño, J., & González-Pinto, A. (2004a). Efficacy of brief interventions for hazardous drinkers in primary care: Systematic review and meta-analyses. *Alcoholism: Clinical and Experimental Research, 28,* 608–618. doi:10.1097/01.ALC.0000122106.84718.67

Ballesteros, J., González-Pinto, A., Querejeta, I., & Ariño, J. (2004b). Brief interventions for hazardous drinkers delivered in primary care are equally effective in men and women. *Addiction, 99,* 103–108. doi:10.1111/j.1360-0443.2004.00499.x

Barbosa, C., Cowell, A., Bray, J., & Aldridge, A. (2015). The cost-effectiveness of alcohol screening, brief intervention, and referral to treatment (SBIRT) in emergency and outpatient medical settings. *Journal of Substance Abuse Treatment, 53,* 1–8. doi:10.1016/j.jsat.2015.01.003

Barbosa, C., Godfrey, C., & Parrott, S. (2010). Methodological assessment of economic evaluations of alcohol treatment: What is missing? *Alcohol and Alcoholism, 45,* 53–63. doi:10.1093/alcalc/agp067

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., . . . Stroup, D. F. (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA, 276,* 637–639. doi:10.1001/jama.1996.03540080059030

Beich, A., Thorsen, T., & Rollnick, S. (2003). Screening in brief intervention trials targeting excessive drinkers in general practice: Systematic review and meta-analysis. *BMJ, 327,* 536–542. doi:10.1136/bmj.327.7414.536

Bernstein, E., Bernstein, J. A., Stein, J. B., & Saitz, R. (2009). SBIRT in emergency care settings: Are we ready to take it to scale? *Academic Emergency Medicine, 16,* 1072–1077. doi:10.1111/j.1553-2712.2009.00549.x

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,* 498–507. doi:10.1111/j.1465-3362.2010.00174.x

Bertholet, N., Daeppen, J.-B., Wietlisbach, V., Fleming, M., & Burnand, B. (2005). Reduction of alcohol consumption by brief alcohol intervention in primary care: Systematic review and meta-analysis. *Archives of Internal Medicine, 165,* 986–995. doi:10.1001/archinte.165.9.986

Boers, M., Kirwan, J. R., Wells, G., Beaton, D., Gossec, L., d'Agostino, M. A., . . . Tugwell, P. (2014). Developing core outcome measurement sets for clinical trials: OMERACT Filter 2.0. *Journal of Clinical Epidemiology, 67,* 745–753. doi:10.1016/j.jclinepi.2013.11.013

Bray, J. W., Cowell, A. J., & Hinde, J. M. (2011). A systematic review and meta-analysis of health care utilization outcomes in alcohol screening and brief intervention trials. *Medical Care, 49,* 287–294. doi:10.1097/MLR.0b013e318203624f

Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., . . . Rose, M., & the PROMIS Cooperative Group. (2007). The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap cooperative group during its first two years. *Medical Care, 45, Supplement 1,* S3–S11. doi:10.1097/01.mlr.0000258615.42478.55

Coffield, A. B., Maciosek, M. V., McGinnis, J. M., Harris, J. R., Caldwell, M. B., Teutsch, S. M., . . . Haddix, A. (2001). Priorities among recommended clinical preventive services. *American Journal of Preventive Medicine, 21,* 1–9. doi:10.1016/S0749-3797(01)00308-7

COMET Initiative. (2017). *How to search the COMET Initiative database.* Retrieved from https://stream.liv.ac.uk/eqyg4t36

Cunningham, J. A., Ansara, D., Wild, T. C., Toneatto, T., & Koski-Jännes, A. (1999). What is the price of perfection? The hidden costs of using detailed assessment instruments to measure alcohol consumption. *Journal of Studies on Alcohol, 60,* 756–758. doi:10.15288/jsa.1999.60.756

Daykin, A., Selman, L. E., Cramer, H., McCann, S., Shorter, G. W., Sydes, M. R., . . . Shaw, A. (2016). What are the roles and valued attributes of a Trial Steering Committee? Ethnographic study of eight clinical trials facing challenges. *Trials, 17,* 307. doi:10.1186/s13063-016-1425-y

Daykin, A., Selman, L. E., Cramer, H., McCann, S., Shorter, G. W., Sydes, M. R., . . . Shaw, A. (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. *Trials, 18,* 612. doi:10.1186/s13063-017-2305-9

Dodd, S., Clarke, M., Becker, L., Mavergames, C., Fish, R., & Williamson, P. R. (2018). A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. *Journal of Clinical Epidemiology, 96,* 84–92. doi:10.1016/j.jclinepi.2017.12.020

Donoghue, K., Patton, R., Phillips, T., Deluca, P., & Drummond, C. (2014). The effectiveness of electronic screening and brief intervention for reducing levels of alcohol consumption: A systematic review and meta-analysis. *Journal of Medical Internet Research, 16,* e142. doi:10.2196/jmir.3193

Eldridge, S. M., Chan, C. L., Campbell, M. J., Bond, C. M., Hopewell, S., Thabane, L., & Lancaster, G. A., & the PAFS consensus group. (2016). CONSORT 2010 statement: Extension to randomised pilot and feasibility trials. *Pilot and Feasibility Studies, 2,* 64. doi:10.1186/s40814-016-0105-8

Field, C. A., Baird, J., Saitz, R., Caetano, R., & Monti, P. M. (2010). The mixed evidence for brief intervention in emergency departments, trauma care centers, and inpatient hospital settings: What should we do? *Alcoholism: Clinical and Experimental Research, 34,* 2004–2010. doi:10.1111/j.1530-0277.2010.01297.x

Freemantle, N. (2001). Interpreting the results of secondary end points and subgroup analyses in clinical trials: Should we lock the crazy aunt in the attic? *BMJ, 322,* 989–991. doi:10.1136/bmj.322.7292.989

Glasziou, P., Altman, D. G., Bossuyt, P., Boutron, I., Clarke, M., Julious, S., . . . Wager, E. (2014). Reducing waste from incomplete or unusable reports of biomedical research. *The Lancet, 383,* 267–276. doi:10.1016/S0140-6736(13)62228-X

Harman, N. L., James, R., Wilding, J., & Williamson, P. R., & the SCORE-IT study team. (2017). SCORE-IT (Selecting Core Outcomes for Randomised Effectiveness trials In Type 2 diabetes): A systematic review of registered trials. *Trials, 18,* 597. doi:10.1186/s13063-017-2317-5

Heather, N. (2014). The efficacy-effectiveness distinction in trials of alcohol brief intervention. *Addiction Science & Clinical Practice, 9,* 13. doi:10.1186/1940-0640-9-13

Heather, N. (2016). Spreading alcohol brief interventions from health care to non-health care settings: Is it justified? *Drugs: Education, Prevention, & Policy, 23,* 359–364. doi:10.1080/09687637.2016.1187113

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, Issue 2,* Article No. CD004148. doi:10.1002/14651858.CD004148.pub4

Kilburn, L. S., Banerji, J., & Bliss, J. M., & the NCRI Breast Clinical Studies Group. (2014). The challenges of long-term follow-up data collection in non-commercial, academically-led breast cancer clinical trials: The UK perspective. *Trials, 15,* 379. doi:10.1186/1745-6215-15-379

Kypri, K. (2007). Methodological issues in alcohol screening and brief intervention research. *Substance Abuse, 28,* 31–42. doi:10.1300/J465v28n03_04

Lin, J. Y., Lu, Y., & Tu, X. (2012). How to avoid missing data and the problems they pose: Design considerations. *Shanghai Archives of Psychiatry, 24,* 181–184. doi:10.3969/j.issn.1002-0829.2012.03.010

McCambridge, J., & Saitz, R. (2017). Rethinking brief interventions for alcohol in general practice. *BMJ, 356,* j116. doi:10.1136/bmj.j116

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., . . . Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ, 340,* c869. doi:10.1136/bmj.c869

Moher, D., Pham, B., Lawson, M. L., & Klassen, T. P. (2003). The inclusion of reports of randomised trials published in languages other than English in systematic reviews. *Health Technology Assessment, 7,* 1–90. doi:10.3310/hta7410

Moyer, A., Finney, J. W., Swearingen, C. E., & Vergun, P. (2002). Brief interventions for alcohol problems: A meta-analytic review of controlled investigations in treatment-seeking and non-treatment-seeking populations. *Addiction, 97,* 279–292. doi:10.1046/j.1360-0443.2002.00018.x

National Institute for Health and Care Excellence. (2010). *NICE Public Health (PH) Guideline 24: Alcohol-use disorders: Prevention.* Retrieved from http://www.nice.org.uk/PH24

O'Donnell, A., Anderson, P., Newbury-Birch, D., Schulte, B., Schmidt, C., Reimer, J., & Kaner, E. (2014). The impact of brief alcohol interventions in primary healthcare: A systematic review of reviews. *Alcohol and Alcoholism, 49,* 66–78. doi:10.1093/alcalc/agt170

Rehm, J., Mathers, C., Popova, S., Thavorncharoensap, M., Teerawattananon, Y., & Patra, J. (2009). Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. *The Lancet, 373,* 2223–2233. doi:10.1016/S0140-6736(09)60746-7

Riddle, D. L., Stratford, P. W., & Bowman, D. H. (2008). Findings of extensive variation in the types of outcome measures used in hip and knee replacement clinical trials: A systematic review. *Arthritis Care and Research, 59,* 876–883. doi:10.1002/art.23706

Saitz, R. (2010). Candidate performance measures for screening for, assessing, and treating unhealthy substance use in hospitals: Advocacy or evidence-based practice? *Annals of Internal Medicine, 153,* 40–43. doi:10.7326/0003-4819-153-1-201007060-00008

Saitz, R., Svikis, D., D'Onofrio, G., Kraemer, K. L., & Perl, H. (2006). Challenges applying alcohol brief intervention in diverse practice settings: Populations, outcomes, and costs. *Alcoholism: Clinical and Experimental Research, 30,* 332–338. doi:10.1111/j.1530-0277.2006.00038.x

Shorter, G. W., Heather, N., Bray, J. W., Berman, A. H., Giles, E. L., O'Donnell, A. J., . . . Newbury-Birch, D. (2019). Prioritization of outcomes in efficacy and effectiveness alcohol brief intervention trials: International multi-stakeholder e-Delphi consensus study to inform a core outcome set. *Journal of Studies on Alcohol and Drugs, 80,* 299–309. doi:10.15288/jsad.2019.80.299

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,* 611. doi:10.1186/s13063-017-2335-3

Shorter, G. W., & Heather, N. Newbury- Birch, D., Giles, E. L., Holloway, A., Stockdale, K. J., . . . O'Donnell, A. J. (2016a). *Outcome Reporting*

for Brief Intervention Trials (ORBIT): COMET Initiative Protocol. Retrieved from http://cometinitiative.org/studies/details/957

Shorter, G. W. Newbury- Birch, D., Heather, N., Giles, E. L., Holloway, A., Bray, J. W., . . . O'Donnell, A. J. (2016b). *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

Thornley, B., & Adams, C. (1998). Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ, 317,* 1181–1184. doi:10.1136/bmj.317.7167.1181

U.S. Preventive Services Task Force. (2004). *Screening and behavioral counseling interventions in primary care to reduce alcohol misuse: Recommendation statement.* Retrieved from https://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/alcohol-misuse-screening-and-behavioral-counseling-interventions-in-primary-care

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, Supplement 3,* 280. doi:10.1186/s13063-017-1978-4

Williamson, P. R., Altman, D. G., Blazeby, J. M., Clarke, M., Devane, D., Gargon, E., & Tugwell, P. (2012). Developing core outcome sets for clinical trials: Issues to consider. *Trials, 13,* 132. doi:10.1186/1745-6215-13-132

World Health Organization. (2016). *Management of substance abuse.* Retrieved from http://www.who.int/substance_abuse/activities/sbi/en