A Systematic Review and Meta-Analysis of the Effectiveness of Brief Interventions for Reducing Antenatal Alcohol Consumption

Joel D. Smith

School of Psychology, University of Adelaide

Author Note

This thesis is submitted in partial fulfilment of the Honours degree of Bachelor of Psychological Science (Honours).

There are no known conflicts of interest to report.

Abstract word count = 250 words. Main text word count = 9471.

Contents

List of Figures

List of Tables

Abstract

Antenatal alcohol consumption increases risk of adverse pregnancy outcomes. It is recommended that pregnant women abstain from alcohol, though 14.5% of pregnant Australians are reported not to. Screening and brief interventions, which involve single, short therapeutic sessions to motivate behaviour change, are recommended for treatment of antenatal alcohol consumption. However, limited evidentiary support for the practice, reports of poor clinical applicability, and the lack of existing meta-analyses of patterns of use data raise concerns about whether these recommendations are justified. This meta-analysis investigates the effectiveness of ≤60-minute single-session brief interventions on reducing alcohol consumption frequency, quantity, and abstinence outcomes in pregnant women screened positive for alcohol use. Seven databases were searched to yield 15716 records. Nine studies were included for review, and eight for analysis. Exclusions were made for polydrug and multi-session screening or interventions. Frequency and quantity outcomes were assessed using Hedges' *g* values and abstinence outcomes using odds ratios. Subgroup analyses and meta-regressions were conducted on potential predictors of effectiveness. Random-effects models were employed. Significant effects in favour of intervention were observed only for meta-analyses of abstinence outcomes. However, no results were deemed clinically significant due to the limited number of studies viable for analysis and their notable risks of bias. Imprecision and high risk of publication biases were also identified. Existing healthcare recommendations were therefore not validated and a need for further research with more consistent methodologies was identified. How research consistently, quality, and ethicality could be improved in future studies is discussed, with a framework provided.

Keywords: Screening, brief intervention, pregnancy, alcohol, antenatal alcohol consumption, systematic review, meta-analysis

Declaration

This thesis contains no material which has been accepted for the award of any other degree of diploma in any University, and, to the best of my knowledge, this thesis contains no material previously published except where due reference is made. I give permission for the digital version of this thesis to be made available on the web, via the University of Adelaide's digital thesis repository, the Library Search and through web search engines, unless permission has been granted by the School to restrict access for a period of time.

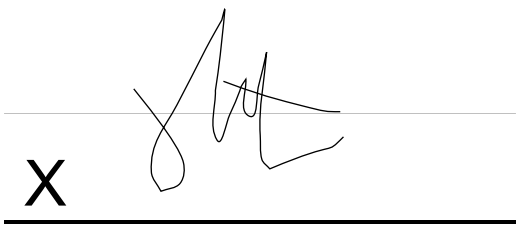
X

Joel Smith
Bachelor of Psychological Science (Honours)

Date: 27/09/2021

Contribution Statement

The research topic of this thesis was developed in conjunction with my supervisors. I completed all thesis writing and formatting, study screenings, data extractions, data transformations and categorisations, author contacting, analyses, result interpretations, and risk of bias assessments. However, my supervisors helped validate study screenings at the title and abstract level, data extractions, data transformations, and author contacting forms. Otherwise, my supervisors provided drafting as per thesis drafting guidelines and so, contributed to some formatting and grammar corrections.



X

Joel Smith
Bachelor of Psychological Science (Honours)

Date: 27/09/2021

Acknowledgements

I would like to acknowledge the contributions of my supervisors, Andrea Gordon, Paul Delfabbro, Matthew Stevens, and Robert Ali, who provided invaluable advice and support throughout my thesis year. I am especially grateful for their patience in teaching me an entirely new form of analysis and research. I would also like to thank the Library Liaison who helped me develop my search terms and the authors contacted for data requests for taking the time to respond and provide new data.

A Systematic Review and Meta-Analysis of the Effectiveness of Brief Interventions for Reducing Antenatal Alcohol Consumption

Antenatal alcohol consumption increases the risk of adverse pregnancy outcomes (DeJong et al., 2019). Despite this, many women continue to drink alcohol during pregnancy (National Health and Medical Research Council [NHMRC], 2020). Screening and brief intervention (SBI) is recommended for identifying and managing alcohol use during pregnancy (World Health Organisation [WHO], 2014). Although, this has limited evidentiary support (Gomez et al., 2020) and subsequently questionable clinical applicability (Chiodo et al., 2019; O'Brien, 2014). The current systematic review provides background on antenatal alcohol consumption and the application of SBIs for its treatment and is founded on a review of notable methodological and conceptual issues in the literature. A meta-analysis will be conducted to investigate the effects of brief interventions on antenatal alcohol consumption frequency, quantity, and abstinence, with an aim of verifying current healthcare recommendations. Further subgroup analyses and meta-regressions will be undertaken to assess variables that may influence these outcomes, with aims of verifying current healthcare recommendations. This study will differ from the recent systematic review and meta-analysis by Gomez et al. (2020) in that it will investigate patterns of consumption regarding frequency and quantity outcomes, assess potential determinants of variations, and will provide an alternative definition of brief intervention judged to be better aligned with existing classifications (Australian Government Department of Health [AGDH], 2004; Rodgers, 2018; WHO, 2014). It will also provide insight into how current research practices can be standardised, made more ethical, and conducted with less risk of bias.

Overview of Antenatal Alcohol Consumption

When consumed, alcohol is absorbed into the bloodstream where ninety-percent is metabolised in the liver, and ten-percent expelled by the lungs, kidneys, and sweat

(Cederbaum, 2012). In pregnant women, the bloodstream is in part absorbed into fetal circulation via the placenta (NHMRC, 2020), including any alcohol concentrated within it (Popova et al., 2017). Unlike its mother, the fetus is limited in its ability to metabolise alcohol (Burd et al., 2012). The pathways for fetal alcohol metabolism in the liver are significantly underdeveloped during the first two months of gestation and, even after, function at around 5-10% capacity compared to a healthy adult (Burd et al., 2012; Cederbaum, 2012). Lung and kidney metabolic pathways are also limited, as much of the alcohol they expel re-enters fetal circulation by swallowing and reabsorption during the recycling of the amniotic fluid (Gilbert, 2006). The fetus is therefore unable to sufficiently metabolise or expel alcohol and is at substantial risk of prolonged alcohol exposure (Burd et al., 2012; Popova et al., 2017).

Effects on the Fetus

Adverse outcomes for alcohol-exposed pregnancies are numerous and ranging in severity. Women who consume alcohol while pregnant are at greater risk of experiencing spontaneous fetal death (Andersen et al., 2012; Sundermann et al., 2021), stillbirth (Kesmodel et al., 2002), preterm birth, low birthweight, and growth retardation (O'Leary et al., 2009; Patra et al., 2011), compared to those who do not. Fetuses exposed to alcohol are also at risk of developing fetal alcohol spectrum disorders (FASD; DeJong et al., 2019). FASD is the umbrella term used to describe the group of physical and cognitive disabilities experienced by a child exposed to alcohol in the womb (Rasmussen et al., 2008). The term was developed from the existing fetal alcohol syndrome (Jones & Smith, 1973), which is now considered the most severe version FASD (DeJong et al., 2019). Symptoms include physical deformities (Carson et al., 2017; Stoler et al., 2004), reduced intelligence, neurological deficits (such as to memory, learning, and judgment), and issues with behaviour, attention, and sleep (Centers for Disease Control, 2021a; DeJong et al., 2019; O'Malley & Nanson, 2002). FASD symptoms and their severity vary greatly by child (Riley et al., 2011;

Wattendorf & Muenke, 2005). FASD symptoms tend to persist throughout childhood and early adulthood, although knowledge is limited for how they present in later-life (Moore & Riley, 2015) and some cognitive deficits have been shown to improve with age (Rangmar et al., 2015).

The risk of adverse pregnancy outcomes varies with consumption frequency and quantity (NHMRC, 2020). Current evidence, while limited, suggests low-level drinking has minimal or no effect on birth outcomes (Carson et al., 2017; Flak et al., 2014; Mamluk et al., 2017). Similar inconclusiveness exists around the effects of low-level consumption on FASD development (Butt et al., 2011). However, it is difficult to determine a consistent effect of low-level drinking on pregnancy outcomes due to discrepancies in how it is classified across studies. For example, moderate-level drinking during early pregnancy was shown to increase risk of preterm and low-weight births when defined as one or more standard drinks per day (Jaddoe et al., 2007), which is fewer drinks than used to define low-level drinking elsewhere (Carson et al., 2017). The NHMRC (2020) considers low-level drinking to be less than 1.4 standard drinks per day, or 10 per week. This definition will be adopted for the current review, although it should be noted that this does not align with definitions from most existing studies (Carson et al., 2017; Flak et al., 2014; Mamluk et al., 2017). Nonetheless, the risk of adverse birth and child outcomes is generally considered to be greater as levels of consumption increases (Butt et al. 2011; NHMRC, 2020). Alcohol use during critical fetal development periods in the second and third trimesters and high-quantity binge drinking are otherwise shown to increase risk (Ikehara et al., 2019; May et al., 2013).

Health Recommendations

Abstinence from alcohol consumption during pregnancy is recommended by Australian and international health agencies (Warren, 2015). Included are the NHMRC (NHMRC, 2020), WHO (Schölin, 2016), and the United States Surgeon General (U.S. Office

of the Surgeon General, 2005). These recommendations have evolved from earlier advice that consuming two standard drinks per day or seven per week was safe during pregnancy (O'Leary, 2007; Warren, 2015).

Epidemiology

Despite current recommendations, the Australian Institute of Health and Welfare (AIHW) reports that 55% of Australian women consume alcohol before knowing they are pregnant, with 14.5% continuing to do so after becoming aware (AIHW, 2021). These rates have steadily decreased since 2007 (AIHW, 2014, 2021), though remain greater than the estimated global average of 9.8% (Popova et al., 2017). Of Australian women who report drinking alcohol during pregnancy, 22% do so frequently (two or more drinking occasions a month; AIHW, 2020), and 1.4% do so in high quantities (more than six standard drinks on a single occasion; AIHW, 2014). However, most women (96-97%) are reported to consume only 1-2 standard drinks total across gestation (AIHW, 2014, 2020), indicating levels of use not decisively linked to adverse pregnancy outcomes (Butt et al., 2011; Carson et al., 2017).

The difference in prevalence rates between women who are aware of their pregnancies and those who are not reflect observed patterns of use across gestation and the phenomenon of spontaneous cessation (Stanesby et al., 2018). Spontaneous cessation refers to when a pregnant woman suddenly stops consuming alcohol upon learning that she is pregnant and is reported to occur in 70-90% pregnancies (AIHW, 2021; Centers for Disease Control, 2021b; Harrison & Sidebottom, 2009; Muggli et al., 2016; Ockene et al., 2002; Pirie et al., 2000; Roberts et al., 2014). Women become increasingly likely to abstain from alcohol across the first trimester, with use rates then remaining generally low and consistent for the rest of pregnancy (Alshaarawy, et al., 2016; Muggli et al., 2016). Spontaneous cessation is attributed to the potent motivational effect that pregnancy has on women to engage in health-positive behaviours (Solomon & Quinn, 2004).

Patterns of use for Australian women of different demographics are mixed. An analysis of 1318 pregnant Australians found that women who drank any alcohol during pregnancy were more likely to be white, smokers, high-SES (earning over AUD$100,000 per year), and have an unplanned pregnancy (Muggli et al., 2016). Differences in maternal age, relationship status, education and parity did not produce meaningful differences in risk (Muggli et al., 2016). Another review corroborated there being lower risks of consumption for women of low-SES (determined by having a government HealthCare card for low-income earners), though also observed decreased risk with maternal age (Anderson et al., 2013). Risk was also shown to be lower in women with ten or less years schooling compared to those who were university-educated, with no significant differences observed between any other education levels (Anderson et al., 2013). Risk has otherwise been shown to increase with age, income, having a partner (Giglia & Binns, 2007), and living in a major city (AIHW, 2020).

Impact and Cost

The impact of antenatal alcohol consumption in Australia is unclear. There exists no estimate of the number of pregnancies affected by alcohol consumption in Australia, although current approximations suggest that 0.02 per 1000 non-indigenous children, and between 0.047-2.76 per 1000 indigenous children experience FASD (Bower et al., 2000; Harris & Bucens, 2003). The social and economic costs of antenatal alcohol consumption to Australia are also unclear due to a lack of throughout investigation (Andersson & Elliot, 2018; National Indigenous Drug & Alcohol Committee, 2012). International estimates of the global burden of FASD suggest an average cost of USD$23,804 (AUD$33,336) per person, although this approximation may not be applicable to the Australian context due to its overreliance on data from the United States, and observations of large variations in cost between countries (Greenmyer et al., 2018). Valuations from New Zealand may provide better insights, where

annual costs of $NZ49-200 million (AUD$47-191.8 million) are estimated from productivity loss alone (Easton et al., 2016).

Treatment of Antenatal Alcohol Consumption

The WHO recommends that all women seeking antenatal care be screened for alcohol use during their earliest antenatal care visit, and that brief interventions be provided should risk be identified (2014). However, SBI alone is not sufficient treatment for pregnant women who are unable to cease use or present with high-level use or dependencies (Glass et al., 2015, 2017). These women should also be referred to more intensive treatment, not limited to extended psychosocial interventions, detoxification services, and, depending on individual circumstances, pharmacological interventions (WHO, 2014).

Screening and Brief Intervention

SBI for alcohol use requires an initial screen for consumption, symptoms of dependency, and or risk-drinking behaviours, and a brief intervention aimed at reducing use (Rodgers, 2018; WHO, 2014). This cost-effective practice (Olmstead et al., 2019) provides clinicians a means of identifying women at risk of alcohol-exposed pregnancies and a platform for providing advice and feedback around use (Haber et al., 2009). SBI does not require extensive time to administer and, if a patient is assessed as having no risk, can be ended without intervention. This ensures that clinicians can assess all patients, including those who may not voluntarily disclose use out of fear of stigma or repercussions, without wasting valuable session time (WHO, 2014). Assessments can otherwise be conducted in general healthcare settings and by non-doctor clinicians, such as nurses (Substance Abuse and Mental Health Services Administration, 2011), thereby reducing the burden on understaffed and underfunded specialist alcohol services.

Numerous alcohol use screening tools have been used to assess antenatal alcohol consumption (WHO, 2014). The effectiveness of these tools is determined by their sensitivity

to detect true positive screens and specificity to reject true negatives (Chang, 2001). The 4-item T-ACE (Tolerance, Annoyed, Cut-down, Eye-opener; Sokol et al., 1989) and 5-item TWEAK (Tolerance, Worry, Eye-opener, Amnesia, K[C]ut down; Russell, 1994) are the most employed screening tools, having been developed specifically for use with pregnant women (Chiodo et al., 2019). These both show high sensitivity to identify risky drinking in pregnant women (Chang 2001, 2010), although the T-ACE has reported low specificity (Carson et al., 2017; Chiodo et al., 2019). Altered T-ACE measures have, however, been developed with higher cut-offs and subsequently greater specificity (Chiodo et al., 2014). The 10-item AUDIT (Alcohol Use Disorders Identification Test; Babor et al., 2001) is also popular in the literature (Carson et al., 2017), despite not being validated for use with pregnant women due to low sensitivity to detect risky drinking in the demographic (Chang, 2010; Chiodo et al., 2019). Otherwise, the WHO-endorsed (WHO, 2010) ASSIST (Alcohol, Smoking and Substance Involvement Screening Test; Ali et al., 2002), has been developed as means of connecting screening to structured intervention. The ASSIST shows fair agreement with the T-ACE for alcohol risk screening in pregnant women (weighted kappa reliability score of 0.2513; Hotham et al., 2013).

Brief interventions are typically in-person, 5-30 minutes in length, and single session (Rodgers, 2018; WHO, 2014). However, some may be computer-delivered, take up to 60 minutes, or be administered as multi-session programs (AGDH, 2004; Gomez et al. 2020). Most brief interventions are influenced by motivational interviewing (MI) techniques (Gomez et al., 2020), which aim to address ambivalence toward drinking behaviours in a non-judgmental manner to motivate and sustain change (Wagner & McMahon, 2004). MI is proven effective with populations who use alcohol (National Institute on Drug Abuse, 2018; Wagner & McMahon, 2004), including pregnant women (Osterman, 2011), and can be easily administered in the limited timeframes predicated by a brief intervention (DiClemente et al.,

2017). The inclination for pregnant women to engage in health-positive behaviours during pregnancy otherwise make them good candidates for these motivational interventions (Solomon & Quinn, 2004).

Measurement in Research

Screening tools assess risk of harm from alcohol consumption, but do not provide specific indicators of frequency and quantity of use. Therefore, in research, screening tools are accompanied by use measurement tools so that changes in use or group differences can be assessed. The most common use measurement tool is the Timeline Follow-back interview (TLFB; Sobell & Sobell, 1992), which is a highly reliable and valid means of assessing retrospective drinking up to twelve-months (Carson et al., 2017). Other common tools include the Brief Drinker Profile (BDP; Miller & Marlatt, 1987a, 1987b) and the Quick Drinking Screen (QDS; Sobell et al., 2003). The QDS is shown to yield similar results to the TLFB for women during the preconception phase (Dum et al., 2009), although neither have been validated for use with pregnant women.

Current Issues in the Literature

The use of brief interventions to manage antenatal alcohol consumption in pregnant women is recommended (WHO, 2014), despite inconclusive support from the literature (Glass et al., 2017; Rodgers, 2018), with persistent observations of small, insignificant effects of treatment (Gomez et al., 2020). These effects are in part a consequence of spontaneous cessation and assessment reactivity. Assessment reactivity refers to how screening alone provides therapeutic benefit to reduce alcohol consumption behaviours (Bernstein et al., 2010; Epstein et al., 2005). These phenomena reduce the ability to observe meaningful differences in alcohol use over time and between groups, as participants who receive assessment-only controls may, without intervention, reduce use at rates substantial enough to conceal treatment benefits (Kypri et al., 2007). These concerns around effectiveness have

contributed to healthcare workers questioning the clinical applicability of brief interventions and subsequent reports of low rates of SBI administration (Chiodo et al., 2019; O'Brien, 2014) and follow-up referrals (Ordean et al., 2020). Concerns have also been raised over clinicians lacking the time and training to correctly administer SBIs, fears of alienating new clients, and the potential for false reporting (O'Brien, 2014; Wouldes et al., 2021). These issues highlight the need for decisive evidence in favour of SBI, not only to validate recommendations, but also to ensure that otherwise sceptical practitioners are supportive enough of SBI to consistently apply it in clinical practice.

Some of these issues have been addressed in a recent systematic review and meta-analysis by Gomez et al. (2020), who synthesised effects of brief interventions on abstinence rates in sample populations. However, Gomez et al. (2020) was limited in that it did not address outcomes related to alcohol use frequency and quantity, despite the critical effects of increasing frequency and quantity of consumption on adverse pregnancy outcomes (Butt et al. 2011; Carson et al., 2017; Flak et al., 2014; Jaddoe et al., 2007; Mamluk et al., 2017; May et al., 2013; NHMRC, 2020). Abstinence outcomes are important to assess, but do not provide information around patterns of use. Gomez et al. (2020) otherwise adopted a definition of brief intervention that does not appear comparable to existing classifications (AGDH, 2004; Rodgers, 2018; WHO, 2014), with multi-session programs and non-individualised interventions included for their review.

Importance of this Review

The provision of brief interventions for antenatal alcohol consumption appears clinically feasible. They apply proven motivational interviewing techniques to elicit behaviour change in a population that is already driven to adopt health-positive behaviours (Osterman, 2011; Solomon & Quinn, 2004). However, the lack of evidence in support of these interventions (Gomez et al., 2020) and their poor clinical applicability (O'Brien, 2014)

raise concerns about the legitimacy of healthcare recommendations such as those made by the WHO (2014). It is therefore important for a systematic review and meta-analysis of the available research to be conducted to assess the effect of brief interventions on antenatal alcohol consumption, especially regarding frequency and quantity outcomes, which play important roles in the health of pregnancies, but are yet to be investigated through meta-analysis. In doing so, insight into means of improving research practices, validating current healthcare recommendations, addressing concerns about statistical insignificance and clinical applicability, and optimising outcomes for pregnant women can be provided.

Objectives

To address concerns around the legitimacy of current healthcare recommendations and the clinical insignificance of existing literature, a systematic review and meta-analysis will be conducted on the effectiveness of brief interventions for reducing antenatal alcohol consumption. This will be assessed per six primary outcomes: alcohol consumption frequency at follow-up, frequency change, quantity at follow-up, quantity change, abstinence rate at follow-up, and abstinence maintenance rate (percent who started and finished abstinent). At follow-up and change data will be assessed so that between- and within-group differences can be analysed, and a more comprehensive understanding of effects established. Secondary assessments of potential predictors and moderators of effectiveness will also be conducted (secondary variables in Appendix A).

Method

The present systematic review and meta-analysis was conducted in accordance with the Cochrane Handbook for Systematic Reviews of Interventions (version 6.2; CHSRI; Higgins et al., 2021d) and the Methodological Expectations of Cochrane Intervention Reviews (MECIR; version February 2021; Higgins et al., 2021a).

Eligibility Criteria

The eligibility criteria employed for this review are presented in Table 1, with justifications included for each. Note that eligibility for review did not necessarily mean a study was eligible for analysis.

The present meta-analysis defined a brief intervention as being single-session and ≤60 minutes in length, per existing classifications (AGDH, 2004; Rodgers, 2018; WHO, 2014).

Eligibility for Analysis

Studies were included for review if they reported at least one primary outcome, even if it were not viable for analysis. No exclusions were made for a lack of secondary data. However, to be eligible for analysis, data needed to be presented using a unit of measurement that either met analysis requirements or could be transformed or categorised to meet them. Primary outcome data was required to compare an intervention and control group. Secondary data was required to represent the total sample, not either group individually.

Table 1

Study Eligibility Criteria with Justifications

Criterion	Sub-criterion	Eligibility criteria	Ineligibility criteria	Justification
Publication details	Year	Published 1974 onwards	Published 1973 and earlier	Shift in understandings in research with the development of the fetal alcohol syndrome diagnosis in 1973 (Jones & Smith, 1973)
	Language	English	Non-English	Lacking translation resources
	Location	Any location	-	Maximise number of viable studies
	Journal	Any peer-reviewed journal	-	Maximise number of viable studies
Study details	Experiment type	Randomised control trial	Non-randomised control trial	Increased risk of bias and reduced ability to determine causal effects from non-randomised control trials (Reeves et al., 2021)
	Pilot status	Pilot or not pilot	-	Maximise number of viable studies, although It is noted that pilot studies possess greater risk of imprecision due to small samples (Leon et al., 2011). Pilot status tested as a secondary outcome rather than being an eligibility criteria
	Risk of bias/ quality	Any level of risk of bias	-	Existing work has identified high risk of bias in most relevant studies (Gomez et al., 2020; WHO, 2014)
Participant details	Species	Human	Animal	Scope is for human women
	Pregnancy status	Pregnant	Non-pregnant	Scope is for pregnant women (preconception and postnatal women excluded)
	Alcohol use	Screened positive for risk of harmful use	Screened negative for risk of harmful use	Women screened negative would not meet WHO (2014) eligibility criteria for brief intervention. Also due to inherent differences in approaches toward onset compared to indicative prevention (O'Connell et al., 2009). Note that harmful use may differ between screening tools.

Criterion	Sub-criterion	Eligibility criteria	Ineligibility criteria	Justification
	Dependency status	Some of sample includes women with dependencies	Full sample is women with dependencies	Maximise number of viable studies. However, it is noted that samples containing any percentage of dependent women are potentially unethical as WHO (2014) guidelines would also recommend referral to more intensive treatment. It is for this reason that samples of only women with dependencies were excluded
	Abstainer status (at base)	Samples with and without abstainers included	-	To investigate the effect of brief intervention on abstinence maintenance
Screening details	Drug assessed	Alcohol use only	Polydrug use	Scope is for alcohol consumption. Required that independent alcohol SBI and outcomes were reported
Intervention details	Number of sessions	Single session	Multi-session	Per existing classifications (AGDH, 2004; WHO, 2014; Rodgers, 2018). Studies testing multi-session programs were excluded, even where single-session data from participant attrition may have existed, as this was not considered comparable to data from intentionally single-session interventions
	Length	≤60 minutes	>60 minutes	Per existing classifications (AGDH, 2004; WHO, 2014; Rodgers, 2018)
	Time of intervention delivery	Antenatal period	Any other period	Scope is for antenatal alcohol consumption
	Recommendation around use	Abstinence, personalised goal setting, or harm minimisation	Any suggestion of some level of consumption being safe	Interventions adhering to outdated classifications of there being safe levels of alcohol consumption in pregnancy (outlined in Warren, 2015) were excluded

Frequency data were measured in drinks per month, where a month was defined as 28-days. This format was to allow easy conversions from per week data. Quantity data were measured as drinks per drinking day, and abstinence data were measured as dichotomous outcomes: currently drinking and not currently drinking. Frequency and quantity data reported in episodes or using biological units of measure, such as blood alcohol content or hair ethyl glucuronide quantification, were excluded due to incompatibility with the defined units. Episodic data was otherwise excluded for being open to significant variation and not possessing a maximum value, such that one episode could reflect one drinking day, or continued drinking every day of assessment. These exclusions did not apply to abstinence outcomes as they were not reliant on these units of measurement. Abstinence maintenance rate data required that baseline abstainers were assessed independent of other participants at follow-up. Viable units for secondary data are provided in Appendix A.

Search Strategy

A high sensitivity search was employed to maximise comprehensiveness (Lefebvre et al., 2021) per recommendations that literature searches should be as sensitive as possible given time allowances (Sampson et al., 2003). However, some considerations were made to adhere to CHSRI guidelines (Lefebvre et al, 2021), including running practice searches to identify and remove terms that were recalling majority irrelevant or overlapping reports. The search was otherwise limited to studies published from 1974 onwards to coincide with the invention of the FAS diagnosis (Jones & Smith, 1973).

Database selection

Seven databases were included to ensure a comprehensive search (Arendt, 2007), sensitive to studies unique to specialised databases (Bramer et al., 2017). Database selection was informed by consultations with the University of Adelaide's Faculty of Health and Medical Sciences Library Liaison, existing meta-analyses on screening and or intervention

for alcohol use in pregnant women (Fergie et al., 2019; Gomez et al., 2020; Jonas et al., 2012), and CHSRI recommendations (Lefebvre et al., 2021). Literature searches were conducted on the PsycINFO, PubMed, Embase, Scopus, CINAHL, and Cochrane Library (all databases) electronic databases during April 2021. Medline was not searched as it uses the same database as PubMed (Lefebvre et al., 2021). The Australian Indigenous HealthInfoNet database (journal articles, online journal articles, reports, and theses only; https://healthinfonet.ecu.edu.au/key-resources/publications/) was searched in June 2021 to account for any indigenous-specific studies that may not have been included in mainstream databases. The formal search process was supplemented by a web search in May 2021, and citation searches of studies included for full-text review and relevant meta-analyses.

Search terms and term sets

Search strategies were structured to combine population, behaviour and treatment term sets (Table 2) using the Boolean operator AND. These term sets were adapted from CHSRI recommendations of population, treatment, and study type term sets (Lefebvre et al., 2021). The study type term set was replaced with the behaviour term set to increase the precision of the search to alcohol related studies. This methodology has previously been demonstrated in a Cochrane review of interventions for pregnant women in outpatient illicit drug programs (Terplan et al., 2015). Search strategies were adapted for each database (Appendix B). Relevant database-specific subject headings were included where possible, but otherwise substituted with a consistent range of keywords related to each term set.

Table 2

Search Term Sets

Term set	What terms related to
Population	Related to pregnancy and included postpartum and breastfeeding terms to increase sensitivity to studies where changes to antenatal drinking were examined postnatally
Behaviour	Behaviour terms related to alcoholic beverages and drinking behaviours and excluded chemical terms (such as ethanol) to limit the capture of animal studies
Treatment	Treatment terms included those related to screening, intervention, and general psychotherapeutic techniques for reducing alcohol consumption, to maximise sensitivity to a variety of methods

Data Collection

Study records were exported to Covidence (Veritas Health Innovation, n.d.), a Cochrane-endorsed software (The Cochrane Collaboration, n.d.) used to aid systematic reviews and meta-analyses. Reports gathered from the Aboriginal HealthInfoNet, web searches and citation searching were not stored in Covidence unless escalated to title and abstract screening. Covidence automatically completed a screen for duplicates. Studies were then screened for relevancy and further duplicates by the first author (JS). Title and abstract screening and full-text reviews were also completed by JS. Studies marked by JS for full-text review were independently screened by the third (MS) and fifth (AG) authors. Disagreements were resolved by consensus (Appendix C), noting that some studies had their full texts reviewed by JS to gather further information on inclusion criteria adherence. Once disagreements were resolved, studies were escalated to data extraction.

Data Extraction and Management

Data were manually extracted from the text, tables, and figures of eligible studies by JS. No extraction software was used. Data items were collected in accordance with study outcomes, eligibility criteria, and referencing and author contacting requirements (listed in

Appendix D) and input into a spreadsheet with the same headings. MS and AG validated individual data items only when requested.

Management of Multiple Study Arms and Measurement Points

When studies included multiple arms assessing independent samples, all arms that met eligibility criteria were considered for analysis. When arms assessed overlapping samples, the arm analysing the full sample without consideration of any moderator variables was included. Where data were collected at multiple time-points, the time-point representing the largest measurement period was selected, bar those including postnatal use, as this would present the greatest opportunity for the intervention to take effect/habits to develop (Lally et al., 2009).

Author Contacting

Data were investigated prior to contacting authors. If further information or validation of existing data were required, then authors were contacted (per CHSRI recommendations, Deeks et al., 2021). First authors were contacted where possible, though second authors were contacted if required. Data requests were made using a template email (Appendix E) and request form (Appendix F) and adapted for the data requirements of each study.

Data Transformation Processes

Data transformations were conducted by JS to convert primary and secondary outcome data to viable units of analysis (transformation guidelines in Appendix G, data categorisation guidelines in Appendix H). Transformations made to primary outcome data were validated by MS.

Analysis Plan

Meta-analyses of primary outcomes and subgroup analyses of secondary data were conducted in Review Manager (RevMan; version 5.4.1; The Cochrane Collaboration, 2020). Meta-regressions of secondary data and failsafe-N calculations for primary outcomes were

undertaken using Meta-Essentials (Suurmond et al., 2017). All statistics were automatically generated by the relevant software. All analyses were conducted by JS.

Meta-Analyses

Meta-analyses were conducted on all primary outcomes with two-or-more studies, although it is noted that meta-analyses with so few studies may not provide sufficient power to observe significant effects (Deeks et al., 2021; Valentine et al., 2010). Outcomes with only one viable study had comparison tables and funnel plots generated, but without interpretation. Primary treatment effects were reported as standardised mean differences (Hedges' *g*) for continuous outcomes: frequency at follow-up, frequency change, quantity at follow-up, and quantity change, and as odds ratios (OR) for dichotomous outcomes: abstinence rate and abstinence maintenance rate. Hedges' *g* reflects the mean standardised difference in effect sizes between groups and was selected as it corrects for small sample biases typical of the alternative, Cohen's *d* (Deeks & Higgins, 2010). ORs were selected over risk ratios, which are somewhat preferred by the CHSRI (Deeks et al., 2021), to allow better comparability to Gomez et al. (2021). Corresponding 95% confidence intervals (CI) were also generated. Random effect models were employed due to variation in study methodologies (Deeks et al., 2021), with inverse variance weighting methods applied to continuous outcome calculations and the Mantel-Haenszel method (Mantel & Haenszel, 1959) to dichotomous outcomes. While inverse variance methods are more common, the Mantel-Haenszel method is more appropriate for meta-analyses of continuous outcomes in smaller samples (Deeks et al., 2021).

Hedges' *g* values were interpreted such that a score of 0.2 represented a small change, 0.50 a medium change, and 0.8 a large change (Lakens, 2013). Outcomes were considered significant when the *p*-value of the test of overall effect, *Z*, was below 0.05 and when the 95% CI did not cross zero, thereby rejecting the possibility that the true value contained the

null (per CHSRI guidelines, Schünemann et al., 2021b). A minimum effect size for which data would be considered significant could not be established for frequency, quantity, and abstinence maintenance outcomes as no meta-analyses exist as precedent. A minimum effect size of 1.61 in favour of intervention was set for the abstinence rate outcome, reflecting the lower CI of the Gomez et al. (2020) meta-analysis. MECIR guidelines state that significance should not be assessed as the presence of significant evidence may not necessarily be evidence of a significant effect (Higgins et al., 2021a). Significance was reported regardless to meet study requirements. Multiple significance indicators were assessed to ensure the robustness of findings. Grading of Recommendations, Assessment, Development and Evaluations (GRADE; Schünemann et al., 2021a) assessments were applied to identify the certainty of evidence for each outcome.

Heterogeneity was assessed using Chi^2 (df, *p*), and or I^2 values. Data were judged to be heterogenous if the Chi^2 value were significant, or the I^2 percentage above 40% (Deeks et al., 2021). Narrative reports of study characteristics and their heterogeneity was also provided. Funnel plots were generated to assess reporting bias and the robustness of significant effects. Failsafe-N values were also calculated should funnel plots include less than ten studies, and therefore be uninterpretable (Sterne et al., 2011). These failsafe-N values were calculated using the Rosenthal method (Rosenthal, 1979), which determines the number of missing insignificant studies required to nullify the observed effect, as this is most common method used in psychological sciences (Ferguson & Brannick, 2012). The size of the fail-safe N was automatically determined in Meta-Essentials using the Rosenthal (1979) ad hoc rule (van Rhee et al., 2015). No sensitivity analyses to assess the fragility of summary effects were conducted, though abstinence rate data were narratively compared to that of Gomez et al. (2020).

Subgroup Analyses and Meta-Regressions

Subgroup analyses and meta-regressions of secondary data were conducted on all viable primary meta-analyses regardless of whether heterogeneity was detected, although it is acknowledged that these analyses are intended to identify sources of heterogeneity (Deeks et al., 2021). This avoided data-dredging or post-hoc analyses being conducted, which the CHSRI recommends against due to increased risks of finding false indicators of heterogeneity (Deeks et al., 2021). It is also noted that secondary analyses with few data sources may not be interpretable (Deeks et al., 2021; Valentine et al., 2010). The same effects models and statistical methods applied to meta-analyses were also employed with secondary analyses.

Effect sizes, significance, and heterogeneity assessments were provided for subgroup analyses as per primary meta-analyses. Subgroup differences were assessed as meaningful when all groups reported significant effects, the 95% CIs of these effects did not overlap (Deeks et al., 2021), and when heterogeneity of subgroups was identified, such that meaningful differences between groups were present (Borenstein et al., 2009). The presence of a significant effect or heterogeneity of one group was not considered indicative of overall subgroup significance (Deeks et al., 2021).

Meta-regression effects were analysed using standardised and unstandardised (with standard errors) regression coefficients, where effect sizes were judged per Hedges' *g* categories (Lakens, 2013). Significance was assessed for the unstandardised regression coefficients as a p-value <0.05.

Risk of Bias Assessments

All study outcomes were assessed for quality and risk of bias using the CHSRI-recommended bias types and risk rating scale (explained in Appendix I; Higgins et al., 2021c), regardless of whether they were valid for analysis. Studies that did not report relevant data were not assessed. The inter-reliability of risk of bias judgements were assessed at the

domain and overall judgement levels against Gomez et al. (2020) ratings, with weighted kappa statistics and level of agreement ratings generated using GraphPad (GraphPad Software, 2014). Level of agreement was assessed per Table 3.

Table 3

Scale for the Interpretation of the Kappa Statistic

Kappa	Level of agreement
≤0.20	None
0.21-0.39	Minimal
0.40-0.59	Weak
0.60-0.79	Moderate
0.80-0.90	Strong
>0.90	Almost perfect

Note. Adapted from "Interrater reliability: The kappa statistic" by M. L. McHugh (2012). Biochemia Medica, 22(3), 276-282. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/.

Management of Missing Data

Missing or non-viable data were managed by contacting authors for further information and or data transformations. If no viable data could be generated, then outcomes and or studies were excluded from analysis. The potential effects of these exclusions will be narratively discussed, with reference to outcome funnel plots and failsafe-N values.

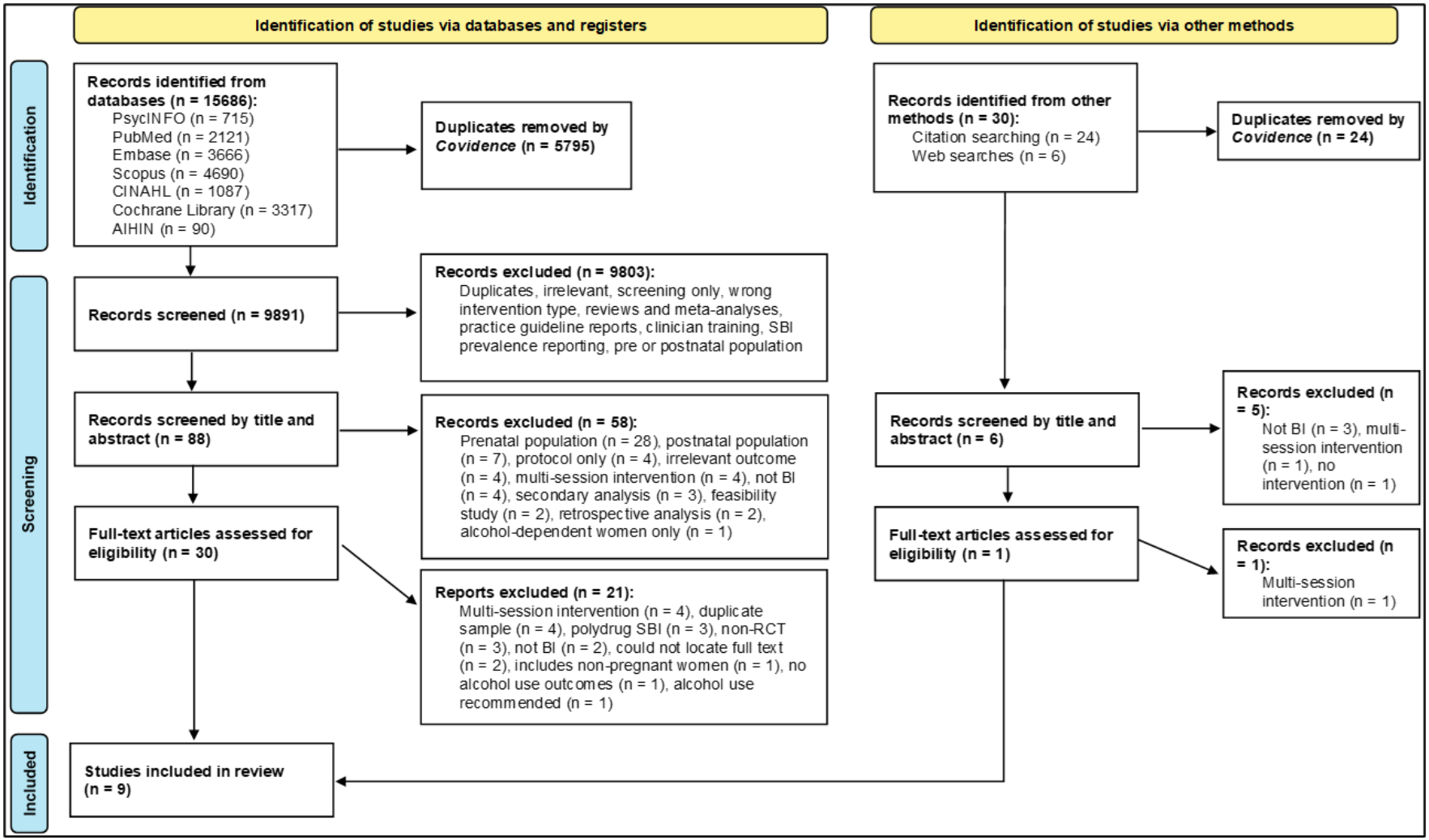
Results

Results of Search

A total of 15686 reports were identified through database searches and 30 using other methods. Covidence automatically removed 5795 duplicates, with a further 9827 studies removed during manual screening for relevancy and duplicates. Ninety-four studies were screened by title and abstract, with thirty-one progressing to full-text review. Nine studies were included for review and eight for analysis (complete flow diagram in Figure 1).

Figure 1

Flow Diagram of the Screening Process



Note. Created based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 Reporting Guidelines (Page et al., 2021).

Author Contacting

Further information and or data validation was sought for eight of the nine studies included in data extraction and review (Appendix J). Contact details for the first three authors of the remaining study (Joya et al., 2016) could not be located, so no communications were made. Five responses were received, with two including new data (Appendix J). Three authors did not respond.

Characteristics of Included Studies

Nine individual randomised-control trials published from 1995-2016 were included for review (Chang et al., 1999, 2005; Handmaker et al., 1999; Joya et al., 2016; Ondersma et al., 2015; Osterman & Dyehouse, 2012; Osterman et al., 2014; Reynolds et al., 1995; Tzilos et al., 2011), representing 1129 participants at base, and 1063 at analysis. Tzilos et al. (2011) was excluded from analyses for lacking viable outcome data. Eight were conducted in the United States and one in Spain (Joya et al., 2016). All studies sampled participants from clinical settings, with only one administering interventions in different locations (two alternative clinical settings and the participants' homes; Handmaker al., 1999). A summary of characteristic distributions is provided in Table 4, with a comprehensive outline of individual study characteristics in Appendix K.

Table 4

Summary of Distributions of Study Characteristics

Study characteristic	Summary of distribution
Alcohol use risk level	Six studies included women with high-level use or dependencies, three excluded them. Six studies included baseline abstainers, one excluded them, and two did not report baseline abstinence
Age	Mean age of participants ranged from 22.4-31.4 years old, though two studies did not report mean ages and the minimum and max ages were unclear
Ethnic majority	Five studies had a black ethnic majority, two had a white majority, and one had a Hispanic majority. One study reported a Spanish/other ethnic majority

Study characteristic	Summary of distribution
Education	Seven studies reported the mean or median level of education to be ≥HSG, with one reporting a fifty-fifty split. Mean percentages of ≥HSGs ranged from 50-96%
Marital status	Rates of marriage were mixed, ranging from 18-80.5%. One study did not report percent married and two reported as percent single without clarifying alternatives
Nulliparity	Nulliparity ranged from 41.4-53% in three studies with differing definitions
Gestation	Mean baseline gestation ranged from 12.2-25 weeks. Three studies did not report
Screening tool	The T-ACE was employed in four studies, the AUDIT in three, and no formal tool in two. All were conducted by research assistants, bar two that were self-administered
Use measurement tool	Use was measured using the ATFB in five studies. The BDP, QDS, and no formal tool were used in one study each. One study did not provide this information. All were conducted by research assistants, bar two that were self-administered
Use measurement period	Use was measured over a period ≥two-months in six studies, and over <two-months in three
Intervention length	Interventions ranged from 20-60 minutes, though six were 30 minutes or above. One study did not report intervention length
Delivery format	Six study interventions were delivered face-to-face, and two by computer (which were self-administered). One study did not report this data. Of the face-to-face interventions, one was conducted by a trained educator, and five by the first author. Two first authors were Doctors of Medicine, two were a psychiatric mental health clinical nurse, and one was a clinical psychology PhD candidate
Primary intervention theory	Six interventions employed motivational interviewing theories and practices, one was education-based, and two theories were not reported
Use recommendation	Three studies recommended abstinence to participants, three suggested personalised goal setting, and three did not report a recommendation
Bonus item	Two studies included a bonus item
Control type	Six interventions were compared to usual care, and three to a reduced care control

Note. HSG = high school graduate; T-ACE = Tolerance, Annoyed, Cut-down, Eye-opener; AUDIT = Alcohol Use Disorders Identification Test; ATFB = Alcohol Timeline Follow-back Interview; BDP = Brief Drinker Profile; QDS = Quick Drinking Screen.

The included studies were generally demographically similar, although employed mostly heterogenous methodologies. Reporting practices were especially varied, with no study, bar those with the same authors, reporting frequency and quantity outcomes using the same unit of measurement. Means of measuring demographic data also varied greatly across reports. These differences necessitated substantial transformations and categorisations.

Notable Inclusions and Exclusions

Notable study inclusions and exclusions are outlined in Appendix L. The partner involvement arm in Chang et al. (2005), the 30-day postpartum timepoint in Osterman et al. (2014), and the second trimester timepoint in Joya et al. (2016), were excluded.

Risk of Bias Assessments

Risk of bias analyses were conducted for eight studies (Appendix M). All outcomes were assessed as high risk. Most bias derived from issues concerning allocation sequence concealment, participant and conductor blinding, the employment of inadequate units of measurement, and not referring to protocols. Bias was greatest for frequency and quantity outcomes, with reduced concerns around post-intervention exclusions, the appropriateness of measures, and selective reporting being noted for abstinence outcomes.

Risk of bias domain assessments were 52.5% similar to those of Gomez et al. (2020; weighted kappa = 0.318, minimum level of agreement; Appendix N). All bar one overall risk of bias assessment (Joya et al., 2016) was the same (88.9%; kappa not calculable). However, these statistics cannot reasonably infer inter-rater reliability due to differences in how studies were assessed (explained in Appendix N).

Data Exclusions, Transformations, Categorisations, and Assumptions

Primary and secondary outcome data were analysed for eligibility, with exclusions, transformations, categorisations and or assumptions made where necessary (records in Appendix O for primary data and Appendix P for secondary data).

Results of Primary Analyses

An overall summary of findings table with Cochrane GRADE assessments (Schünemann et al., 2021a) is presented in Table 5. Funnel plots were generated for all outcomes (Appendix Q) but were not assessable due to a lack of studies (Sterne et al., 2011).

Table 5

Meta-Analysis Summary of Findings Table

Outcome type	Relative effect [95% CI]	Number of participants (studies)[1]	Certainty of the evidence (GRADE)	Reason for GRADE score
FFU	Hedges' *g* 0.00 [-0.21, 0.21]	393 (2)	⊕⊖⊖⊖ **very low**	Downgraded as: all studies have high risk of bias; insignificant *p* value, CI overlaps zero, suggesting inconsistency; number of events <400 and no effect of intervention suggest imprecision. No indirection
FC	-	54 (1)	-	Meta-analysis not viable
QFU	Hedges' *g* -0.03 [-0.37, 0.31]	132 (2)	⊕⊖⊖⊖ **very low**	Same as FFU, though imprecision judgment due to negative effect not no-effect
QC	-	54 (1)	-	Meta-analysis not viable
AR	OR 2.13 [1.31, 3.28]*	313 (4)	⊕⊕⊖⊖ **low**	Downgraded as: all studies have high risk of bias; number of events <400 and large CIs suggest imprecision, however, effect size is fine. No inconsistency or indirection.
AMR	OR 2.28 [1.25, 4.16]*	243 (2)	⊕⊕⊖⊖ **low**	Same as AR

Note. GRADE scoring based on Schünemann et al. (2021a). Exact *p*-values are described later. FFU = frequency at follow-up; FC = frequency change, QFU = quantity at follow-up; QC = quantity change; AR = abstinence rate; AMR = abstinence maintenance rate; OR = odds ratio.

[1] Studies that reported results with a standard deviation of 0.00 were not analysable and excluded from these figures.

* $p \leq 0.01$ (all other results are insignificant)

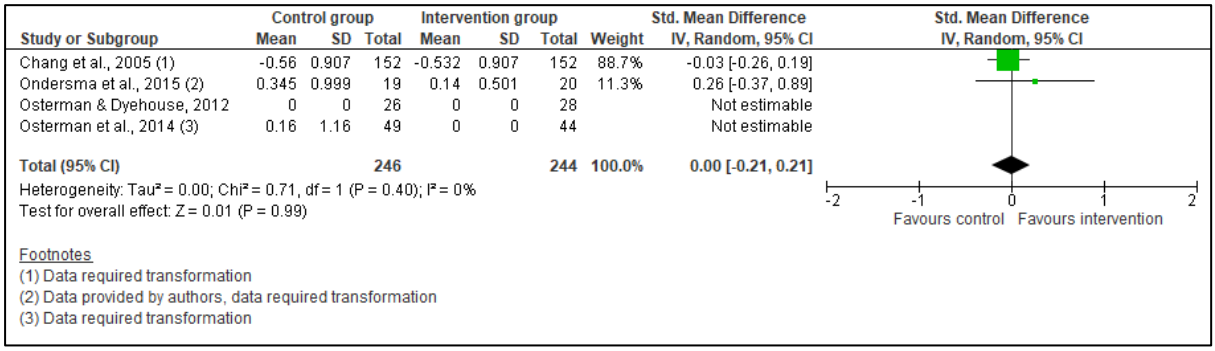
Frequency Outcomes

Four studies were included for meta-analysis of the effect of brief interventions on frequency at follow-up outcomes comparing interventions to controls, representing 490 participants. Two of these studies reported a result with a standard deviation of 0.00 that could not have their effect size estimated. Therefore, two studies were included, representing 393 participants. Meta-analysis found a Hedges' *g* of 0.00 [-0.21, 0.21] drinking days per month, indicating no effect of intervention (Figure 2). This effect was this not significant (Z = 0.01, p = 0.99, CI crossed zero). No heterogeneity was detected (I^2 = 0%).

Only one study provided viable frequency change data, representing 56 participants. Meta-analysis could therefore not be conducted (comparison table in Appendix R).

Figure 2

Comparison Table for Frequency at Follow-Up Outcome



Quantity Outcomes

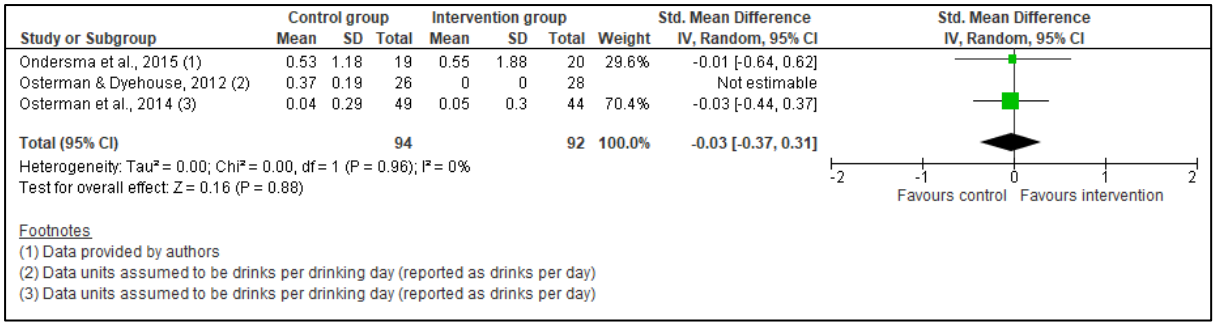
Three studies were included for meta-analysis of the effect of brief interventions on quantity at follow-up outcomes comparing interventions to controls, representing 186 participants. One study effect could not be estimated due to it reporting a standard deviation of 0.00. Therefore, two studies were included, representing 132 participants. A small, negative Hedges' *g* of -0.03 [-0.37, 0.31] drinks per drinking day was reported with an effect

of $Z = 0.16$ (Figure 3). This indicated that control groups reported 0.03 less drinks per drinking day than intervention groups. This effect was not significant ($p = 0.88$, CI crossed zero). Data were not heterogenous ($I^2 = 0\%$).

Only one study provided viable quantity change data, representing 56 participants, so meta-analysis could not be conducted (comparison table in Appendix R).

Figure 3

Comparison Table for Quantity at Follow-Up Outcome



Abstinence Outcomes

Four studies were included for meta-analysis of the effect of brief interventions on abstinence rate outcomes comparing interventions to controls, representing 313 participants. Meta-analysis found 2.13 [1.31, 3.48] times greater odds of achieving abstinence for interventions versus controls, with significant effect ($Z = 3.03$, $p \leq 0.002$, OR greater than pre-set minimum effect for significance; Figure 4). No heterogeneity was reported ($I^2 = 0\%$). High risk of publication bias was detected (failsafe-N = 9), suggesting that only a small number of insignificant data inputs would be required to nullify the observed effect. Large CIs suggest imprecision.

Two studies were included for meta-analysis of the effect of brief interventions on abstinence maintenance rate outcomes comparing interventions to controls, representing 243

participants. An OR of 2.28 [1.25, 4.16] in favour of the intervention group was reported, indicating that intervention participants were 2.28 times more likely to maintain abstinence across the reporting period (Figure 5). This effect was significant ($Z = 2.68$, $p = 0.007$) and the included data were not heterogenous ($I^2 = 0\%$). This data had high risk of publication bias (failsafe-N = 4). Therefore, few insignificant studies would be required to make the overall effect insignificant. Large CIs indicate imprecision.

Figure 4

Comparison Table for Abstinence Rate Outcomes

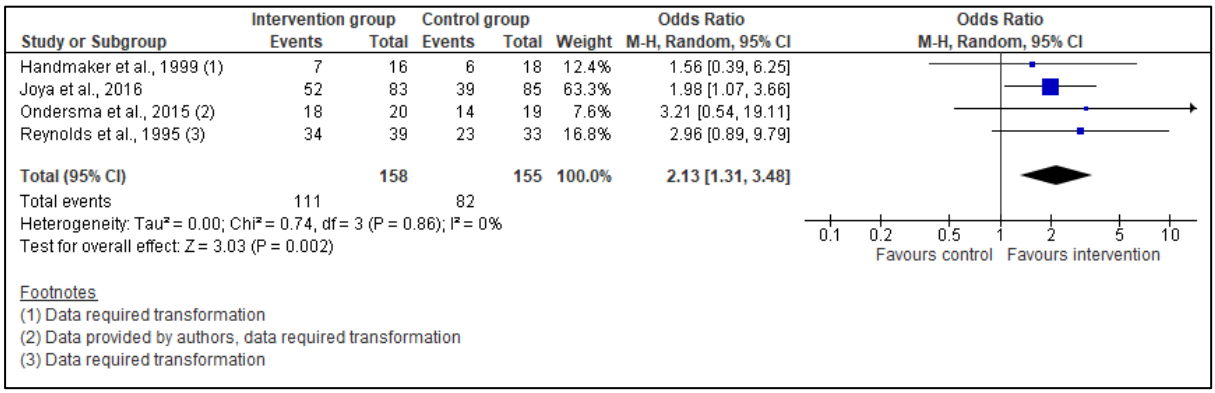
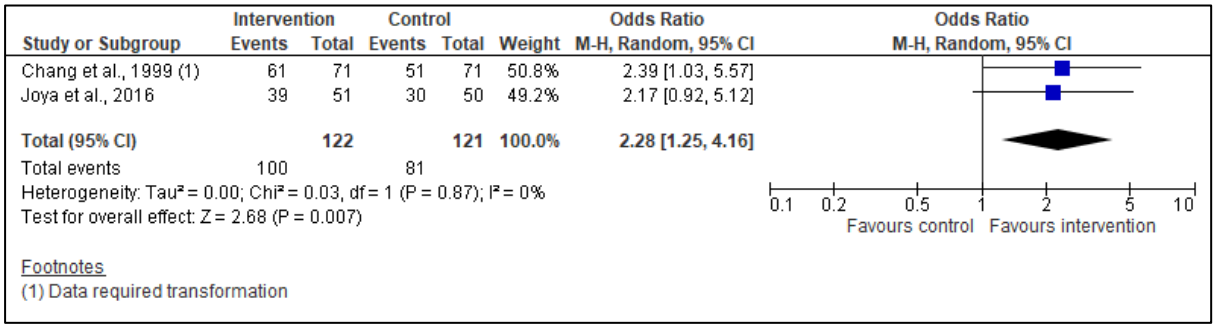


Figure 5

Comparison Table for Abstinence Maintenance Rate Outcomes



Results of Secondary Analyses

Thirteen subgroup analyses were conducted, assessing all categorical secondary data at least once (summary in Appendix S, comparison tables in Appendix T). Subgroup analyses for frequency at follow-up, quantity at follow-up and abstinence maintenance rate outcomes compared two studies. Nine of ten abstinence rate subgroup analyses assessed at least three studies. Few subgroup analyses reported significant effects. All comparisons reported heterogeneity, indicating that no compared groups were distinct and that observed effects were meaningless. Subgroup analyses with significant effects for all comparison groups (subgroup analyses 4.8 and 4.8, Appendix T) otherwise reported almost complete overlap of 95% CIs, further indicating a lack of meaningful effect.

Seventeen meta-regressions were conducted, covering each continuous secondary outcome at least once (summary in Appendix U, regression graphs in Appendix V). Frequency at follow-up, quantity at follow-up, and abstinence maintenance rate outcomes compared data from two studies. Abstinence rate outcomes compared between two to three studies. No meta-regressions reported significant effects. All effects were small with large CIs, suggesting imprecision.

Discussion

Summary of Findings

The current systematic review and meta-analysis assessed the effect of brief interventions on antenatal alcohol consumption. To achieve this, frequency, quantity, and abstinence outcomes were sourced from existing literature. This allowed a comprehensive understanding of the statistical and clinical significance of the effects of brief interventions on antenatal alcohol consumption to be established, with intent to investigate the validity of current recommendations for treatment. The systematic review otherwise aimed to identify

and critically assess study methodologies and biases to inform recommendations and frameworks for standardising research practices and improving the quality of the literature.

Findings of Primary Analyses

It was anticipated that brief interventions would reduce antenatal alcohol consumption for all frequency, quantity, and abstinence measurements. This expectation arose from their common employment of motivational interviewing techniques (Gomez et al., 2020), which have been proven effective with this demographic (Osterman, 2011), and the trend for pregnant women to engage in health positive behaviours (Solomon & Quinn, 2004) such as those prescribed by brief interventions for antenatal alcohol consumption (Rodgers, 2018; WHO, 2014).

Meta-analyses of participant alcohol use frequency and quantity at follow-up produced small, insignificant effects. Insignificance was assessed both by effect *p*-values being above 0.05 and CIs overlapping zero. These outcomes were given very low certainty of evidence ratings, a consequence of study biases, effects being of no magnitude or negative direction, and there being a limited number of studies to assess (Schünemann et al., 2021a). Accordingly, these findings were neither statistically nor clinically significant. Frequency and quantity change outcomes were not assessable due to a lack of data, again highlighting how few studies were available. These findings align with much of the existing literature which has failed to observe significant or meaningful effects of brief intervention on patterns or levels of consumption (Chang et al., 1999; Handmaker et al., 1999; Ondersma et al., 2015; Osterman & Dyehouse, 2012; Osterman et al., 2014; Reynolds et al., 1995; Rubio et al., 2014; Sheehan et al., 2014; Tzilos et al., 2011). Significant effects of brief intervention in reducing levels of total alcohol consumption have, however, been observed in pregnant women (Tzilos Wernette et al., 2018).

Abstinence outcomes indicated that pregnant women provided brief interventions had significantly increased odds of achieving and maintaining abstinence compared to controls. Significance was assessed by the *p*-values being less than 0.05, the CIs not crossing zero, and, for the abstinence rate outcome only, the OR exceeding the pre-set minimum effect size for significance (OR = 1.61). Both findings were rated as being of low evidence certainty due to the high risk of bias and small sample sizes of included studies. These outcomes were otherwise assessed to have high risk of publication bias and imprecision (indicated by large CIs). These statistically significant results are therefore of low clinical significance, as they may not reflect genuine effects. However, it should be noted these findings corroborate previous work that has identified statistically significant effects of brief interventions on improving abstinence rates (O'Connor & Whaley, 2007) and abstinence maintenance rates (Chang et al., 1999) in pregnant women, although, much of the literature has reported insignificant effects on these outcomes (Joya et al., 2016; Ondersma et al., 2015; Reynolds et al., 1995; van der Wulp et al., 2014). Abstinence rate outcomes were otherwise comparable to those of Gomez et al. (2020), which reported a slightly greater OR of 2.31 [1.61, 3.32]. These differences are attributed to the unique data sources assessed by each study. While exclusion criteria were similar for both studies, Gomez et al. (2020) appeared to define brief intervention differently and included studies utilising multi-session interventions (van der Wulp et al., 2014 [health counselling group]), interventions without an interactive component (Crawford-Williams et al., 2016; van der Wulp et al., 2014 [tailored letters group]), and studies with polydrug screening (Yonkers et al., 2012).

Findings of Secondary Analyses

Subgroup analyses and meta-regressions were intended to be conducted on all secondary data types (Appendix A) for all primary outcomes. However, only a fraction of these were viable for analysis due to insufficient data at the primary and secondary outcome

levels. All subgroup analyses reported homogeneity, suggesting that generated groups were not statistically different and therefore incomparable. All meta-regressions reported small effects with very large CIs, indicating a high probability of imprecision. No secondary analyses reached significance or produced meaningful results. These findings aligned with expectations that the limited number of available data sources would prevent meaningful or significant results from being observed (Deeks et al., 2021; Valentine et al., 2010).

However, systematic review identified brief interventions to be more effective on pregnant women with greater levels of baseline use and that studies including them possessed greater power to detect significant effects (Osterman & Dyehouse, 2012; Osterman et al., 2014). It was therefore anticipated that including women with high-level use and dependencies, employing the low-sensitivity AUDIT screen (Chang, 2010), and allowing self-administration of use measurement tools (observed to elicit greater responses for stigmatised behaviours; Freeman et al., 2019; Tzilos et al., 2011), would increase probable effect sizes and power. Including baseline abstainers and having samples of greater baseline gestations, where participants are more likely to have experienced spontaneous cessation (Alshaarawy et al., 2016; Muggli et al., 2016; O'Connor & Whaley, 2007) were therefore expected to report lesser baseline use levels and subsequently smaller, insignificant effects.

Existing research otherwise indicated that effects would be greater in studies that measured use over longer periods, allowing more time for behaviour changes to be adopted (Lally et al., 2009), employed motivational interviewing techniques (Osterman, 2011), recommended personalised goal setting (Chang et al., 2000), and assessed non-White ethnic majorities (Sheehan et al., 2014). Lesser effects were anticipated in studies with older (Sheehan et al., 2014; van der Wulp et al., 2014) and more educated (Chang et al., 2005) samples. However, note that insignificant effects have also been observed for ethnicity

(Chang et al., 2005) and education (van der Wulp et al., 2014). Nulliparity was not anticipated to significantly influence intervention outcomes (Chang et al., 2005).

Expectations were uncertain for pilot status and the use of informal screening and or use measurement tools due to increased risks of bias (Leon et al., 2011; NHMRC, 2019), and of sampling location, marital status, use measurement tool, intervention delivery format and conductor, inclusion of a bonus item, and control type, due to a lack literature assessing these variables.

Implications

This systematic review and meta-analysis failed to validate current recommendations for the use of brief interventions to manage antenatal alcohol consumption. While findings indicated that participants receiving brief interventions had significantly greater odds of achieving and maintaining abstinence compared to controls, they were reliant on a small number of studies with high risk of biases, and were identified to be imprecise, at risk of publication bias, and of low overall evidence certainty. Otherwise, no significant or meaningful findings were observed for frequency, quantity, and secondary outcomes. It is subsequently argued that none of the reported findings are clinically significant and that significant observations were reflections of significant evidence, but not genuine significant effects (Higgins et al., 2021a). Therefore, these results should not be used to validate or justify current or future healthcare policies and further, more standardised, and better-quality research is required should this be achieved.

Limitations

Lack of Viable Data

The most notable limitation to this systematic review and meta-analysis was how few studies were available and viable for analysis. Meta-analyses with fewer studies have lesser power to detect statistically significant effects and achieve clinical significance (Schünemann

et al., 2021b). This was reflected by reports of statistically insignificant frequency and quantity outcomes, and clinically insignificant abstinence outcomes in this study. Additionally, limitations to the number of viable studies prevented two primary meta-analyses from being conducted, and rendered all secondary analyses, to varying degrees, meaningless.

Small Sample Sizes and Low Power

Much of the available literature on the effects of brief interventions for antenatal alcohol consumption is reliant on small sample sizes, which are of lower power to detect significant effects. The ability to achieve sufficient power to detect significant effects is reliant on sample size and the size of potential effects (Röhrig et al., 2010). A study with a small sample size would therefore possess power only to detect large effects. However, small effects of less than one standardised mean difference are commonly reported for frequency and quantity outcomes in pregnant women receiving brief interventions. This is a consequence of spontaneous cessation and low baseline consumption levels (Osterman & Dyehouse, 2012; Osterman et al., 2014), and of assessment reactivity concealing the effects of brief interventions and preventing differences between groups from being observed (Kypri et al., 2007). To detect small effects such as these, Röhrig et al. (2010) suggests that samples of 400 or more per group would be required. This requirement has not been achieved by any of the existing research on brief interventions for antenatal alcohol consumption, suggesting that the literature in general is of low power. Issues around power seem to affect abstinence outcomes to a lesser extent due to their dichotomous measurement, with moderate effects of intervention being reported (Ondersma et al., 2015; Reynolds et al., 1995). Delays in intervention effectiveness have otherwise been observed, where significant differences between groups in AUDIT scores were detected twelve-months post-intervention, but not during the antenatal period (Tsai et al., 2009; Tzilos et al., 2011). It is possible that this delay

reflects the diminishing influence of spontaneous cessation in the postpartum, and of assessment reactivity as time passes. This poses a major challenge to the achievement of clinical significance, as it would suggest that brief interventions that are delivered antenatally only become more effective than assessment alone after the prevention of adverse pregnancy outcomes is no longer possible.

Meta-analyses such as the current paper, which assess primarily underpowered studies, are more likely to report smaller effects (Turner et al., 2013). This likely explains the small effects observed for frequency and quantity outcomes and may have contributed to the abstinence rate OR being lower than that of Gomez et al. (2020).

Small Sample Sizes and Biases

The consequences of small sample sizes extend to study and publication biases, and therefore imprecision. High risk of bias was assessed for all included studies, corroborating existing reports about the quality of the literature on brief interventions for antenatal alcohol consumption (Gomez et al., 2020; Schölin & Fitzgerald, 2019; WHO, 2014). Biased studies are more likely to be affected by errors and issues regarding internal validity (Greco et al., 2013), report inflated effects (Turner et al., 2013), and achieve significance (Dechartres et al., 2013). Meta-analyses that include studies at high risk of bias, such as this one, are consequentially at greater risk of producing invalid results that cannot establish clinical significance (Boutron et al., 2021). In the current study, this was reflected by meaningless effect magnitudes and directions for frequency and quality at follow-up outcomes, and by the large CIs for both abstinence ORs. The literature on brief interventions for antenatal alcohol consumption, in which high risk of biases is prevalent, may also be founded on equally invalid reports of effectiveness.

Publication biases were also observed for significant abstinence outcomes. This too can be attributed to the small sample sizes of the available literature (Turner et al., 2013).

Smaller-sample studies tend to lack sufficient power to detect significant effects, leading to the reporting of false-negative/type II errors (Nayak, 2010; NHMRC, 2019; Röhrig et al., 2010). Studies that report insignificant findings are less likely to be published, especially when sample sizes are small (Turner et al., 2013). This validates reports of publication bias in the current study and suggests the literature on brief interventions for antenatal alcohol consumption more generally, in which sample sizes are typically small, is also prone to publication biases.

However, some of the biases most prevalent to studies of brief interventions for antenatal alcohol consumption can be attributed to inherent limitations in psychological research, such as the reduced ability to establish placebo controls to blind participants and experimenters to allocation (Gomez et al., 2020), and an overreliance on potentially biased self-reporting (Freeman et al., 2019; Tzilos et al., 2011). These unavoidable biases are common in studies of brief interventions for antenatal alcohol consumption, rendering it difficult to conduct trials without high risk of bias when doing so per CHSRI guidelines (Higgins et al., 2021c). These guidelines argue that, should any domain be judged as having high risk of bias, such as the domain related to participant blinding, then so should the overall study or outcome.

Inconsistent Study Methodologies

Considerable diversity in study methodologies and reporting practices was identified through systematic review. Meta-analyses assessing studies with large variability often encounter heterogenous or invalid outcome data (McKenzie & Brennan, 2021). Issues with heterogeneity were not observed by this meta-analysis; however, inconsistent methodologies rendered data extraction difficult and necessitated extensive transformations prior to synthesis. These transformations, while aligned with CHSRI guidelines (Higgins et al., 2021b), introduced potential errors and imprecision due to the inherent subjectivity of

deciding what data needs transformation and how. It was otherwise noted that many studies were missing or failed to comprehensively report certain analyses and outcomes and failed to succinctly report sample sizes at each stage of experimentation (i.e., Chang et al., 1999 did not appear to define analysable sample sizes per group). Many data sources were excluded from analysis for these reasons, despite relevant outcomes being reported to some degree. Issues with older studies employing author-developed, informal screening and use measurement tools were also noted. These are argued to decrease the clinical significance of findings as they are not validated for use in pregnant women who drink alcohol.

Unethical Practices

This systematic review of the literature otherwise identified a major ethical concern regarding the inclusion of women with high-level alcohol use and dependencies, with consequences for the validity of many existing investigations of the effects of brief interventions on antenatal alcohol consumption. The WHO (2014) recommends women who present with high-use or dependencies be provided referrals to treatment after brief intervention. In research, such women would therefore require referrals additional to interventions and so, would need to be excluded to control for variance. However, the effects of brief interventions have been shown to increase with the inclusion of high use and dependent women (Handmaker et al., 1999; Joya et al., 2016; Ondersma et al., 2015; Reynolds et al., 1995; Tzilos et al., 2011; Sheehan et al., 2014), leading some (Osterman & Dyehouse, 2012; Osterman et al., 2014) to include them so significant effects could be observed in low-power studies. This presents a paradox whereby, to detect statistically significant effects and validate current healthcare recommendations, researchers need to ignore those same recommendations. This sacrifices the clinical significance of the research, as it would no longer reflect recommended practice, and would unethically prevent women with higher levels of use from receiving necessary referrals to treatment (WHO, 2014).

Lack of Sensitivity Analyses

Sensitivity analyses were not planned for this meta-analysis, although this would have provided valuable information about the robustness of results (Deeks et al., 2021). This is especially true given the small number of studies included for analyses, and the prevalence of potentially subjective decision making around eligibility criteria, screening, and the extensive transformations, categorisations, and assumptions required to standardise data for analysis (Deeks et al., 2021; Shrier et al., 2008). Subject practices are in part reflected by disparities in outcome magnitudes between this study and Gomez et al. (2020), who employed different eligibility criteria.

Recommendations

To address issues regarding low power, small sample sizes, and small effects, future research should aim to sample 100 participants per group minimum, preferably more, by recruiting from multiple locations. This standard is notably lower than those recommended by Röhrig et al. (2010), but more achievable. Screening and assessments should be reduced in length and depth to minimise the impact of assessment reactivity (Bernstein et al., 2010). Otherwise, more research required to explore the effects of spontaneous cessation and intervention effect delay on brief intervention effectiveness. To do so, studies can measure alcohol consumption prior to pregnancy to assess spontaneous cessation in early gestation and control for this in antenatal use measurements. Reporting into the postnatal period will otherwise provide insight into the power of current research practices to detect significant effects antenatally given effect delays. The inclusion of women with high-level use and dependencies is not a recommended means of increasing effects, due to ethical concerns. It is otherwise argued that doing so reduces the clinical significance of findings, rending the benefits to statistical significance redundant. Effect sizes should instead be managed by increasing sample sizes and addressing biases.

It is recommended that study biases are managed by; reducing the quantity of information provided to participants about study details to help with blinding; reporting all outcomes and analyses; utilising protocols; and employing only validated screen and use measurement tools. Publication biases are likely too pervasive to address at the individual-study level, so authors should focus on reducing methodological biases to improve the certainty of their findings. However, it is noted that existing Cochrane standards for assessing risk of bias (Higgins et al., 2021c) are themselves biased against psychological interventions, albeit due to reasonable recommendations for ensuring study quality. The need for an adapted risk of bias assessment tool specific to systematic reviews and meta-analyses of psychological interventions should therefore be developed and employed alongside existing tools to give indication of relative and standardised biases.

Otherwise, it is recommended that the definition of brief intervention be clarified to prevent mixed understandings across the literature. The current study argues that brief interventions are single-session and less than sixty-minutes, though shorter, thirty-minute-or-less alternatives may be more appropriate for revised definitions, given the short length of specialist clinician visits.

Method and Data Reporting Framework

To aid in the standardisation and risk of bias management of future research, a framework has been developed (Appendix W). This framework was designed to incorporate the beneficial practices identified in the systematic review and disparage impractical ones, and to provide a framework specific to brief interventions for antenatal alcohol consumption and associated features and reporting standards. It also includes a checklist to help authors keep track of whether indicators of bias have been addressed and reported on. This was based on CHSRI bias assessments (Higgins et al., 2021c) despite earlier criticisms, as they are, at current, most optimal.

The framework suggests that frequency, quantity, and abstinence outcomes be assessed at follow-up and as values of change, per the current meta-analysis. While abstinence outcomes were identified to be less affected by small effect sizes, suggesting that they may be a preferable outcome type, assessing all six indicators of use will ensure that comprehensive understandings on the effects of brief interventions on antenatal alcohol consumption can be developed. Flexibility was provided around units of measurement for frequency outcomes. Any unit that could be converted into 28-day month was considered viable, with measurement by episodes explicitly discouraged. This will ensure that studies can be easily synthesised, without requiring they be identical, thereby reducing reliance on potentially biased transformations in future meta-analyses. Personalised goal setting is the preferred advice to provide participants, as this is more effective than advising abstinence (Chang et al., 2005).

Conclusion

This systematic review and meta-analysis assessed the effectiveness of brief interventions for reducing antenatal alcohol consumption for frequency, quantity, and abstinence outcomes. No results were deemed clinically significant, despite abstinence rate and abstinence maintenance rate outcomes being statistically significant, due to the limited number of studies included for analyses, publication biases, and the imprecision around observed effects. Therefore, the findings of this report did not validate existing healthcare recommendations, nor can they inform future policy. Major flaws in the literature were identified regarding the inability for studies to achieve statistical significance without sacrificing clinical significance, due to interference from prevailing patterns of antenatal alcohol consumption. A need for further research that is standardised and of less risk of bias was identified, with recommendations and a framework for achieving this provided.

References

Alcohol and Drug Foundation. (2017). What is a standard drink?. https://adf.org.au/insights/what-is-a-standard-drink/

Ali, R., Awwad, E., Babor, T., Bradley, F., Butau, T., Farrell, M., Formigoni, M. L. O. S., Isralowitz, R., Boerngen de Lacerda, R., Marsden, J., McRee, B., Monteiro, M., Pal, H., Rubio-Stipec, M., & Vendetti, J. (2002). The alcohol, smoking and substance involvement screening test (ASSIST): Development, reliability and feasibility. *Addiction, 97*(9), 1183-1194. https://doi.org/10.1046/j.1360-0443.2002.00185.x

Alshaarawy, O., Breslau, N., & Anthony, J. C. (2016). Monthly estimates of alcohol drinking during pregnancy: United States, 2002–2011. *Journal of Studies on Alcohol and Drugs, 77*(2), 272-276. https://dx.doi.org/10.15288%2Fjsad.2016.77.272

Andersen, A. M. N., Andersen, P. K., Olsen, J., Grønbæk, M., & Strandberg-Larsen, K. (2012). Moderate alcohol intake during pregnancy and risk of fetal death. *International Journal of Epidemiology, 41*(2), 405-413. https://doi.org/10.1093/ije/dyr189

Anderson, A. E., Hure, A. J., Forder, P., Powers, J. R., Kay-Lambkin, F. J., & Loxton, D. J. (2013). Predictors of antenatal alcohol use among Australian women: a prospective cohort study. *BJOG: An International Journal of Obstetrics & Gynaecology, 120*(11), 1366-1374. https://doi.org/10.1111/1471-0528.12356

Andersson, E., & Elliot, E. (2018). Economic costs of fetal alcohol specturm disorder (FASD). *Journal of Paediatrics and Child Health, 52*(S2), 7. https://doi.org/10.1111/jpc.13946

Arendt, J. (2007). How do psychology researchers find studies to include in meta-analyses?. *Behavioral & Social Sciences Librarian, 26*(1), 1-23. https://doi.org/10.1300/J103v26n01_01

Australian Government Department of Health. (2004). *Working With Young People on AOD Issues: Facilitator's Guide.* Commonwealth of Australia. https://www1.health.gov.au/internet/publications/publishing.nsf/Content/drugtreat-pubs-front9-fa-toc

Australian Institute of Health and Welfare. (2014). *National Drug Strategy Household Survey detailed report: 2013* (Cat. No. PHE 183). https://www.aihw.gov.au/reports/illicit-use-of-drugs/2013-ndshs-detailed/summary

Australian Institute of Health and Welfare. (2020). *Australia's Children* (Cat. No. CWS 69). https://www.aihw.gov.au/reports/children-youth/australias-children/contents/health/smoking-and-drinking-in-pregnancy

Australian Institute of Health and Welfare. (2021). *Alcohol, Tobacco & Other Drugs in Australia* (Cat. No. PHE 221). https://www.aihw.gov.au/reports/alcohol/alcohol-tobacco-other-drugs-australia

Babor, T. F., Higgins-Biddle, J. C., Saunders, J. B., & Monteiro, M. G. (2001). *AUDIT: The Alcohol Use Disorders Identification Test*. World Health Organisation. https://www.who.int/publications/i/item/audit-the-alcohol-use-disorders-identification-test-guidelines-for-use-in-primary-health-care

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. https://doi.org/10.1111/j.1465-3362.2010.00174.x

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons. https://doi.org/10.1002/9780470743386

Boutron, I., Page, M. J., Higgins, J. P. T., Altman, D. G., Lundh, A., & Hróbjartsson, A. (2021). Chapter 7: Considering bias and conflicts of interest among the included

studies. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (version 6.2). Cochrane. www.training.cochrane.org/handbook

Bower, C., Silva, D., Henderson, T. R., Ryan, A., & Rudy, E. (2000). Ascertainment of birth defects: the effect on completeness of adding a new source of data. *Journal of Paediatrics and Child Health, 36*(6), 574–576. https://doi.org/10.1046/j.1440-1754.2000.00575.x

Bramer, W. M., Rethlefsen, M. L., Kleijnen, J., & Franco, O. H. (2017). Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Systematic Reviews, 6*(1), 1-12. https://doi.org/10.1186/s13643-017-0644-y

Burd, L., Blair, J., & Dropps, K. (2012). Prenatal alcohol exposure, blood alcohol concentrations and alcohol elimination rates for the mother, fetus and newborn. *Journal of Perinatology, 32*(9), 652-659. https://doi.org/10.1038/jp.2012.57

Butt, P., Beirness, D., Gliksman, L., Paradis, C., & Stockwell, T. (2011). *Alcohol and health in Canada: a summary of evidence and guidelines for low-risk drinking*. Canadian Centre on Substance Abuse. https://prod.ccsa.ca/sites/default/files/2019-04/2011-Summary-of-Evidence-and-Guidelines-for-Low-Risk%20Drinking-en.pdf

Carson, G., Cox, L. V., Crane, J., Croteau, P., Graves, L., Kluka, S., ... & Wood, R. (2017). No. 245-Alcohol use and pregnancy consensus clinical guidelines. *Journal of Obstetrics and Gynaecology Canada, 39*(9), e220–e254. https://doi.org/10.1016/j.jogc.2017.06.005

Cederbaum, A. I. (2012). Alcohol metabolism. *Clinics in Liver Disease, 16*(4), 667-685. https://doi.org/10.1016/j.cld.2012.08.002

Centers for Disease Control and Prevention. (2020). What is stillbirth?. https://www.cdc.gov/ncbddd/stillbirth/facts.html

Centers for Disease Control and Prevention. (2021a). Fetal alcohol spectrum disorders (FASDs). https://www.cdc.gov/ncbddd/fasd/facts.html

Centers for Disease Control and Prevention. (2021b). Prevalence of FASDs. https://www.cdc.gov/ncbddd/fasd/data.html

Chang, G. (2001). Alcohol-screening instruments for pregnant women. *Alcohol Research & Health, 25*(3), 204-209. https://pubmed.ncbi.nlm.nih.gov/11810959/

Chang, G., Fisher, N. D., Hornstein, M. D., Jones, J. A., & Orav, E. J. (2010). Identification of risk drinking women: T-ACE screening tool or the medical record. *Journal of Women's Health, 19*(10), 1933-1939. https://dx.doi.org/10.1089%2Fjwh.2009.1911

Chang, G., Goetz, M. A., Wilkins-Haug, L., & Berman, S. (2000). A brief intervention for prenatal alcohol use: An in-depth look. *Journal of Substance Abuse Treatment, 18*(4), 365-369. http://dx.doi.org/10.1016/S0740-5472(99)00105-1

Chang, G., McNamara, T. K., Orav, E. J., Koby, D., Lavigne, A., Ludman, B., Vincitorio, N. A., & Wilkins-Haug, L. (2005). Brief intervention for prenatal alcohol use: A randomized trial. *Obstetrics and Gynecology, 105*(5), 991–998. https://doi.org/10.1097/01.AOG.0000157109.05453.84

Chang, G., Wilkins-Haug, L., Berman, S., & Goetz, M. A. (1999). Brief intervention for alcohol use in pregnancy: A randomized trial. *Addiction, 94*(10), 1499-1508. DOI: 10.1046/j.1360-0443.1999.941014996.x

Chiodo, L. M., Cosmian, C., Pereira, K., Kent, N., Sokol, R. J., & Hannigan, J. H. (2019). Prenatal alcohol screening during pregnancy by midwives and nurses. *Alcoholism: Clinical and Experimental Research, 43*(8), 1747-1758. https://doi.org/10.1111/acer.14114

Chiodo, L. M., Delaney-Black, V., Sokol, R. J., Janisse, J., Pardo, Y., & Hannigan, J. H. (2014). Increased cut-point of the TACER-3 screen reduces false positives without

losing sensitivity in predicting risk alcohol drinking in pregnancy. *Alcoholism, Clinical and Experimental Research, 38*(5), 1401–1408. https://doi.org/10.1111/acer.12368

Covidence Systematic Review Software [Computer program]. Veritas Health Innovation, n.d. http://www.covidence.org/

Crawford-Williams, F., Fielder, A., Mikocka-Walus, A., Esterman, A., & Steen, M. (2016). A public health intervention to change knowledge, attitudes and behaviour regarding alcohol consumption in pregnancy. *Evidence Based Midwifery, 14*(1), 4-10.

Dechartres, A., Trinquart, L., Boutron, I., & Ravaud, P. (2013). Influence of trial sample size on treatment effect estimates: Meta-epidemiological study. *BMJ, 346*, f2304. https://doi.org/10.1136/bmj.f2304

Deeks, J. J., & Higgins, J. P. (2010). Statistical algorithms in review manager 5. *Statistical Methods Group of The Cochrane Collaboration, 1*(11). https://training.cochrane.org/handbook/current/chapter-10-statistical-algorithms-revman-5-1

Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2021). Chapter 10: Analysing data and undertaking meta-analyses. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (version 6.2). Cochrane. www.training.cochrane.org/handbook

DeJong, K., Olyaei, A., & Lo, J. O. (2019). Alcohol use in pregnancy. *Clinical Obstetrics and Gynecology, 62*(1), 142-155. https://dx.doi.org/10.1097%2FGRF.0000000000000414

DiClemente, C. C., Corno, C. M., Graydon, M. M., Wiprovnick, A. E., & Knoblach, D. J. (2017). Motivational interviewing, enhancement, and brief interventions over the last

decade: A review of reviews of efficacy and effectiveness. *Psychology of Addictive Behaviors, 31*(8), 862-887. https://doi.org/10.1037/adb0000318

Dum, M., Sobell, L. C., Sobell, M. B., Heinecke, N., Voluse, A., & Johnson, K. (2009). A quick drinking screen for identifying women at risk for an alcohol-exposed pregnancy. *Addictive Behaviors, 34*(9), 714-716. https://doi.org/10.1016/j.addbeh.2009.04.001

Easton, B., Burd, L., Rehm, J., & Popova, S. (2016). Productivity losses associated with fetal alcohol spectrum disorder in New Zealand. *The New Zealand Medical Journal, 129*(1440), 72-83. https://www.otago.ac.nz/wellington/otago620018.pdf

Epstein, E. E., Drapkin, M. L., Yusko, D. A., Cook, S. M., McCrady, B. S., & Jensen, N. K. (2005). Is alcohol assessment therapeutic? Pretreatment change in drinking among alcohol-dependent women. *Journal of Studies on Alcohol, 66*(3), 369–378. https://doi.org/10.15288/jsa.2005.66.369

Fergie, L., Campbell, K. A., Coleman-Haynes, T., Ussher, M., Cooper, S., & Coleman, T. (2019). Identifying effective behavior change techniques for alcohol and illicit substance use during pregnancy: A systematic review. *Annals of Behavioral Medicine, 53*(8), 769-781. https://doi.org/10.1093/abm/kay085

Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods, 17*(1), 120–128. https://doi.org/10.1037/a0024445

Finding_SDs [Calculator]. Drahota, A., & Beller, E., n.d. https://training.cochrane.org/resource/revman-calculator

Flak, A. L., Su, S., Bertrand, J., Denny, C. H., Kesmodel, U. S., & Cogswell, M. E. (2014). The association of mild, moderate, and binge prenatal alcohol exposure and child

neuropsychological outcomes: A meta-analysis. *Alcoholism: Clinical and Experimental Research, 38*(1), 214-226. https://doi.org/10.1111/acer.12214

Freeman, J., Condon, C., Hamilton, S., Mutch, R. C., Bower, C., & Watkins, R. E. (2019). Challenges in accurately assessing prenatal alcohol exposure in a study of fetal alcohol spectrum disorder in a youth detention center. *Alcoholism: Clinical and Experimental Research, 43*(2), 309-316. https://doi.org/10.1111/acer.13926

Giglia, R. C., & Binns, C. W. (2007). Patterns of alcohol intake of pregnant and lactating women in Perth, Australia. *Drug and Alcohol Review, 26*(5), 493–500. https://doi.org/10.1007/s10995-017-2318-8

Gilbert, W. M. (2006). Amniotic fluid dynamics. *NeoReviews, 7*(6), e292-e299. https://doi.org/10.1542/neo.7-6-e292

Glass, J. E., Andréasson, S., Bradley, K. A., Finn, S. W., Williams, E. C., Bakshi, A. S., ... & Saitz, R. (2017). Rethinking alcohol interventions in health care: a thematic meeting of the International Network on Brief Interventions for Alcohol & Other Drugs (INEBRIA). *Addiction Science & Clinical Practice, 12*(1), 14. https://doi.org/10.1186/s13722-017-0079-8

Glass, J. E., Hamilton, A. M., Powell, B. J., Perron, B. E., Brown, R. T., & Ilgen, M. A. (2015). Specialty substance use disorder services following brief alcohol intervention: a meta-analysis of randomized controlled trials. *Addiction, 110*(9), 1404–1415. https://doi.org/10.1111/add.12950

Gomez, K. U., Goodwin, L., Jackson, L., Jones, A., Chisholm, A., & Rose, A. K. (2020). Are psychosocial interventions effective in reducing alcohol consumption during pregnancy and motherhood? A systematic review and meta-analysis. *Addiction, 116*(7), 1638-1663. https://doi.org/10.1111/add.15296

Greco, T., Zangrillo, A., Biondi-Zoccai, G., & Landoni, G. (2013). Meta-analysis: Pitfalls and hints. *Heart, Lung and Vessels, 5*(4), 219-225. https://dx.doi.org/10.1371%2Fjournal.pone.0059202

Greenmyer, J. R., Klug, M. G., Kambeitz, C., Popova, S., & Burd, L. (2018). A multicountry updated assessment of the economic impact of fetal alcohol spectrum disorder: Costs for children and adults. *Journal of Addiction Medicine, 12*(6), 466–473. https://doi.org/10.1097/ADM.0000000000000438

Haber, P., Lintzeris, N., Proude, E., & Lopatko, O. (2009). *Guidelines for the treatment of alcohol problems* (Cat. No. P3 -5625). https://www.health.gov.au/resources/publications/guidelines-for-the-treatment-of-alcohol-problems

Handmaker, N. S., Miller, W. R., & Manicke, M. (1999). Findings of a pilot study of motivational interviewing with pregnant drinkers. *Journal of Studies on Alcohol, 60*(2), 285-287. https://doi.org/10.15288/jsa.1999.60.285

Harris, K. R., & Bucens, I. K. (2003). Prevalence of fetal alcohol syndrome in the Top End of the Northern Territory. *Journal of Paediatrics and Child Health, 39*(7), 528–533. https://doi.org/10.1046/j.1440-1754.2003.00208.x

Harrison, P. A., & Sidebottom, A. C. (2009). Alcohol and drug use before and during pregnancy: An examination of use patterns and predictors of cessation. *Maternal and Child Health Journal, 13*(3), 386-394. https://doi.org/10.1007/s10995-008-0355-z

Higgins, J. P. T., Lasserson, T., Chandler, J., Tovey, D., Thomas, J., Flemyng, E., & Churchill, R. (2021a). *Methodological Expectations of Cochrane Intervention Reviews (MECIR)* (version February 2021). Cochrane. https://community.cochrane.org/mecir-manual

Higgins, J. P. T., Li, T., & Deeks, J. J. (2021b). Chapter 6: Choosing effect measures and computing estimates of effect. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (version 6.2). Cochrane. www.training.cochrane.org/handbook

Higgins, J. P. T., Savović, J., Page, M. J., Elbers, R. G., & Sterne, J. A. C. (2021c). Chapter 8: Assessing risk of bias in a randomized trial. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (version 6.2). Cochrane. www.training.cochrane.org/handbook

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2021d). *Cochrane Handbook for Systematic Reviews of Interventions* (version 6.2). Cochrane. www.training.cochrane.org/handbook

Hotham, E., Ali, R., White, J., Sullivan, T., & Robinson, J. (2013). Investigation of the alcohol, smoking, and substance involvement screening test (the ASSIST) version 3.0 in pregnancy. *Addictive Disorders & Their Treatment, 12*(3), 123-135. https://doi.org/10.1097/ADT.0b013e3182636904.

Ikehara, S., Kimura, T., Kakigano, A., Sato, T., Iso, H., Japan Environment Children's Study Group, ... & Katoh, T. (2019). Association between maternal alcohol consumption during pregnancy and risk of preterm delivery: the Japan Environment and Children's Study. *BJOG: An International Journal of Obstetrics & Gynaecology, 126*(12), 1448-1454. https://doi.org/10.1111/1471-0528.15899

Jaddoe, V. W. V., Bakker, R., Hofman, A., Mackenbach, J. P., Moll, H. A., Steegers, E. A., & Witterman, J. C. M. (2007). Moderate alcohol consumption during pregnancy and

the risk of low birth weight and preterm birth. The generation R study. *Annals of Epidemiology, 17*(10), 834-840. https://doi.org/10.1016/j.annepidem.2007.04.001

Jonas, D. E., Garbutt, J. C., Amick, H. R., Brown, J. M., Brownley, K. A., Council, C. L., Viera, A. J., Wilkins, T. M., Schwartz, C. J., Richmond, E. M., Yeatts, J., Evans, T. S., Wood, S. D., & Harris, R. P. (2012). Behavioral counseling after screening for alcohol misuse in primary care: A systematic review and meta-analysis for the U.S. Preventive Services Task Force. *Annals of Internal Medicine, 157*(9), 645–654. https://doi.org/10.7326/0003-4819-157-9-201211060-00544

Jones, K., & Smith, D. (1973). Recognition of the fetal alcohol syndrome in early infancy. *The Lancet, 302*(7836), 999-1001. https://doi.org/10.1016/s0140-6736(73)91092-1

Kesmodel, U., Wisborg, K., Olsen, S. F., Henriksen, T. B., & Secher, N. J. (2002). Moderate alcohol intake during pregnancy and the risk of stillbirth and death in the first year of life. *American Journal of Epidemiology, 155*(4), 305–312. https://doi.org/10.1093/aje/155.4.305

Kypri, K., Langley, J. D., Saunders, J. B., & Cashell-Smith, M. L. (2007). Assessment may conceal therapeutic benefit: findings from a randomized controlled trial for hazardous drinking. *Addiction, 102*(1), 62–70. https://doi.org/10.1111/j.1360-0443.2006.01632.x

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Lally, P., Van Jaarsveld, C. H., Potts, H. W., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology, 40*(6), 998-1009. https://doi.org/10.1002/ejsp.674

Lefebvre, C., Glanville, J., Briscoem, S., Littlewood, A., Marshall, C., Metzendorf, M., Noel-Storr, A., Rader, T., Shokraneh, F., Thomas, J., & Wieland, S. (2021). Chapter 4:

Searching for and selecting studies. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (version 6.2). Cochrane. www.training.cochrane.org/handbook

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research, 45*(5), 626-629. https://dx.doi.org/10.1016%2Fj.jpsychires.2010.10.008

Mamluk, L., Edwards, H. B., Savović, J., Leach, V., Jones, T., Moore, T. H., ... & Zuccolo, L. (2017). Low alcohol consumption and pregnancy and childhood outcomes: time to change guidelines indicating apparently 'safe'levels of alcohol during pregnancy? A systematic review and meta-analyses. *BMJ Open, 7*(7), e015410. https://dx.doi.org/10.1136%2Fbmjopen-2016-015410

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*(4), 719-748. https://doi.org/10.1093/jnci/22.4.719

May, P. A., Blankenship, J., Marais, A. S., Gossage, J. P., Kalberg, W. O., Joubert, B., Cloete, M., Barnard, R., De Vries, M., Hasken, J., Robinson, L. K., Adnams, C. M., Buckley, D., Manning, M., Parry, C. D., Hoyme, H. E., Tabachnick, B., & Seedat, S. (2013). Maternal alcohol consumption producing fetal alcohol spectrum disorders (FASD): Quantity, frequency, and timing of drinking. *Drug and Alcohol Dependence, 133*(2), 502–512. https://doi.org/10.1016/j.drugalcdep.2013.07.013

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276-282. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/

McKenzie, J. E., & Brennan, S. E. (2021). Chapter 12: Synthesizing and presenting findings using other methods. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li,

M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (version 6.2). Cochrane. www.training.cochrane.org/handbook

Miller, W. R., & Marlatt, G. A. (1987a). *The Brief Drinker Profile*. Psychological Assessment Resources.

Miller, W. R., & Marlatt, G. A. (1987b). *The Follow-up Drinker Profile*. Psychological Assessment Resources.

Moore, E. M., & Riley, E. P. (2015). What happens when children with fetal alcohol spectrum disorders become adults?. *Current Developmental Disorders Reports, 2*(3), 219-227. https://dx.doi.org/10.1007%2Fs40474-015-0053-7

Muggli, E., O'Leary, C., Donath, S., Orsini, F., Forster, D., Anderson, P. J., ... & Halliday, J. (2016). "Did you ever drink more?" A detailed description of pregnant women's drinking patterns. *BMC Public Health, 16*(1), 1-13. https://doi.org/10.1186/s12889-016-3354-9

National Health and Medical Research Council. (2019). *Guidelines for Guidelines: Assessing Risk of Bias*. NHMRC Publications. https://nhmrc.gov.au/guidelinesforguidelines/develop/assessing-risk-bias

National Health and Medical Research Council. (2020). *Australian Guidelines to Reduce Health Risks From Drinking Alcohol* (Cat. No. DS 14). https://www.nhmrc.gov.au/about-us/publications/australian-guidelines-reduce-health-risks-drinking-alcohol

National Indigenous Drug & Alcohol Committee. (2012). *Addressing Fetal Alcohol Spectrum Disorder in Australia*. Australian National Council on Drugs. https://apo.org.au/node/31035

National Institute on Drug Abuse. (2018). *Principles of Drug Addiction Treatment: A Research-Based Guide* (3rd Edition, NIH Pub. No. 09-4180). National Institute on

Drug Abuse. https://www.drugabuse.gov/publications/principles-drug-addiction-treatment-research-based-guide-third-edition/evidence-based-approaches-to-drug-addiction-treatment/behavioral-therapies/motivational-enhancement-therapy

Nayak, B. K. (2010). Understanding the relevance of sample size calculation. *Indian Journal of Ophthalmology, 58*(6), 469-470. https://dx.doi.org/10.4103%2F0301-4738.71673

Ockene, J. K., Ma, Y., Zapka, J. G., Pbert, L. A., Goins, K. V., & Stoddard, A. M. (2002). Spontaneous cessation of smoking and alcohol use among low-income pregnant women. *American Journal of Preventive Medicine, 23*(3), 150-159. https://doi.org/10.1016/S0749-3797(02)00492-0

O'Brien, P. L. (2014). Performance measurement: A proposal to increase use of SBIRT and decrease alcohol consumption during pregnancy. *Maternal and Child Health Journal, 18*(1), 1-9. https://dx.doi.org/10.1007%2Fs10995-013-1257-2

O'Connell, M. E., Boat, T., Warner, K. E. (2009). *Preventing mental, emotional, and behavioral disorders among young people: progress and possibilities*. National Academies Press. https://www.ncbi.nlm.nih.gov/books/NBK32775/

O'Connor, M. J., & Whaley, S. E. (2007). Brief intervention for alcohol use by pregnant women. *American Journal of Public Health, 97*(2), 252-258. https://dx.doi.org/10.2105%2FAJPH.2005.077222

O'Leary, C. M., Nassar, N., Kurinczuk, J. J., & Bower, C. (2009). The effect of maternal alcohol consumption on fetal growth and preterm birth. *BJOG: An International Journal of Obstetrics & Gynaecology, 116*(3), 390-400. https://doi.org/10.1111/j.1471-0528.2008.02058.x

Olmstead, T. A., Yonkers, K. A., Ondersma, S. J., Forray, A., Gilstad-Hayden, K., & Martino, S. (2019). Cost-effectiveness of electronic-and clinician-delivered screening,

brief intervention and referral to treatment for women in reproductive health centers. *Addiction, 114*(9), 1659-1669. https://dx.doi.org/10.1111%2Fadd.14668

O'Malley, K. D., & Nanson, J. O. (2002). Clinical implications of a link between fetal alcohol spectrum disorder and attention-deficit hyperactivity disorder. *The Canadian Journal of Psychiatry, 47*(4), 349-354. https://doi.org/10.1177%2F070674370204700405

Ondersma, S. J., Beatty, J. R., Svikis, D. S., Strickler, R. C., Tzilos, G. K., Chang, G., Divine, G. W., Taylor, A. R., & Sokol, R. J. (2015). Computer-delivered screening and brief intervention for alcohol use in pregnancy: A pilot randomized trial. *Alcoholism, Clinical and Experimental Research, 39*(7), 1219–1226. https://doi.org/10.1111/acer.12747

Ordean, A., Forte, M., Selby, P., & Grennell, E. (2020). Screening, brief intervention, and referral to treatment for prenatal alcohol use and cigarette smoking: A survey of academic and community health care providers. *Journal of Addiction Medicine, 14*(4), e76-82. https://dx.doi.org/10.1097%2FADM.0000000000000588

Osterman, R. (2011). Feasibility of using motivational interviewing to decrease alcohol consumption during pregnancy. *Journal of Addictions Nursing, 22*(3), 93-102. https://doi.org/10.3109/10884602.2011.585723

Osterman, R. L., Carle, A. C., Ammerman, R. T., & Gates, D. (2014). Single-session motivational intervention to decrease alcohol use during pregnancy. *Journal of Substance Abuse Treatment, 47*(1), 10-19. https://doi.org/10.1016/j.jsat.2014.01.009

Osterman, R. L., & Dyehouse, J. (2012). Effects of a motivational interviewing intervention to decrease prenatal alcohol use. *Western Journal of Nursing Research, 34*(4), 434-454. https://doi.org/10.1177/0193945911402523

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J.,

Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ, 71*(372). https://doi.org/10.1136/bmj.n71

Patra, J., Bakker, R., Irving, H., Jaddoe, V. W., Malini, S., & Rehm, J. (2011). Dose-response relationship between alcohol consumption before and during pregnancy and the risks of low birthweight, preterm birth and small for gestational age (SGA)-a systematic review and meta-analyses. *BJOG: An International Journal of Obstetrics and Gynaecology, 118*(12), 1411–1421. https://doi.org/10.1111/j.1471-0528.2011.03050.x

Pirie, P. L., Lando, H., Curry, S. J., McBride, C. M., & Grothaus, L. C. (2000). Tobacco, alcohol, and caffeine use and cessation in early pregnancy. *American Journal of Preventive Medicine, 18*(1), 54-61. https://doi.org/10.1016/S0749-3797(99)00088-4

Popova, S., Lange, S., Probst, C., Gmel, G., & Rehm, J. (2017). Estimation of national, regional, and global prevalence of alcohol use during pregnancy and fetal alcohol syndrome: a systematic review and meta-analysis. *The Lancet Global Health, 5*(3), e290-e299. https://doi.org/10.1016/S2214-109X(17)30021-9

Quantify agreement with kappa [Calculator]. GraphPad Software, 2014. https://www.graphpad.com/quickcalcs/kappa1/?K=3

Rangmar, J., Sandberg, A. D., Aronson, M., & Fahlke, C. (2015). Cognitive and executive functions, social cognition and sense of coherence in adults with fetal alcohol syndrome. *Nordic Journal of Psychiatry, 69*(6), 1754-1760. https://doi.org/10.3109/08039488.2015.1009487

Rasmussen, C., Andrew, G., Zwaigenbaum, L., & Tough, S. (2008). Neurobehavioural outcomes of children with fetal alcohol spectrum disorders: A Canadian perspective. *Paediatrics & Child Health, 13*(3), 185-191. https://doi.org/10.1093/pch/13.3.185

Reeves, B. C., Deeks, J. J., Higgins, J. P. T., Shea, B., Tugwell, P., & Wells, G. A. (2021). Chapter 24: Including non-randomized studies on intervention effects. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (version 6.2). Cochrane. www.training.cochrane.org/handbook

Review Manager (RevMan) [Computer program]. Version 5.4.1. The Cochrane Collaboration, 2020. https://training.cochrane.org/online-learning/core-software-cochrane-reviews/revman

Reynolds, K. D., Coombs, D. W., Lowe, J. B., Peterson, P. L., & Gayoso, E. (1995). Evaluation of a self-help program to reduce alcohol consumption among pregnant women. *International Journal of the Addictions, 30*(4), 427-443. https://doi.org/10.3109/10826089509048735

Riley, E. P., Infante, M. A., & Warren, K. R. (2011). Fetal alcohol spectrum disorders: an overview. *Neuropsychology Review, 21*(2), 73. https://doi.org/10.1007/s11065-011-9166-x

Roberts, S. C., Wilsnack, S. C., Foster, D. G., & Delucchi, K. L. (2014). Alcohol use before and during unwanted pregnancy. *Alcoholism, Clinical and Experimental Research, 38*(11), 2844–2852. https://doi.org/10.1111/acer.12544

Rodgers, C. (2018). Brief interventions for alcohol and other drug use. *Australian Prescriber, 41*(4), 117-121. https://doi.org/10.18773/austprescr.2018.031

Röhrig, B., du Prel, J. B., Wachtlin, D., Kwiecien, R., & Blettner, M. (2010). Sample size calculation in clinical trials: Part 13 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International, 107*(31-32), 552-556. https://dx.doi.org/10.3238%2Farztebl.2010.0552

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638-64. https://dx.doi.org/10.1037/0033-2909.86.3.638

Rubio, D. M., Day, N. L., Conigliaro, J., Hanusa, B. H., Larkby, C., McNeil, M., Cohen, E., Jones, B., Watt-Morse, M., Gilmour, C., Lancet, M., & Kraemer, K. L. (2014). Brief motivational enhancement intervention to prevent or reduce postpartum alcohol use: A single-blinded, randomized controlled effectiveness trial. *Journal of Substance Abuse Treatment, 46*(3), 382–389. https://doi.org/10.1016/j.jsat.2013.10.009

Russell, M. (1994). New assessment tools for risk drinking during pregnancy: T-ACE, TWEAK, and others. *Alcohol Health and Research World, 18*(1), 55-61. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6876474/

Sampson, M., Barrowman, N. J., Moher, D., Klassen, T. P., Pham, B., Platt, R., St John, P. D., Viola, R., & Raina, P. (2003). Should meta-analysts search Embase in addition to Medline?. *Journal of Clinical Epidemiology, 56*(10), 943–955. https://doi.org/10.1016/s0895-4356(03)00110-0

Schölin, L. (2016). *Prevention of harm caused by alcohol exposure in pregnancy. Rapid review and case studies from Member States*. World Health Organisation. https://www.euro.who.int/en/health-topics/disease-prevention/alcohol-use/publications/2016/prevention-of-harm-caused-by-alcohol-exposure-in-pregnancy.-rapid-review-and-case-studies-from-member-states-2016

Schölin, L., & Fitzgerald, N. (2019). The conversation matters: a qualitative study exploring the implementation of alcohol screening and brief interventions in antenatal care in Scotland. *BMC Pregnancy and Childbirth, 19*(1), 1-11. https://doi.org/10.1186/s12884-019-2431-3

Schünemann, H. J., Higgins, J. P. T., Vist, G. E., Glaszious, P., Akl, E. A., Skoetz, N., & Gouyatt, G. H. (2021a). Chapter 14: Completing 'Summary of findings' tables and

grading the certainty of the evidence. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (version 6.2). Cochrane. www.training.cochrane.org/handbook

Schünemann, H. J., Vist, G. E., Higgins, J. P. T., Santesso, N., Deeks, J. J., Glasziou, P., Akl, E. A., & Guyatt, G. H. (2021b). Chapter 15: Interpreting results and drawing conclusions. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (version 6.2). Cochrane. www.training.cochrane.org/handbook

Sheehan, J., Gill, A., & Kelly, B. D. (2014). The effectiveness of a brief intervention to reduce alcohol consumption in pregnancy: A controlled trial. *Irish Journal of Psychological Medicine, 31*(3), 175-189. https://doi.org/10.1017/ipm.2014.28

Shrier, I., Boivin, J. F., Platt, R. W., Steele, R. J., Brophy, J. M., Carnevale, F., ... & Rossignol, M. (2008). The interpretation of systematic reviews with meta-analyses: An objective or subjective process?. *BMC Medical Informatics and Decision Making, 8*(1), 1-8. https://doi.org/10.1186/1472-6947-8-19

Sobell, L. C., Agrawal, S., Sobell, M. B., Leo, G. I., Young, L. J., Cunningham, J. A., & Simco, E. R. (2003). Comparison of a quick drinking screen with the timeline followback for individuals with alcohol problems. *Journal of Studies on Alcohol, 64*(6), 858–861. https://doi.org/10.15288/jsa.2003.64.858

Sobell L. C., & Sobell M. B. (1992). Timeline follow-back. In R. Z. Litten, & J. P. Allen (Eds.), *Measuring alcohol consumption* (pp. 41-72). Humana Press. https://doi.org/10.1007/978-1-4612-0357-5_3

Sokol, R. J., Martier, S. S., & Ager, J. W. (1989). The T-ACE questions: practical prenatal detection of risk-drinking. *American Journal of Obstetrics and Gynecology, 160*(4), 863-870. https://doi.org/10.1016/0002-9378(89)90302-5

Solomon, L. J., & Quinn, V. P. (2004). Spontaneous quitting: Self-initiated smoking cessation in early pregnancy. *Nicotine & Tobacco Research, 6*(Suppl_2), S203-S216. https://doi.org/10.1080/14622200410001669132

Stanesby, O., Cook, M., & Callinan, S. (2018). *Examining trends in alcohol consumption during pregnancy in Australia, 2001-2016*. Centre for Alcohol Policy Research. https://fare.org.au/examining-trends-in-alcohol-consumption-during-pregnancy-in-australia-2001-2016/

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., ... & Higgins, J. P. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ, 343*(7818), 302-307. https://doi.org/10.1136/bmj.d4002

Stoler, J. M., & Holmes, L. B. (2004). Recognition of facial features of fetal alcohol syndrome in the newborn. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics, 127C*(1), 21–27. https://doi.org/10.1002/ajmg.c.30012

Substance Abuse and Mental Health Services Administration. (2011). *White Paper on Screening, Brief Intervention and Referral to Treatment.* https://www.samhsa.gov/sites/default/files/sbirtwhitepaper_0.pdf

Sundermann, A. C., Edwards, D. R. V., Slaughter, J. C., Wu, P., Jones, S. H., Torstenson, E. S., & Hartmann, K. E. (2021). Week-by-week alcohol consumption in early pregnancy and spontaneous abortion risk: A prospective cohort study. *American Journal of Obstetrics and Gynecology, 224*(1), 97.e1-97.e16. https://doi.org/10.1016/j.ajog.2020.07.012

Suurmond R, van Rhee, H, Hak T. (2017). Introduction, comparison and validation of Meta-Essentials: A free and simple tool for meta-analysis. *Research Synthesis Methods, 8*(4), 537-553. https://doi.org/10.1002/jrsm.1260.

Tamashiro, E. M., Milanez, H. M., & Azevedo, R. C. S. D. (2020). " Because of the baby": reduction on drug use during pregnancy. *Revista Brasileira de Saúde Materno Infantil, 20*(1), 313-317. https://doi.org/10.1590/1806-93042020000100017

Terplan, M., Ramanadhan, S., Locke, A., Longinaker, N., & Lui, S. (2015). Psychosocial interventions for pregnant women in outpatient illicit drug treatment programs compared to other interventions. *Cochrane Database of Systematic Reviews*, (4), CD006037. https://doi.org/10.1002/14651858.CD006037.pub3

The Cochrane Collaboration. (n.d.). Covidence. https://community.cochrane.org/help/tools-and-software/covidence

Tsai, Y. F., Tsai, M. C., Lin, Y. P., & Chen, C. Y. (2009). Brief intervention for problem drinkers in a Chinese population: A randomized controlled trial in a hospital setting. *Alcoholism, Clinical and Experimental Research, 33*(1), 95–101. https://doi.org/10.1111/j.1530-0277.2008.00815.x

Turner, R. M., Bird, S. M., & Higgins, J. P. (2013). The impact of study size on meta-analyses: Examination of underpowered studies in Cochrane reviews. *PloS One, 8*(3), e59202. https://dx.doi.org/10.1371%2Fjournal.pone.0059202

Tzilos, G. K., Sokol, R. J., & Ondersma, S. J. (2011). A randomized phase I trial of a brief computer-delivered intervention for alcohol use during pregnancy. *Journal of Women's Health, 20*(10), 1517-1524. https://doi.org/10.1089/jwh.2011.2732

Tzilos Wernette, G., Plegue, M., Kahler, C. W., Sen, A., & Zlotnick, C. (2018). A pilot randomized controlled trial of a computer-delivered brief intervention for substance

use and risky sex during pregnancy. *Journal of Women's Health, 27*(1), 83–92. https://doi.org/10.1089/jwh.2017.6408

U.S. Office of the Surgeon General. (2005). Surgeon General's advisory on alcohol and pregnancy. http://come-over.to/FAS/SurGenAdvisory.htm

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics, 35*(2), 215-247. https://doi.org/10.3102%2F1076998609346961

van der Wulp, N. Y., Hoving, C., Eijmael, K., Candel, M. J., van Dalen, W., & De Vries, H. (2014). Reducing alcohol use during pregnancy via health counseling by midwives and internet-based computer-tailored feedback: A cluster randomized trial. *Journal of Medical Internet Research, 16*(12), e3493. https://doi.org/10.2196/jmir.3493

van Rhee, H. J., Suurmond, R., & Hak T. (2015). *User manual for Meta-Essentials: Workbooks for meta-analyses* (10th ed.). Erasmus Research Institute of Management. http://www.erim.eur.nl/research-support/meta-essentials

Wagner, C. C., & McMahon, B. T. (2004). Motivational interviewing and rehabilitation counseling practice. *Rehabilitation Counseling Bulletin, 47*(3), 152-161. https://doi.org/10.1177%2F00343552040470030401

Warren, K. R. (2015). A review of the history of attitudes toward drinking in pregnancy. *Alcoholism: Clinical and Experimental Research, 39*(7), 1110-1117. https://doi.org/10.1111/acer.12757

Waterson, E. J., & Murray-Lyon, I. M. (1990). Preventing fetal alcohol effects; a trial of three methods of giving information in the antenatal clinic. *Health Education Research, 5*(1), 53-61. https://doi.org/10.1093/her/5.1.53

Wattendorf, D. J., & Muenke, M. (2005). Fetal alcohol spectrum disorders. *American Family Physician, 72*(2), 279-285. https://www.aafp.org/afp/2005/0715/p279.html

World Health Organization. (2010). *The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST)*. World Health Organisation. https://www.who.int/publications/i/item/978924159938-2

World Health Organization. (2014). *Guidelines for the Identification and Management of Substance Use and Substance Use Disorders in Pregnancy*. World Health Organisation. https://www.who.int/publications/i/item/9789241548731

Wouldes, T. A., Crawford, A., Stevens, S., & Stasiak, K. (2021). Evidence for the effectiveness and acceptability of e-SBI or e-SBIRT in the management of alcohol and illicit substance use in pregnant and post-partum women. *Frontiers in Psychiatry, 12*, 634805. https://doi.org/10.3389/fpsyt.2021.634805

Yonkers, K. A., Forray, A., Howell, H. B., Gotman, N., Kershaw, T., Rounsaville, B. J., & Carroll, K. M. (2012). Motivational enhancement therapy coupled with cognitive behavioral therapy versus brief advice: a randomized trial for treatment of hazardous substance use in pregnancy and after delivery. *General Hospital Psychiatry, 34*(5), 439-449. https://doi.org/10.1016/j.genhosppsych.2012.06.002

Appendix A: Secondary Outcome Variables

Secondary outcome	Data type	Unit of measurement	Note
Pilot status	CAT	Pilot or non-pilot	
Eligibility of women with dependencies	CAT	Included or excluded	-
Eligibility of baseline abstainers	CAT	Included or excluded	-
Gestation eligibility cut-off	CONT	Weeks since conception	-
Location of sampling	CAT	Specialist clinic, general hospital, or other	Location of sampling is also the location of intervention delivery unless otherwise stated
Age	CONT	Mean years old	
Ethnic majority	CAT	White, African American, Hispanic, or other	Common for literature to detail ethnic majority only
Education	CONT	Percent high-school graduate or equivalent or above (≥HSG)	Common measurement in literature
Marital status	CONT	Percent married	-
Gestation	CONT	Mean weeks since conception	-
Nulliparity	CONT	Percent expecting their first birth	A birth was considered as a livebirth or the loss of a child past 20 weeks, with this cut-off reflecting a change in definition from miscarriage to stillbirth (Centers for Disease Control, 2020)
Screening tool	CAT	T-ACE, AUDIT, or no formal tool	-
Use measurement tool	CAT	TLFB, BDP, QDS, or no formal tool	-

Secondary outcome	Data type	Unit of measurement	Note
Screen and use measurement tool conductor	CAT	Self-administered or administered by research assistant	-
Use measurement period	CAT	Period ≥2-months or <2-months	Two-month cut-off represents the average time required to adopt a new behaviour (Lally et al., 2009)
Standard drink measure	CONT	grams of alcohol per standard drink	-
Brief intervention length	CONT	Minutes	-
Intervention conductor	CAT	Self-administered, administered by first author, or by trained educator	Any extra/alternative conductors and the qualifications of the first authors will also be provided, but not included for categorisations
Delivery format	CAT	Face-to-face or computer-delivered	-
Primary intervention theory	CAT	MI or education-based	Any secondary theories will also be provided, but not included for categorisation
Use recommendation	CAT	Abstinence recommended or personalised goal setting recommended	-
Bonus item	CAT	Bonus item or no bonus item	Refers to whether the intervention included a non-psychotherapeutic addition (i.e., follow-up mailing)
Control type	CAT	Usual care or reduced care	-

Note. For all variables, data is intended to be from full samples. All listed variables are intended to be included in analyses. Data will be categorised as needed. CONT = continuous outcome; CAT = categorical outcome; HSG = high school graduate; T-ACE = Tolerance, Annoyed, Cut-down, Eye-opener; AUDIT = Alcohol Use Disorders Identification Test; ATFB = Alcohol Timeline Follow-back Interview; BDP = Brief Drinker Profile; QDS = Quick Drinking Screen.

Appendix B: Database Search Strategies

PsycINFO Search Strategy

Term set	Terms
Population	(antenatal OR breast feeding OR maternity OR neonatal OR perinatal OR post-delivery OR postpartum OR postnatal OR pregnant OR prenatal OR trimester).ti,ab OR antenatal period.sh OR breastfeeding.sh OR exp obstetrics OR perinatal period.sh OR postnatal period.sh OR exp pregnancy OR prenatal care.sh
AND Behaviour	(alcohol OR drinking).ti,ab OR exp alcohol drinking patterns OR exp alcoholic beverage OR drinking behavior.sh
AND Treatment	(abstinence OR alcohol, smoking and substance involvement screening test OR alcohol treatment OR "alcohol use disorders identification test" OR ASSIST OR AUDIT OR behaviour modification OR brief advice OR brief intervention OR CAGE OR CBT OR cessation OR cognitive behaviour therapy OR counselling OR cut down, annoyed, guilty, eye-opener OR physician advice OR MI OR motivational intervention OR motivational therapy OR SBI OR screen OR screening OR single session OR T-ACE OR tolerance, annoyed, cut down, eye-opener OR tolerance, worried, eye-openers, amnesia, kut down OR TWEAK).ti,ab OR exp behavior modification OR exp cognitive behavior therapy OR exp cognitive therapy OR computer assisted therapy.sh OR counseling.sh OR drug usage screening.sh OR exp psychotherapy OR motivational interviewing.sh OR sobriety.sh

PubMed Search Strategy

Term set	Terms
Population	“antenatal”[tiab] OR “breastfeeding”[tiab] OR “breast feeding”[mh:noexp] OR “hospitals, maternity”[mh] OR “maternity”[tiab] OR “neonatal”[tiab] OR “obstetrics”[tiab] OR “obstetrics and gynecology department, hospital”[mh] OR “perinatal”[tiab] OR “perinatal care”[mh] OR “post-delivery”[tiab] OR “postpartum”[tiab] OR “postpartum period”[mh] OR “postnatal”[tiab] OR “postnatal care”[mh] OR “pregnancy”[mh:noexp] OR “pregnancy trimesters”[mh] OR “pregnant”[tiab] OR “pregnant women”[mh] OR “prenatal”[tiab] OR “trimester”[tiab]
AND Behaviour	“alcohol”[tiab] OR “alcoholic beverages”[mh] OR “drinking”[tiab] OR “drinking behavior”[mh]
AND Treatment	“abstinence”[tiab] OR "alcohol, smoking and substance involvement screening test"[tiab] OR “alcohol treatment”[tiab] OR "alcohol use disorders identification test"[tiab] OR “ASSIST"[tiab] OR "AUDIT"[tiab] OR “behavior modification”[tiab] OR “behaviour modification”[tiab] OR “brief advice”[tiab] OR “brief intervention”[tiab] OR "CAGE"[tiab] OR “CBT”[tiab] OR “cessation”[tiab] OR

Term set	Terms
	"cognitive behavior therapy"[tiab] OR "cognitive behaviour therapy"[tiab] OR "cognitive therapy"[tiab] OR "computer assisted therapy"[tiab] OR "counseling"[mh] OR "counselling"[tiab] OR "cut down, annoyed, guilty, eye-opener"[tiab] OR "drug therapy, computer assisted"[mh] OR "physician advice"[tiab] OR "psychotherapy"[mh:noexp] OR "MI"[tiab] OR "motivational intervention"[tiab] OR "motivational interviewing"[tiab] OR "motivational therapy"[tiab] OR "SBI"[tiab] OR "screen"[tiab] OR "screening"[tiab] OR "T-ACE"[tiab] OR "therapy, computer assisted"[mh:noexp] OR "tolerance, annoyed, cut down, eye-opener"[tiab] OR "tolerance, worried, eye-openers, amnesia, kut down"[tiab] OR "TWEAK"[tiab]

Embase Search Strategy

Term set	Terms
Population	("antenatal" OR "breastfeeding" OR "breast feeding" OR "maternity" OR "neonatal" OR "obstetrics" OR "perinatal" OR "post-delivery" OR "postpartum" OR "postnatal" OR "pregnant" OR "prenatal" OR "trimester"):ti,ab OR "pregnancy"/exp OR "prenatal care"/exp
AND Behaviour	("alcohol" OR "drinking"):ti,ab OR "alcohol consumption"/de OR "alcoholic beverage"/exp OR "drinking behavior"/de
AND Treatment	("abstinence" OR "alcohol, smoking and substance involvement screening test" OR "alcohol use disorders identification test" OR "ASSIST" OR "AUDIT" OR "behavio*r modification" OR "brief advice" OR "alcohol treatment" OR "CAGE" OR "CBT" OR "cessation" OR "cognitive behavio*r therapy" OR "cognitive therapy" OR "counselling" OR "cut down, annoyed, guilty, eye-opener" OR "physician advice" OR "MI" OR "motivational intervention" OR "motivational interviewing" OR "motivational therapy" OR "SBI" OR "screen" OR "screening" OR "single session" OR "T-ACE" OR "tolerance, annoyed, cut down, eye-opener" OR "tolerance, worried, eye-openers, amnesia, kut down" OR "TWEAK"):ti,ab OR "alcohol abstinence"/de OR "brief intervention"/de OR "computer assisted therapy"/de OR "counseling"/exp OR "psychotherapy"/exp

Scopus Search Strategy

Term set	Terms
Population	TITLE-ABS-KEY ("antenatal" OR "breastfeeding" OR "breast feeding" OR "maternity" OR "neonatal" OR "obstetrics" OR "perinatal" OR "post-delivery" OR "postpartum" OR "postnatal" OR "pregnancy" OR "pregnant" OR "prenatal" OR "trimester")

Term set	Terms
AND Behaviour	TITLE-ABS-KEY (“alcohol” OR “alcoholic beverage” OR “beer” OR “drinking” OR “wine”)
AND Treatment	TITLE-ABS-KEY (“abstinence” OR "alcohol, smoking and substance involvement screening test" OR “alcohol treatment” OR "alcohol use disorders identification test" OR “ASSIST” OR "AUDIT" OR “behavior modification” OR “brief advice” OR “brief intervention” OR “CAGE” OR “CBT” OR “cessation” OR “cognitive behavior therapy” OR “cognitive therapy” OR “computer assisted therapy” OR “counseling” OR "cut down, annoyed, guilty, eye-opener" OR “physician advice” OR “psychotherapy” OR “MI” OR “motivational intervention” OR “motivational interviewing” OR “motivational therapy” OR “SBI” OR “screen” OR “screening” OR “single session” OR “T-ACE” OR "tolerance, annoyed, cut down, eye-opener" OR "Tolerance, worried, eye-openers, amnesia, kut down" OR “TWEAK”)

CINAHL Search Strategy

Term set	Terms
Population	TI (“antenatal” OR “breastfeeding” OR “breast feeding” OR “maternity” OR “neonatal” OR “obstetrics” OR “perinatal” OR “post-delivery” OR “postpartum” OR “postnatal” OR “pregnant” OR “prenatal” OR “trimester”) OR AB (“antenatal” OR “breastfeeding” OR “breast feeding” OR “maternity” OR “neonatal” OR “obstetrics” OR “perinatal” OR “post-delivery” OR “postpartum” OR “postnatal” OR “pregnant” OR “prenatal” OR “trimester”) OR MH “obstetric patients” OR MH “postnatal period+” OR MH “pregnancy+” OR MH “pregnancy trimesters+”
AND Behaviour	TI (“alcohol” OR “drinking”) OR AB (“alcohol” OR “drinking”) OR MH “alcoholic beverages+” OR MH “drinking behavior+”
AND Treatment	TI (“abstinence” OR “alcohol treatment” OR "alcohol, smoking and substance involvement screening test" OR "alcohol use disorders identification test" OR "ASSIST" OR "AUDIT" OR “behavio#r modification” OR “brief advice” OR “brief intervention” OR “CAGE” OR “CBT” OR “cessation” OR “cognitive behavio#r therapy” OR "cut down, annoyed, guilty, eye-opener" OR “physician advice” OR “MI” OR “motivational intervention” OR “motivational therapy” OR “SBI” OR “screen” OR “screening” OR “single session” OR "T-ACE" OR "tolerance, annoyed, cut down, eye-opener" OR "tolerance, worried, eye-openers, amnesia, kut down" OR "TWEAK") OR AB (“abstinence” OR “alcohol treatment” OR "alcohol, smoking and substance involvement screening test" OR "alcohol use disorders identification test" OR "ASSIST" OR "AUDIT" OR “behavio#r modification” OR “brief advice” OR “brief intervention” OR “CAGE” OR “CBT” OR “cessation” OR “cognitive behavio#r therapy” OR "cut down, annoyed, guilty, eye-opener" OR “physician advice” OR “MI” OR “motivational intervention” OR “motivational therapy” OR “SBI” OR “screen” OR

Term set	Terms
	"screening" OR "single session" OR "T-ACE" OR "tolerance, annoyed, cut down, eye-opener" OR "tolerance, worried, eye-openers, amnesia, kut down" OR "TWEAK") OR MH "alcohol rehabilitation programs+" OR MH "cognitive therapy+" OR MH "counseling+" OR MH "drug therapy, computer assisted" OR MH "psychotherapy+" OR MH "motivational interviewing"

Cochrane Library Search Strategy

Term set	Terms
Population	("antenatal" OR "breastfeeding" OR "maternity" OR "neonatal" OR "obstetrics" OR "perinatal" OR "post-delivery" OR "postpartum" OR "postnatal" OR "pregnant" OR "prenatal" OR "trimester"):ti,ab,kw OR MeSH descriptor: [breast feeding] this term only OR MeSH descriptor: [hospitals, maternity] explode all trees OR MeSH descriptor: [obstetrics and gynecology department, hospital] explode all trees OR MeSH descriptor: [perinatal care] explode all trees OR MeSH descriptor: [postpartum period] explode all trees OR MeSH descriptor: [postnatal care] explode all trees OR MeSH descriptor: [pregnancy] this term only OR MeSH descriptor: [pregnancy trimesters] explode all trees OR MeSH descriptor: pregnant women] explode all trees
AND Behaviour	("alcohol" OR "alcoholic beverage" OR "drinking"):ti,ab,kw OR MeSH descriptor: [drinking behavior] explode all trees
AND Treatment	("abstinence" OR "alcohol treatment" OR "behavior modification" OR "brief advice" OR "brief intervention" OR "CBT" OR "cessation" OR "cognitive behavior therapy" OR "cognitive therapy" OR "computer assisted therapy" OR "counselling" OR "physician advice" OR "MI" OR "motivational intervention" OR "motivational interviewing" OR "motivational therapy"):ti,ab,kw ("Alcohol, Smoking and Substance Involvement Screening Test" OR "Alcohol Use Disorders Identification Test" OR "ASSIST" OR "AUDIT" OR "CAGE" OR "Cut down, Annoyed, Guilty, Eye-Opener" OR "SBI" OR "screen" OR "T-ACE" OR "Tolerance, Annoyed, Cut down, Eye-opener" OR "TWEAK" OR "Tolerance, Worried, Eye openers, Amnesia, Kut down"):ti,ab,kw OR MeSH descriptor: [counseling] explode all trees OR MeSH descriptor: [drug therapy, computer assisted] explode all trees OR MeSH descriptor: [psychotherapy] explode all trees MeSH descriptor: [therapy, computer assisted] this term only

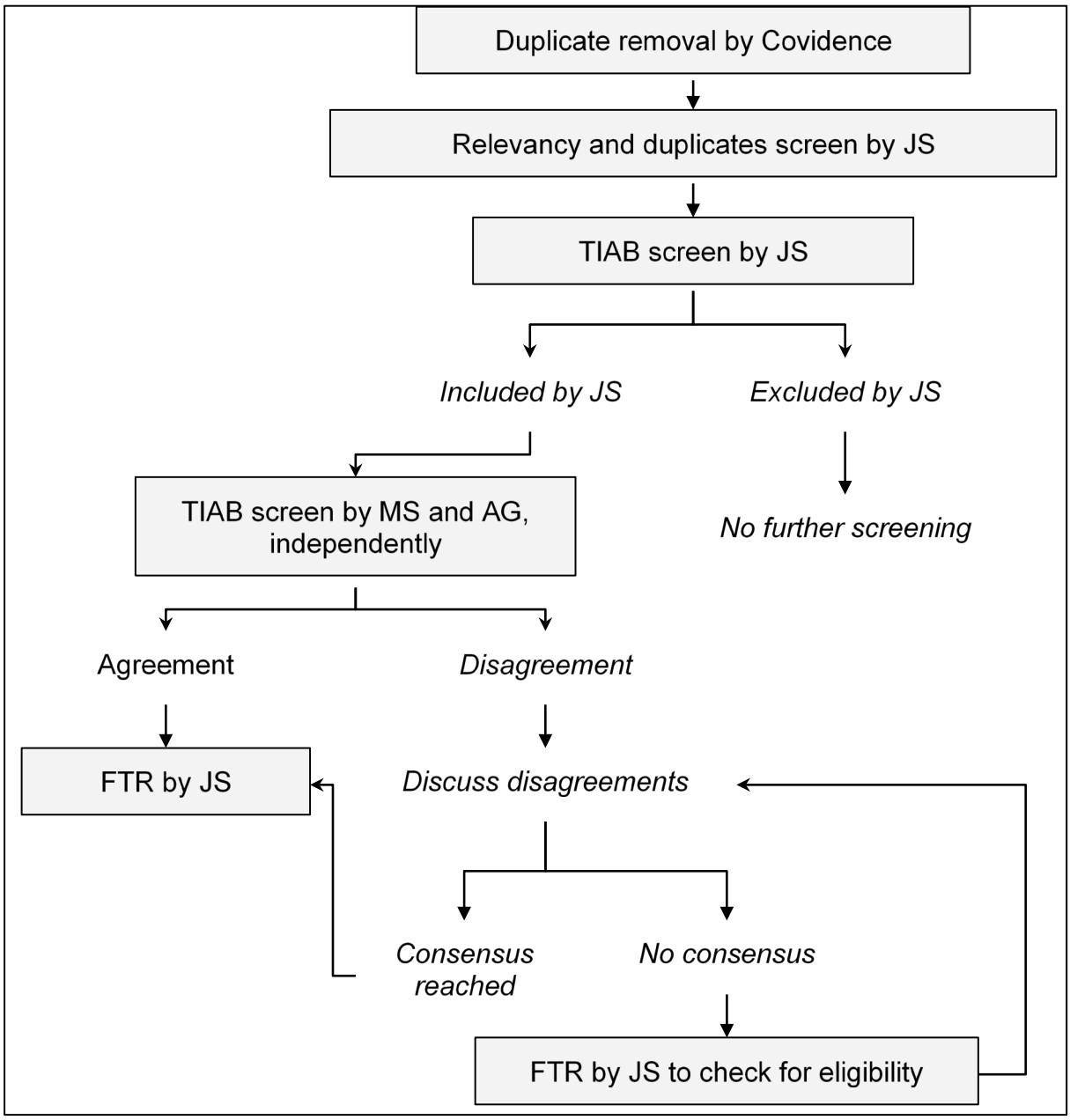
Note. All Cochrane Library databases were included in the search.

Aboriginal HealthInfoNet Search Strategy

Term set	Terms
Population	Pregnancy OR reproduction OR birth
AND Behaviour	Alcohol

Note. Aboriginal HealthInfoNet was searched using topic headers only. Sources included were journal articles, online journal articles, reports, and theses.

Appendix C: Disagreement Resolution Process for Screening



Note. JS = first author; MS = third author; AG = fifth author; TIAB = title and abstract; FTR = full-text review.

Appendix D: Data Items Extracted

Data were collected where provided for:

- publication details (as required for citations, references),
- author contact emails (for author contacting),
- location of publication,
- funding sources,
- declarations of interest among primary authors,
- pilot status,
- sample size at analysis (intervention group [IG] and control group [CG]),
- mean (with SD) drinking days at base (any unit, IG and CG),
- mean (with SD) drinking days at follow-up (any unit, IG and CG),
- mean (with SD) change in drinking days (any unit, IG and CG),
- mean (with SD) drinks per drinking period at base (any unit, IG and CG),
- mean (with SD) drinks per drinking period at follow-up (any unit, IG and CG),
- mean (with SD) change in drinks per drinking period (any unit, IG and CG),
- number or percentage of participants abstinent at base (IG and CG),
- number or percentage of participants abstinent at follow-up (IG and CG),
- number or percentage of participants who maintained abstinence through intervention period (IG and CG),
- location of sampling and brief intervention delivery
- age (any unit, total sample or IG and CG),
- gestation at base (total sample or IG and CG),
- education (any unit, total sample or IG and CG),
- marital status (any unit, total sample or IG and CG),
- nulliparity (any unit, total sample or IG and CG),

- ethnic majority (total sample or IG and CG),
- eligibility status of participants with high-level use or dependencies,
- eligibility status of participants who were abstinent as of base,
- maximum gestation eligibility cut-off,
- what was considered one standard drink (any unit),
- screening tool,
- screening conductor,
- use measurement tool at base,
- use measurement conductor at base,
- use measurement tool at follow-up,
- use measurement conductor at follow-up,
- brief intervention length,
- brief intervention delivery method (face-to-face or computer-delivered),
- brief intervention theoretical foundations,
- brief intervention conductor and their qualifications,
- brief intervention bonus items (defined in Appendix A),
- whether abstinence or harm minimisation was recommended to participants,
- measurement period of the study, and
- control type.

Appendix E: Template Author Contact Email

Yellow text was adapted for each email.

Dear Dr A. Author,

My name is Joel Smith. I am a current Honours Psychology student at the University of Adelaide, South Australia.

For my Honours Thesis, I am conducting a meta-analysis of the effectiveness of brief alcohol interventions on reducing alcohol use in pregnant women. My supervisors are Dr Andrea Gordon, Professor Paul Delfabbro, Dr Robert Ali, and Dr Matthew Stevens. The primary outcomes that I am investigating are below, comparing brief intervention groups to control groups.

#	Outcome	Effect measurement
1	**Frequency**, reported as drinking days per month (28 days) at follow-up	**Cohen's d**, requires mean standardised effect (with standard deviation) and group sample size
2	**Quantity**, reported as drinks per drinking day at follow-up	
3	**Frequency change**, reported as change in drinking days per month across measurement period	
4	**Quantity change**, reported as change in drinks per drinking day across measurement period.	
5	**Abstinence maintenance**	**Odds Ratio**, requires number of participants who abstained from base to follow-up and group sample size

I otherwise intend to run several secondary analyses comparing these outcomes against differences in sample demographics, screening and brief intervention formats, and methodologies.

I am hoping to include your below study/ studies in my meta-analysis.

- A. Author. (2021). Brief intervention for antenatal alcohol consumption.

While I have already extracted the available data from your published work, I was hoping that you might be able to provide further information so that I can standardise the outcomes into Cohen's d/ Odds ratio values.

If possible, could you please provide me with the information listed on the attached spreadsheet? The spreadsheet includes spots for demographics data/ methodology information, and the information that I have already extracted for you to check if desired. If you do not have the requested data, or did not record it, please type NA in the relevant box. If you have the data, but in a different unit of measurement to that requested, please provide the data in whatever unit of measurement is available.

Please let me know if you are able to provide the requested data. If you are unable to provide the data, I understand, but if you could please let me know, that would be much appreciated.

If you have any questions or concerns, please let me know also. Thanks in advance!

Kind regards,

Joel Smith

Bachelor of Psychology (Honours) Student, University of Adelaide, South Australia

Appendix F: Example Data Request Form

Example is from the Chang et al. (2005) data request form.

Citation (short)		Chang et al., 2005		
Title		Brief Intervention for Prenatal Alcohol Use: A Randomized Trial		**SD = Standard Deviation
Data point		**Relevant data (please complete missing data here)**	**Notes from J Smith**	**NOTE: You mentioned in the study that were two lots of analyses, one comparing intervention to control, and one comparing the effects of intervention with and without partner involvement. Could you please provide the intervention vs control data, i.e. data where both participants with partner involvement and without were included**
Sample sizes at analysis	Intervention group (IG) Control group (CG)	152 152		
Mean drinking days per month (28 days) at follow-up (with SD)	IG CG		If you collected this data in any other format (i.e. per episode), that information would also be appreciated	
Mean change in drinking days per month (28 days) across reporting period (with SD)	IG CG		In your study, you mention % reduction in drinking days. Does the raw number for change in drinking days exist (so, not a percentage)? If you have 'per month' data, that would be great, otherwise, 'per episode' data is also appreciated	
Mean drinks per drinking day at follow-up (with SD)	IG CG		If you collected this data in any other format (i.e. per episode), that information would also be appreciated	
Mean change in drinks per drinking day across reporting period (with SD)	IG CG	-0.39 -0.4	Please provide the SD values if available	
Number of participants who maintained abstinence	IG CG		This does not appear to be mentioned in your report. Please provide the data only if it is available	
How 'drink' was defined			For example, you mentioned it being any consumption, even a sip, in Chang et al., 1999	
Mean age (with SD)	Whole sample			
% High-school graduate or above	Whole sample			
% married	Whole sample	80.5		
% nulliparous	Whole sample			
How was 'nulliparous' defined			For example, nulliparous = no live or still births, or miscarriages past 20 wks of gestation. Please include any other definitions that you used (if actually measured)	
Mean weeks of gestation at base (with SD)	Whole sample			
Gestation eligibility cut-off		28 wks		
Ethnic majority	Whole sample	White		
Were women with alcohol dependency included?		No		
Were women who screened positive but were abstinent at baseline use measurment included?		Yes		

Note. The formatting was changed slightly (i.e., column width reduced) to fit this figure to the page.

Appendix G: Data Transformation Process

Primary Data Transformations

#[1]	Outcome	Transformation	Transformation process
1	FFU, FC	Converting mean (SD) drinking days per invalid reporting period to drinking days per 28-day month	1. Calculate number of days in reporting period (i.e., seven days for per week data) 2. Divide this value by 28 to get the number of months in this period (i.e., 0.25 months for per week data) 3. Divide the reported mean drinking days by this number of months to generate drinking days per month values To transform associated SD values: 4. Divide SD values by number of months (same process as step 3)
2	FFU, FC	Converting percent drinking days per reporting period to drinking days per 28-day month	1. Calculate number of days in reporting period (i.e., seven days for per week data) 2. Apply the percentage of drinking days to the number of days in the reporting period to generate the number of drinking days per reporting period (i.e., for 2% of drinking days per 158-day reporting period, multiply 158 by 0.02) 3. Divide the value from step 1 by 28 to get the number of months in this period (i.e., 5.64 months for 158-day reporting period) 4. Divide the calculated mean drinking days (i.e., 158 * 0.02) by this number of months to generate drinking days per month values To transform associated SD values: 5. Divide SD values by number of months (same process as step 4)
3	FFU, FC, QFU, QC	Estimating SD values where none are reported	1. Input available data (t-test *p*-value, t-test value, and or standard error values) into the Cochrane Finding_SDs calculator (Drahota & Beller, n.d.)
4	FC, QC	Calculating mean (SD) change in use across reporting period	1. Subtract mean use at baseline by mean use at follow-up (so, follow-up negative baseline) To transform associated SD values: 2. Calculate the SD of change per Equation Figure 1 from CHSRI section 6.5.2.8 (Higgins et al., 2021b). If the required within-group correlation coefficient is not provided, then that of a similar study can be used or calculated and used per Equation Figure 1 (Higgins et al., 2021b)

#[1]	Outcome	Transformation	Transformation process
5	AR, AMR	Convert percent abstinent to number of abstainers	1. Apply percentage of abstainers to relevant sample size (i.e., 44% abstinence in a sample of *n* = 16 computes to seven abstainers)

Note. Some data may not require the full transformation or require multiple transformations. FFU = frequency at follow-up; FC = frequency change, QFU = quantity at follow-up; QC = quantity change; AR = abstinence rate; AMR = abstinence maintenance rate; SD = standard deviation.

[1] Refers to transformation number. This number will be used to track the transformations made in the results.

Equation Figure 1

Equation to Calculate Standard Deviations of Change Data

$$\mathrm{SD}_{E,\mathrm{change}} = \sqrt{\mathrm{SD}^2_{E,\mathrm{baseline}} + \mathrm{SD}^2_{E,\mathrm{final}} - \left(2 \times \mathrm{Corr} \times \mathrm{SD}_{E,\mathrm{baseline}} \times \mathrm{SD}_{E,\mathrm{final}}\right)} \quad (1)\ \text{Calculate change SD}$$

$$\mathrm{Corr}_E = \frac{\mathrm{SD}^2_{E,\mathrm{baseline}} + \mathrm{SD}^2_{E,\mathrm{final}} - \mathrm{SD}^2_{E,\mathrm{change}}}{2 \times SD_{E,\mathrm{baseline}} \times SD_{E,\mathrm{final}}} \quad (2)\ \text{Calculate correlation coefficient (then apply to [1])}$$

Note. Calculations shown for one group only. Corr = within-group correlation coefficient; SD = standard deviation; $_E$ = experimental group. Adapted from "6.5.2.8 Imputing standard deviations for changes from baseline" by J. P. T. Higgins, T. Li, & J. J. Deeks, in J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (version 6.2), (2021b), Cochrane. www.training.cochrane.org/handbook.

Secondary Data Transformations

#[1]	Transformation	Transformation process
6	Converting mL pure alcohol to g pure alcohol (standard drink measure conversion)	1. Multiply mL pure alcohol by 0.8, where 1mL pure alcohol is equivalent to 0.8g pure alcohol (Alcohol and Drug Foundation, 2017)
7	Converting percent-based demographics data from per group statistics to total sample statistics	1. Apply the percentage of participants fitting the demographic in the IG to the number of participants total in the IG group (i.e., if 43.4% of the 152 IG are nulliparous, then multiply 152 by 0.434). 2. Repeat step 1 for CG data 3. Sum the number of participants fitting the demographic the IG and CG groups together 4. Divide the number of total participants fitting the demographic by the total number of participants in the study sample (and multiply by 100 to make this value a percentage)
8	Converting mean (with SD) demographics data from per group statistics to total sample statistics	1. Calculated per CHSRI section 6.5.2.10 (Higgins et al., 2021b), shown in Equation Figure 2.

[1] Refers to transformation number. This number will be used to track the transformations made in the results. SD = standard deviation; IG = intervention group; CG = control group.

Inverting and Combining Data

#[1]	Transformation	Transformation process
9	Inversion of percentage values	1. If data were provided for the inverse outcome, the percentage was inverted (i.e., from 67% not abstaining to 33% abstaining for abstinence rate data)
10	Summing of demographic reporting levels	1. If data were presented in multiple levels that reflected the same outcome defined by this review, then it was combined (i.e., summing 42.5% HSG and 21.2% university educated to get 63.7% ≥HSG)

[1] Refers to transformation number. This number will be used to track the transformations made in the results. HSG = high school graduate.

Equation Figure 2

Equation to Calculate Total Sample Data from Intervention Group and Control Group Data

	Combined groups
Sample size	$N_1 + N_2$
Mean	$\frac{N_1 M_1 + N_2 M_2}{N_1 + N_2}$
SD	$\sqrt{\frac{(N_1 - 1)\mathrm{SD}_1^2 + (N_2 - 1)\mathrm{SD}_2^2 + \frac{N_1 N_2}{N_1 + N_2}(M_1^2 + M_2^2 - 2M_1 M_2)}{N_1 + N_2 - 1}}$

Note. N = sample size; M = mean, SD = standard deviation; $_1$ = group one; $_2$ = group two. Adapted from "6.5.2.10 Combining groups" by J. P. T. Higgins, T. Li, & J. J. Deeks, in J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (version 6.2), (2021b), Cochrane. www.training.cochrane.org/handbook.

Appendix H: Data Categorisation Process

Data type	Viable categories	Categorisations
Location of sampling	Specialist clinic, general hospital, or other[1]	• Specialist clinic = obstetrics, prenatal care, and maternity clinics • General hospital = hospital settings without specialisations listed
Ethnic majority	White, African American, Hispanic, or other[1]	• White = white or Caucasian • African American = African American, non-Hispanic African American, or black • Hispanic = South American Hispanic (assumed of American studies)
Use measurement period	Period ≥2-months or <2-months	• ≥2-months = any measurement period greater than two months, including base-to-delivery • <2-months = any period less than two months
Primary intervention theory	MI or education-based	• MI = MI, including studies with secondary theories • Education-based = Education, including studies with secondary theories
Bonus item	Bonus item or no bonus item	• Bonus item = any added item beyond initial therapeutic session • No bonus item = no bonus item
Control type	Usual care or reduced care	• Usual care = usual care • Reduced care = more than usual care, but less than intervention group

Note. Data categorisations were made per the categories described in Appendix A. Some categorical data did not require categorisation. MI = motivational interviewing.

[1] Any other non-viable category reported.

Appendix I: Risk of Bias Assessment Guidelines

Risk of bias assessments were made per CHSRI guidelines (Higgins et al., 2021c). Assessments were made at three levels: bias indicator, domain, and outcome levels. Bias indicators were used to assess evidence of risk of different bias domains. In turn, these domains reflected the overall level of bias present for an outcome. Five biases were assessed:

(1) Bias arising from the randomisation process.

(2) Bias due to deviations from intended interventions.

(3) Bias due to missing outcome data.

(4) Bias in measurement of the outcome.

(5) Bias in selection of the reported results.

Domains 1-2 were assessed once per study, as they relate to study processes. Domains 3-5 were assessed per relevant outcome, as they relate to the measurement and reporting of specific outcomes. How specific bias indicators were judged are provided on the next page, with an outline of how domains and outcomes were judged provided the page after.

Bias Indicator Assessment Guidelines

Bias domain	Bias indicator	Low risk	Some risk	High risk
1	Allocation sequence	Randomised	Not stated	Not randomised
	Allocation sequence concealment	Concealed	Not stated	Not concealed
	Significant baseline different between groups (demographic or alcohol use)	None	Not stated	Some
2	Blinding of allocation (participants)	Blinded	Unclear	Not blinded
	Blinding of allocation (intervention conductors)	Blinded	Unclear	Not blinded
	Deviations from intended intervention	Integrity assessed and no deviations	Not stated	Deviations
3	Exclusions for analysis[1]	No exclusions	<5% exclusion rate	≥5% exclusion rate or any significant differences in those excluded
4	Appropriateness of measurement tool and unit of measurement	Both appropriate	-	Either or both inappropriate
	Differences in measurement between groups	None	-	Some
	Blinding of allocation (assessor)[2]	Blinded	Unclear	Not blinded

Bias domain	Bias indicator	Low risk	Some risk	High risk
5	Adherence to protocol	Adhered	Unclear	Not adhered
	Selective reporting (outcomes)	None	-	Some
	Selective reporting (analyses)	None	-	Some

[1] 5% cut-off rate based on CHSRI (Higgins et al., 2021c).

[2] Where the assessor was whoever ran follow-up assessments or conducted hair EtG.

Domain and Outcome Assessment Guidelines

	Low risk	Some risk	High risk
Domain	No bias indicators judged as having some or high risk	Two or less bias indicators judged as having some risk without the overall risk of the domain being questioned	Any bias indicators judged as having high risk, or three bias indicators judged as having some risk, or two bias indicators judged as having some risk to a degree that the overall risk of the domain is questioned
Outcome	As above, but replacing bias indicators with domains	As above, but replacing bias indicators with domains	As above, but replacing bias indicators with domains

Appendix J: Records of Author Contacting and New Data

Study citation	Author contacted	Author response	Notes
(Chang et al., 1999)	First	No further data provided	-
(Chang et al., 2005)	First	No further data provided	-
(Handmaker et al., 1999)	Second, who forwarded to first for response	No further data provided	Could not locate first author's email, so initially contacted second author
(Joya et al., 2016)	NA	NA	Could not locate contacts for first, second, or third authors, so no contact made
(Ondersma et al., 2015)	First, who forwarded to second for response	Further primary and secondary data provided	-
(Osterman & Dyehouse, 2012)	First then second after no response	No response	Contacted first author, who did not respond, so second author was then contacted
(Osterman et al., 2014)	First then second after no response	No response	Contacted first author, who did not respond, so second author was then contacted
(Reynolds et al., 1995)	First	No further data provided	-
(Tzilos et al., 2011)	First	Further secondary data provided	-

The new data provided for Ondersma et al. (2015) was: mean drinking days in the past 90-days (0.45 [1.61] for intervention group, 1.11 [3.21] for control); mean drinks per drinking day in the past 90-days (0.55 [1.88] for intervention group, 0.53 [1.18] for control); and that a standard drink was defined as 14g of pure alcohol (J. Beatty, personal communication, July 31, 2021). The new data provided for Tzilos et al. (2011) was: a standard drink was defined as 14g pure alcohol; the gestation cut-off was 36 weeks; and that women with high-level use or dependencies were eligible for the study (G. Tzlios Wernette, personal communication, July 23, 2021).

Appendix K: Characteristics of Included Studies

Characteristics of Included Studies I: Study Details

Data type	Study citation								
	(Chang et al., 1999)	(Chang et al., 2005)	(Handmaker et al., 1999)	(Joya et al., 2016)	(Ondersma et al., 2015)	(Osterman & Dyehouse, 2012)	(Osterman et al., 2014)	(Reynolds et al., 1995)	(Tzilos et al., 2011)
Location	United States	United States	United States	Spain	United States	United States	United States	United States	United States
Pilot status	No	No	Yes	No	Yes	No	No	No	Yes
Funding	NIAAA	NIH	NIH	See note	NIH	UC	NIH	NIAAA	NIAAA
Declarations	-	-	-	None	See note	None	-	-	None
Women DEP	No	No	Yes[1]	Yes[1]	Yes	No	Yes	Yes	Yes[†]
Women BA	See notes	Yes	-	Yes	Yes	-	Yes	No	Yes
MGest (wks)	28	28	-	-	28	36	36	36	36

Note. Chang et al. (1999) included baseline abstainers only for abstinence outcome analyses. Joya et al. (2016) funded by the Fondo Europeo de DesarrolloRegional, Instituto Carlos III, Generalitat de Catalunya, RecerCaixa and Fundación Mutua Madrileña. Ondersma et al. (2015) first author part-owns the company that developed their intervention. Women DEP = women with high-level use or dependencies; women BA = baseline abstinent women; MGest (wks) = max gestation (weeks); NIAAA = National Institute on Alcohol Abuse and Alcoholism; NIH = National Institutes of Health; UC = University of Cincinnati.

[1] Women with high-level use were included, but dependency was not formally assessed.

[†] Data was not presented in the text and was instead gathered from authors (Appendix J).

Characteristics of Included Studies II: Demographics at Baseline

Data type	Study citation								
	(Chang et al., 1999)	(Chang et al., 2005)	(Handmaker et al., 1999)	(Joya et al., 2016)	(Ondersma et al., 2015)	(Osterman & Dyehouse, 2012)[1]	(Osterman et al., 2014)	(Reynolds et al., 1995)	(Tzilos et al., 2011)
n at base (IG, CG)	250 (123, 127)	304 (152, 152)	42 (20, 22)	168 (83, 85)	48 (24, 24)	67 (39, 28)	122 (61, 60)	78 (42, 36)	50 (27, 23)
Sampling location[2]	Specialist clinic‡	Specialist clinic‡	See notes	General hospital	Specialist clinic‡	Specialist clinic‡	Specialist clinic‡	Specialist clinic‡	Specialist clinic‡
Age (yrs)	x̄ 30.7 (±5.4)	Med 31.4§	x̄ 24 (±5.76)	x̄ 31.1 (±5.48)‡	54.2% 18-25, 33.3% 26-33, 12.5% 34-37§	x̄ 24.9§	x̄ 25.6 (±4.81)‡	x̄ 22.4	x̄ 26 (±5.2)
Ethnic m	White‡	White	Hispanic	Other‡	AA	AA‡	AA‡	AA	AA
Education	96% ≥HSG‡	Med HSG§	x̄ 12 (±2.71) yrs§	94.7% ≥HSG‡	66.7% ≥HSG	75.4% ≥HSG§	69% ≥HSG‡	-	50% ≥HSG
Married (%)	74%	80.5%	38%‡	-	20.8%	83% single§	14.7%‡	35%‡	26% single IG and CG§
Nulli (%; definition)	53% (no definition)	41.9% (no children)‡	-	41.4% (no pregnancies)‡	-	x̄ 1.58 deliveries§	x̄ 1.95 (±2.03) deliveries§	-	-

Data type	Study citation								
	(Chang et al., 1999)	(Chang et al., 2005)	(Handmaker et al., 1999)	(Joya et al., 2016)	(Ondersma et al., 2015)	(Osterman & Dyehouse, 2012)[1]	(Osterman et al., 2014)	(Reynolds et al., 1995)	(Tzilos et al., 2011)
Gest (wks)	x̄ 16 (±4.6)	Med IG 11, Med CG 12[§]	-	-	x̄ 12.2 (±5.4)	x̄ 20.71[§]	x̄ 23.4 (±8.69)[‡]	x̄ 12.3	x̄ 25 (±8)

Note. Handmaker et al. (1999) recruited from obstetrics clinics though administered the intervention at a children's hospital, participant homes, and a psychology clinic. Data reflects total sample unless otherwise stated. Bracketed data reflects standard deviations (excluding nulliparity definitions). yrs = years; wks = weeks; ethnic m = ethnic majority; nulli = nulliparity; gest = gestation; x̄ = mean; Med = median; IG = intervention group; CG = control group; HSG = high school graduate; AA = African American.

[1] Demographics data sampled ineligible and eligible women, so all demographics data were excluded.

[2] Sampling location also refers to the location where the intervention was delivered, expect for Handmaker et al. (1999).

[‡] Data required some transformation or categorisation (Appendix O).

[§] Data not viable for analysis due to ineligible units of analysis (Appendix O). Or, in the case of Osterman and Dyehouse (2012), for including ineligible women in data.

Characteristics of Included Studies III: Screening and Brief Intervention, Use Measurement, and Control Details

Data type	Study citation								
	(Chang et al., 1999)	(Chang et al., 2005)	(Handmaker et al., 1999)	(Joya et al., 2016)	(Ondersma et al., 2015)	(Osterman & Dyehouse, 2012)[1]	(Osterman et al., 2014)	(Reynolds et al., 1995)	(Tzilos et al., 2011)
Screen	T-ACE (-)	T-ACE (RA)	NFT (-)	AUDIT (-)	T-ACE (SA)	AUDIT (-)	AUDIT (RA)	NFT (SA)	T-ACE (RA)
UM at base	ATFB (RA)	ATFB (RA)	BDP (RA)	[1]	None	NA	QDS (RA)	NFT (SA)	ATFB (RA)
UM at FU	ATFB (RA)	ATFB (RA)	BDP (RA)	[1]	ATFB (SA)	ATFB (-)	QDS (RA)	NFT (SA)	ATFB (RA)
UM report period	≥2-months‡	≥2-months‡	≥2-months‡	≥2-months‡	≥2-months‡	<2-months‡	<2-months‡	≥2-months‡	<2-months‡
S. drink unit	See notes	-	12g‡	[1]	14g†	-	-	-	14g†
BI length in minutes	45 (FA_1)	25 (FA_1 or nurse)	60 (FA_2)	-	20 (SA)	30 (FA_3)	30 (FA_3)	30 (TE)	30 (SA)
BI delivery	F2F	F2F	F2F	-	By computer	F2F	F2F	F2F	By computer
Primary BI theory	-	-	MI	MI	MI	MI	MI‡	Education‡	MI
Use rec.	Abs	Abs	Abs	-	PGS	PGS	-	-	PGS
Bonus item	No	No	No	No	Yes‡	No	No	Yes‡	No

Data type	Study citation								
	(Chang et al., 1999)	(Chang et al., 2005)	(Handmaker et al., 1999)	(Joya et al., 2016)	(Ondersma et al., 2015)	(Osterman & Dyehouse, 2012)[1]	(Osterman et al., 2014)	(Reynolds et al., 1995)	(Tzilos et al., 2011)
Control	Usual care	Usual care	Reduced care‡	Usual care	Reduced care‡	Usual care	Usual care	Usual care, see notes	Reduced care‡

Note. Chang et al. (1999) considered any consumption, even a sip, as a drink. Ondersma et al. (2015) bonus item was three tailored mailings and reduced care was a less-interactive BI on infant nutrition. Reynolds et al. (1995) bonus item was a self-help manual with a one-minute call to check manual progress and usual care included providing referrals for women with dependencies. Handmaker et al. (1999) reduced care was referrals and letters about the risk of antenatal alcohol consumption. Tzilos et al. (2011) reduced care was a computer activity about television preferences. UM = use measurement; FU = follow-up; s. drink = standard drink; BI = brief intervention; use rec. = recommendation around use; T-ACE = Tolerance, Annoyed, Cut-down, Eye-opener; NFT = no formal tool; AUDIT = Alcohol Use Disorders Identification Test; ATFB = Alcohol Timeline Follow-back Interview; BDP = Brief Drinker Profile; QDS = Quick Drinking Screen; SA = self-administered; RA = administered by research assistant; FA = administered by first author ($_1$ = Doctor of Medicine; $_2$ = clinical psychology PhD candidate; $_3$ = psychiatric mental health clinical nurse); TE = administered by trained educator; F2F = face-to-face; MI = motivational interview; Abs = abstinence recommended; PGS = personlised goal setting recommended.

[1] Joya et al. (2016) conducted a modified ATFB but did not present results. Use measurement was conducted using ethyl glucuronide quantification of hair samples instead.

† Data was not presented in the text and was instead gathered from authors (Appendix J)

‡ Data required some transformation or categorisation (Appendix O).

Characteristics of Included Studies IV: Measurement of Relevant Outcomes

Data type	Study citation								
	(Chang et al., 1999)	(Chang et al., 2005)	(Handmaker et al., 1999)	(Joya et al., 2016)	(Ondersma et al., 2015)	(Osterman & Dyehouse, 2012)	(Osterman et al., 2014)	(Reynolds et al., 1995)	(Tzilos et al., 2011)
n at analysis (IG, CG)	247 (unclear)	304 (152, 152)	34 (16, 18)	168 (83, 85)	39 (20, 19)	56 (28, 26), see notes	93 (44, 49)	72 (39, 33)	50 (27, 23)
FFU	Yes (p-EP) [§]	Yes (%-DD)	Yes (%-DA) [§]	-	Yes (p-RP) [†]	Yes (p-week)	Yes (p-week)	-	-
FFC	-	-	-	-	-	Yes (p-week)	Yes (p-week)	-	-
QFU	-	Yes (p-EP) [§]	Yes (p-RP) [§]	-	Yes (p-RP) [†]	Yes (p-day) [2]	Yes (p-day)[2]	-	-
QC	Yes (p-DD) [§]	-	-	-	-	Yes (p-day) [2]	Yes (p-day)[2]	-	-
AR	No	-	Yes	Yes	Yes	-	-	Yes	-
AMR	Yes	-	Yes[§]	Yes	-	-	-	-	-
Alt. arm[1]	No	Yes	No	Yes	No	No	Yes	No	No

Note. Units are bracketed. Sample sizes are lower for AMR values, as max only considers baseline abstainers. Osterman & Dyehouse (2012) *n* = 53 for change data (intervention and control group sizes are unclear). Alt. arm = alternate study arm; FFU = frequency at follow-up; FC = frequency change, QFU = quantity at follow-up; QC = quantity change; AR = abstinence rate; AMR = abstinence maintenance rate; IG = intervention group; CG = control group; p = per; EP = episode; DD = drinking day (drinking days when following %-); DA = days abstinent; RP = reporting period.

[1] Alternative arm details and judgments provided in method.

[2] Reported as per day, not specified whether this is per drinking day or per measurement day, but drinking day is presumed (justification in Appendix O).

[†] Data was not presented in the text and was instead gathered from authors (Appendix J).

[§] Data not viable for analysis due to ineligible units of analysis or methodological concerns (Appendix O).

Appendix L: Notable Inclusions and Exclusions from Review

Notable Inclusions for Review

Study citation	Reason why inclusion was notable	Why study was included
(Joya et al., 2016)	Brief intervention length not specified	Intervention was nonetheless defined as brief so assumed to be ≤60 minutes
(Ondersma et al., 2015)	Included interaction beyond initial session (sent three personalised letters to participants throughout their remaining gestation)	Being entirely different to the initial session, this was considered a bonus item (defined in Appendix A), rather than an independent second session
(Reynolds et al., 1995)	Included interaction beyond initial session (provided a self-help manual and a one-minute phone call to discuss manual progress)	Same as above
	Referred women with high use for further treatment, potentially improving effect size	The intervention and control groups did not significantly differ in baseline levels of use (Reynolds et al., 1995), so both should have been equally affected. It was deemed that impacts on overall effects would be no different to those observed due to studies including reduced care controls instead of usual care controls, so was included and considered a methodological difference of the study

Notable Exclusions from Review

Study citation	Reason why exclusion was notable	Why study was excluded[1]
(Crawford-Williams et al., 2016)	Included in Gomez et al. (2020) analysis	Intervention was a booklet only and therefore did not have an interactive psychotherapeutic element
(O'Connor & Whaley, 2007)	Multi-session program but with potential for single-session data from attrition	Individual sessions of multi-session programs deemed incomparable to intentionally single-session interventions
(Rubio et al., 2014)	Same as above	Same as above
(Tamashiro et al., 2020)	Same as above	Same as above
(van der Wulp et al., 2014)	Health counselling group included in Gomez et al. (2020) analysis	Intervention was over three sessions
	Tailored letter group included in Gomez et al. (2020) analysis and similar to Ondersma et al. (2015), which was included	Unlike Ondersma et al. (2015), there was no interactive psychotherapeutic element to intervention
(Yonkers et al., 2012)	Included in Gomez et al. (2020) analysis	Screening tool was adapted for polydrug assessment

[1] Reasons for exclusions correspond with eligibility and ineligibility criteria.

Appendix M: Risk of Bias Tables

Chang et al. (1999) Risk of Bias Table

Bias domain	Risk	Support for judgement
1	Some	Allocation sequence generation involved simple randomisation conducted by computer algorithm. Status of allocation sequence concealment is unclear. Significant baseline differences in pre-assessment drinks per drinking day only (higher in control)
2	High	Unclear whether participants knew of allocation (participants might have deduced due to assessment only control). Intervention conductor was aware of assigned intervention during trial (control was assessment-only, so any intervention was indicative of being in the intervention group). Deviations from intended intervention unclear.
3F	High	Post-intervention exclusions made for those who could not complete follow-up (2/107 subjects, 1.8%, values per group not stated). Further exclusions were made for subjects who did not change in use over the measurement period. Potential risk of inflated effects
4F	High	Appropriate measurement tool, but inappropriate unit (per episode). Measurement did not differ between groups. Outcome assessor was the participant (data self-reported), unclear whether they were blinded to allocation
5F	High	Unclear whether analysis was conducted according to a protocol. No evidence of selective reporting of outcomes and insignificant were results reported. Potential selective reporting of analyses as change data was not presented, only at-follow-up data
3Q	High	Same as 3F
4Q	High	Same as 4F, though inappropriate unit is ‘any consumption, even a sip’
5Q	High	Unclear whether analysis was conducted according to a protocol. Potential selective reporting of outcomes in that significance was not reported. Potential selective reporting of analyses as at-follow-up data was not presented, only change data
3A	Some	Post-intervention exclusions made for those who could not complete follow-up (1/143 subjects, 0.7%, values per group not stated)

Bias domain	Risk	Support for judgement
4A	Some	Same as 4F, though unit is appropriate
5A	Some	Unclear whether analysis was conducted according to a protocol. No evidence of selective reporting of outcomes or analyses
Overall F	High	4/5 domains judged high risk of bias
Overall Q	High	4/5 domains judged high risk of bias
Overall A	High	3/5 domains judged high risk of bias

Note. Risk of bias rating system and domain types in Appendix I. Frequency at follow-up, quantity change, abstinence rate, and abstinence maintenance rate reported. F = frequency outcomes; Q = quantity outcomes; A = abstinence outcomes.

Chang et al. (2005) Risk of Bias Table

Bias domain	Risk	Support for judgement
1	Some	Allocation sequence generation involved simple randomisation. Unclear how this was conducted. Status of allocation concealment unclear. No significant differences between groups regarding baseline alcohol use (no information on baseline demographic differences)
2	High	Participants and intervention conductor were aware of assigned intervention during trial (participants were informed and control was assessment-only, so any intervention was indicative of being in the intervention group). Unclear whether nurse intervention conductors were blinded to allocation. Deviations from intended intervention unclear
3F	Low	All subjects were included in analysis, even the 5% who did not complete follow-up
4F	High	Appropriate measurement tool, but inappropriate unit (per episode). Measurement did not differ between groups. Outcome assessor was the participant (data self-reported), who was not blinded to allocation

Bias domain	Risk	Support for judgement
5F	High	Unclear whether analysis was conducted according to a protocol. Potential selective reporting of outcomes in that significance was not reported. Potential selective reporting of analyses as at-follow-up data was not reported for each group, just for the overall sample. Change data was reported for each group
3Q	Low	Same as 3F
4Q	Some	Same as 4F, though unit is appropriate
5Q	High	Same as 5F
Overall F	High	3/5 domains judged high risk of bias
Overall Q	High	2/5 domains judged high risk of bias

Note. Risk of bias rating system and domain types in Appendix I. Frequency at follow-up and quantity at follow-up reported. F = frequency outcomes; Q = quantity outcomes.

Handmaker et al. (1999) Risk of Bias Table

Bias domain	Risk	Support for judgement
1	Some	Allocation sequence generation involved stratified randomisation using prepared envelopes. Stratification was to ensure 2:1 ratio of light/ moderate to heavy drinkers. Status of allocation concealment unclear. No information on baseline differences
2	High	Unclear whether participants knew of allocation (control was a reduced intervention, so blind allocation was possible). Intervention conductor was aware of assigned intervention during trial (control did not involve conductors, so any intervention was indicative of being in the intervention group). Deviations from intended intervention unclear

Bias domain	Risk	Support for judgement
3F	High	High risk of biases from post-intervention exclusions made for those who could not complete follow-up (8/42 subjects, 19%; five from intervention group, three from control). Though, dropouts did not significantly differ from retained subjects on any baseline indices
4F	Some	Appropriate measurement tool. Measurement did not differ between groups. Outcome assessor was the participant (data self-reported), unclear whether they were blinded to allocation
5F	High	Unclear whether analysis was conducted according to a protocol. No evidence of selective reporting of outcomes and insignificant results reported. Potential selective reporting of analyses as change data was not presented, only at-follow-up data
3Q	High	Same as F3
4Q	High	Same as F4, though note that use of standard ethanol content as a use measure was not appropriate
5Q	High	Same as F5
3A	High	Same as F3
4A	Some	Same as F4
5A	Some	Unclear whether analysis was conducted according to a protocol. No evidence of selective report of outcomes or analyses
Overall F	High	3/5 domains judged high risk of bias
Overall Q	High	4/5 domains judged high risk of bias
Overall A	High	2/5 domains judged high risk of bias

Note. Risk of bias rating system and domain types in Appendix I. Frequency at follow-up, quantity change, abstinence rate, and abstinence maintenance rate reported. F = frequency outcomes; Q = quantity outcomes; A = abstinence outcomes.

Joya et al. (2016) Risk of Bias Table

Bias domain	Risk	Support for judgement
1	Some	Allocation sequence generation involved simple randomisation conducted by computer algorithm. Status of allocation sequence concealment is unclear. No significant differences between groups regarding baseline demographics (no information on baseline alcohol use differences)
2	High	Unclear whether participants or intervention conductors knew of allocation (likely that conductors knew, and participants might have deduced due to usual care control). Deviations from intended intervention unclear
3A	Low	All subjects were included in analysis
4A	Low	Appropriate measurement tool. Measurement did not differ between groups. Hair analysis conductors were blinded to allocation
5A	Some	Unclear whether analysis was conducted according to a protocol. No evidence of selective reporting of outcomes or analyses and insignificant results reported
Overall A	High	1/5 domains judged high risk of bias

Note. Risk of bias rating system and domain types in Appendix I. Abstinence rate and abstinence maintenance rate reported. A = abstinence outcomes.

Ondersma et al. (2015) Risk of Bias Table

Bias domain	Risk	Support for judgement
1	Some	Allocation sequence generation involved simple randomisation conducted by computer algorithm. Status of allocation sequence concealment is unclear. No significant differences between groups regarding any baseline demographics measures
2	Some	Participants were made aware of study details, but it is unclear if this included details of intervention. If not, then participants (who were also self-conductors) may have been blinded to allocation. No deviations from intended intervention (pre-set computer program)
3F	Low	All subjects were included in analysis
4F	Some	Appropriate measurement tool. Measurement did not differ between groups. Unclear whether participants, who self-administered use measurement tools, were aware of allocation
5F	High	Unclear whether analysis was conducted according to a protocol. Potential selective reporting of outcomes as raw frequency scores were not provided (retrieved from contacting authors), though insignificant summary results were. No evidence of selective reporting of analyses
3Q	Low	Same as 3F
4Q	Some	Same as 3F
5Q	High	Unclear whether analysis was conducted according to a protocol. Potential selective reporting of outcomes and analyses, as no quantity data was provided in text despite being recorded (retrieved from contacting authors)
3A	Low	Same as 3F
4A	Some	Same as 3F

Bias domain	Risk	Support for judgement
5A	Some	Unclear whether analysis was conducted according to a protocol. No evidence of selective reporting of outcomes or analyses and insignificant results reported
Overall F	High	1/5 domains judged high risk of bias
Overall Q	High	1/5 domains judged high risk of bias
Overall A	High[1]	4/5 domains judged some risk of bias, so deemed high risk overall

Note. Risk of bias rating system and domain types in Appendix I. Frequency at follow-up, quantity change, abstinence rate, and abstinence maintenance rate reported. F = frequency outcomes; Q = quantity outcomes; A = abstinence outcomes.

[1] Multiple domains were assessed as posing some risk of bias to a degree that the overall outcome was judged high risk also.

Osterman and Dyehouse (2012) Risk of Bias Table

Bias domain	Risk	Support for judgement
1	Some	Allocation sequence generation involved simple randomisation. Status of allocation sequence concealment is unclear. No significant differences between groups regarding baseline demographics (no information on baseline alcohol use differences)
2	High	Unclear whether participants knew of allocation (participants might have deduced due to assessment only control). Intervention conductor was aware of assigned intervention during trial (control did not involve conductors, so any intervention was indicative of being in the intervention group). No deviations from intended interventions, half of interventions were assessed by motivational interviewing experts

Bias domain	Risk	Support for judgement
3F	High	High risk of biases from post-intervention exclusions made for those who could not complete follow-up (11/67 subjects, 16.4%; 10 from intervention group, 1 from control). Though, dropouts did not significantly differ from retained subjects on any baseline indices. A further three exclusions were made for outliers for change data. The groups that these exclusions came from is not reported.
4F	Some	Appropriate measurement tool. Measurement did not differ between groups. Unclear whether research assistants who gathered data were blinded to allocation
5F	Some	Unclear whether analysis was conducted according to a protocol. No evidence of selective reporting of outcomes or analyses, and insignificant data reported
3Q	High	Same as 3F
4Q	High	Same as 4F, unclear whether unit of drinks per day is drinks per drinking day or per reporting day
5Q	Some	Same as 5F
Overall F	High	2/5 domains judged high risk of bias
Overall Q	High	3/5 domains judged high risk of bias

Note. Risk of bias rating system and domain types in Appendix I. Frequency at follow-up, frequency change, quantity at follow-up and quantity change reported. F = frequency outcomes; Q = quantity outcomes.

Osterman et al. (2014) Risk of Bias Table

Bias domain	Risk	Support for judgement
1	Low	Allocation sequence generation involved simple randomisation conducted by computer algorithm. Allocation sequence concealed until after baseline assessment. No significant differences between groups regarding baseline demographics measures (no information on baseline alcohol use differences)
2	High	Participants were made aware of their group assignment. Intervention conductor was aware of assigned intervention during trial (control did not involve conductors, so any intervention was indicative of being in the intervention group). No deviations from intended interventions, half of interventions were assessed by motivational interviewing experts
3F	High	Post-intervention exclusions made for those who could not complete follow-up (29/122 subjects, 23.8%, 18 from intervention group, 11 from control). A further two exclusions were made per group for outlier data (4/122, 3.28%)
4F	Some	Appropriate measurement tool. Measurement did not differ between groups. Unclear whether research assistants who gathered data were blinded to allocation
5F	High	Unclear whether analysis was conducted according to a protocol. No evidence of selective reporting of outcomes and insignificant data reported. Potential selective reporting of analyses, as change data was not reported
3Q	High	Same as 3F
4Q	High	Same as 4F, unclear whether unit of drinks per day is drinks per drinking day or per reporting day
5Q	High	Same as 5F
Overall F	High	3/5 domains judged high risk of bias
Overall Q	High	4/5 domains judged high risk of bias

Note. Risk of bias rating system and domain types in Appendix I. Frequency at follow-up, frequency change, quantity at follow-up and quantity change reported. F = frequency outcomes; Q = quantity outcomes.

Reynolds et al. (1995) Risk of Bias Table

Bias domain	Risk	Support for judgement
1	Some	Allocation sequence generation involved simple randomisation. Unclear how this was conducted. Status of allocation sequence concealment is unclear. No significant differences between groups regarding baseline demographics or consumption levels
2	High	Participants were made aware of study details, but it is unclear if this included details of intervention. If not, then participants may have been blinded to allocation. Intervention conductor was aware of assigned intervention during trial (control did not involve conductors, so any intervention was indicative of being in the intervention group). Deviations from intended intervention unclear
3A	High	Post-intervention exclusions made for those who could not complete follow-up (6/78 subjects, 7.7%, three per group). Dropouts differed from inclusions by religion, baseline gestation, and baseline beer consumption (significance of these differences unclear)
4A	Some	Appropriate measurement tool. Measurement did not differ between groups. Unclear whether participants, who self-administered use measurement tools, were aware of allocation
5A	Some	Unclear whether analysis was conducted according to a protocol. No evidence of selective reporting of outcomes or analyses and insignificant results reported
Overall A	High	2/5 domains judged high risk of bias

Note. Risk of bias rating system and domain types in Appendix I. Abstinence rate reported. A = abstinence outcomes.

Appendix N: Risk of Bias Assessment Comparisons with Gomez et al. (2020)

All risk of bias assessments conducted in this review were also conducted by Gomez et al. (2020). Gomez et al. (2020) assessed risk of bias on the study level, whereas this review did so on the outcome level. To allow comparison, the highest rating of each domain across outcomes was considered for the ratings from this review (i.e., high risk for frequency, but low for abstinence means high overall rating).

Overall Similarity of Agreements/Disagreements with Gomez et al. (2020)

Bias domain	Whether or not risk of bias assessments match between studies								Similarity
	(Chang et al., 1999)	(Chang et al., 2005)	(Handmaker et al., 1999)	(Joya et al., 2016)	(Ondersma et al., 2015)	(Osterman & Dyehouse, 2012)	(Osterman et al., 2014)	(Reynolds et al., 1995)	
1	☑	☒	☒	☑	☑	☑	☑	☒	5/8
2	☑	☒	☒	☒	☑	☒	☒	☑	3/8
3	☒	☑	☑	☑	☑	☑	☒	☒	5/8
4	☑	☑	☒	☑	☒	☑	☑	☒	5/8
5	☒	☒	☒	☑	☒	☑	☒	☑	3/8
	3/5	2/5	1/5	4/5	3/5	4/5	2/5	2/5	21/40
Overall bias	☑	☑	☑	☒	☑	☑	☑	☑	7/8

Note: ☑ = same judgment; ☒ = different judgement.

Specific Similarity of Agreements/Disagreements with Gomez et al. (2020): Domain Judgements

Judgement	Low	Some	High	Total
Low	5	0	0	5
Some	3	8	3	14
High	5	8	8	21
Total	13	16	11	40

Specific Similarity of Agreements/Disagreements with Gomez et al. (2020): Overall Judgements

Judgement	Low	Some	High	Total
Low	0	0	0	0
Some	0	0	0	0
High	0	1	7	8
Total	0	1	7	8

Appendix O: Primary Data Exclusions, Transformations and Assumptions

Primary Data Exclusions and Transformations

Study citation	Outcome	Excluded/ transformed	Exclusion/transformation notes
(Chang et al., 1999)	FFU	EXCL	Excluded due to episodic units, no standard drink measure, no post-exclusion sample sizes
	QFU	EXCL	Excluded due to episodic units, no standard drink measure, no post-exclusion sample sizes
	AMR	TRANS	Required number of baseline abstainers. Baseline ARs of 58% IG, 56% CG. Conversions per transformation #5[1]
(Chang et al., 2005)	FFU	TRANS	FFU reported as 1.9% IG and 2% CG drinking days per 158-day reporting period with an effect of brief intervention of FFU of $b = 0.802$ (SE ±0.587). Mean drinking days per month conversions made per transformation #2. Associated SD values also converted per transformation #2, but first required generation using the reported SE values per transformations #3
	QFU	EXCL	Excluded due to having episodic units
(Handmaker et al., 1999)	FFU	EXCL	Excluded due to providing covariate test statistics only
	QFU	EXCL	Excluded due to providing covariate test statistics only
	AR	TRANS	Reported as 67% IG, 56% CG not abstaining at follow-up. Values inverted to 44% IG, 33% CG abstaining per conversion #9. Conversions to AR then made per transformation #5
	AMR	EXCL	Excluded due to no base abstinence rates being reported, and authors noting high risk of false reporting by participants
(Joya et al., 2016)	-	-	-
(Ondersma et al., 2015)	FFU	TRANS	Extra data provided by authors. FFU reported as 0.45 (±1.61) IG, 1.11 (±3.21) CG drinking days per 90-day reporting period. Conversions per transformation #1
(Osterman & Dyehouse, 2012)	FC	TRANS	FC reported as -0.875 (±0.919) IG, -1.38 (±1.25) CG drinking days per week. Conversions per transformation #1[1]

Study citation	Outcome	Excluded/ transformed	Exclusion/transformation notes
(Osterman et al., 2014)	FFU	TRANS	FFU reported as 0.00 (±0.00) IG, 0.00 (±0.00) CG drinking days per week. Conversion intended to be made per transformation #1, but were not necessary given all-zero values
	FC	EXCL	Excluded. Conversions intended to be made per transformation #4. However, transformation #4 requires a correlation coefficient, which was not provided. Instead the change SD needed to be estimated using the correlation statistics of another, similar study (Higgins et al., 2021b). The only other study with FC data was Osterman and Dyehouse (2012). Osterman and Dyehouse (2012) reported SD values of 0.00, rendering the correlation coefficient calculation impossible, and their reported coefficients are for between-group effects, not within-group as required. As such, no correlation statistic was available to generate the FC SDs.
	QC	EXCL	Excluded due to the same reasons as Osterman et al. (2014) FC data
(Reynolds et al., 1995)	AR	TRANS	Reported as 88% IG, 69% CG abstinent at follow-up. Conversions per transformation #5

Note. All conversions were made per predefined transformations. If viable outcomes were included but not listed here, then they did not require any transformations. No secondary data exclusions or transformations made for Joya et al. (2016). Tzilos et al. (2011) data is not included here as the study was not eligible for analysis. FFU = frequency at follow-up; FC = frequency change, QFU = quantity at follow-up; QC = quantity change; AR = abstinence rate; AMR = abstinence maintenance rate; EXCL = excluded; TRANS = transformed; IG = intervention group; CG = control group; SD = standard deviation.

[1] Assumptions were made to this data. Assumptions listed on next page.

Assumptions Made for Primary Outcome Data

Study citation	Assumption
(Chang et al., 1999)	(1) For abstinence maintenance rate data, baseline sample sizes per group were calculated to be 71 per group. This did not align with reports of 143 participants total. It is unclear why these values did not align, perhaps due to rounding errors. The 71 per group values were retained with a likely reduction in data quality noted
(Joya et al., 2016)	(1) For abstinence rate and abstinence maintenance rate data, EtG analysis method may have picked up some postnatal use for third trimester data. However, said data was reported as third trimester use only, so assumed to meet eligibility
(Osterman & Dyehouse, 2012)	(1) It is unclear whether FC and QC data considered baseline reports of past 30-day or past 12-months alcohol use, though the size of the values would suggest the latter. These data should be interpreted with caution, as they likely capture higher levels of use from before participants knew they were pregnant, or even started trying (2) Three exclusions were made for change data. It was not reported which groups there were excluded from so one participant was removed from each baseline sample to mitigate these exclusions as best as possible (result was 28 IG, 26 CG). Potential reductions in data quality are noted (3) Quantity data was reported per day, though it is unclear whether this represented drinking days or days in the measurement period (authors did not respond for clarification). It was assumed that the drinking day interpretation was correct, as these values were used to generate a composite consumption score (drinking days * drinks per day) in, which would not make sense with per measured day units
(Osterman & Dyehouse, 2014)	(1) Same as Osterman and Dyehouse (2012) assumption (3). Though, Osterman et al. (2014) did not calculate a composite score, so it had to be assumed that the same unit definitions were used based on their similar methodologies and conducted by the same first author

Note. EtG = hair ethyl glucuronide quantification; FC = frequency change; QC = quantity change; IG = intervention group; CG = control group.

Appendix P: Secondary Data Exclusions, Transformations, and Categorisations

Secondary Data Exclusions and Transformations

Study citation, IG, CG *n* at base	Outcome	Excluded/ transformed	Exclusion/transformation notes
(Chang et al., 1999), 123, 127	Education	TRANS	Education reported as 11% HSG, 29% some college, 34% 4-year college degree, and 22% completed graduate or professional school for total sample. These four demographics were combined to generate ≥HSG data per conversion #10
(Chang et al., 2005), 152, 152	Age	EXCL	Excluded due to being reported as total sample median age
	Education	EXCL	Excluded due to being reported as total sample median education level
	Nulliparity	TRANS	Nulliparity reported as 43.4% IG, 40.4% CG. Conversions per transformation #7
	Base gestation	EXCL	Excluded due to being reported as median base gestation per group
(Handmaker et al., 1999), 20, 22	Education	EXCL	Excluded due to being reported as total sample mean years of education
	Marital status	TRANS	Marital status reported as 62% unmarried. Inverted to 38% married per conversion #9
	Standard drink measure	TRANS	Standard drink reported as 15mL pure alcohol. Conversions per transformation #6
(Joya et al., 2016), 83, 85	Age	TRANS	Age reported as $\bar{x}$ 32.3 (±5.0) IG, 29.9 (±5.7) CG. Conversions per transformation #8
	Education	TRANS	Education reported as 42.5% HSG and 21.2% university educated IG, 48.2% HSG and 19.7% university educated CG. HSG and university educated sample were combined per transformation #10 and then converted to total sample data per transformation #7
	Nulliparity	TRANS	Nulliparity reported as 58.6% IG, 58.6% CG having previous pregnancies. Values inverted to reflect no previous pregnancies per transformation #9 and then converted per transformation #7
(Ondersma et al., 2015), 24, 24	Age	EXCL	Excluded due to being reported as total sample age ranges
(Osterman & Dyehouse, 2012), 39, 38	ALL DATA	EXCL	All secondary data was excluded as values provided included eligible participants

Study citation, IG, CG *n* at base	Outcome	Excluded/ transformed	Exclusion/transformation notes
(Osterman et al., 2014), 62, 60	Age	TRANS	Age reported as x̄ 25.3 (±4.67) IG, 25.6 (±4.98) CG. Conversions per transformation #8
	Ethnic majority	TRANS	Ethnic majority reported as 61.3% IG, 55% CG black. Only needed to confirm majority, not necessarily generate the total sample number/percentage. Conversion per transformations #7
	Education	TRANS	Education reported as 24.2% HSG, 42% some college, 1.6% 4-year college degree, and 1.6% college graduate IG, 25% HSG, 41.7% some college, 3.3% 4-year college degree, and 0% college graduate CG. These four demographics were combined per transformation #10 and then converted per transformation #7
	Marital status	TRANS	Marital status reported as 16.1% IG, 13.3% CG married. Conversions per transformation #7
	Nulliparity	EXCL	Excluded due to nulliparity being reported as x̄ 1.95 (±2.03) births
	Base gestation	TRANS	Base gestation reported as x̄ 23.6 (±8.72) IG, 23.1 (±8.72) CG weeks. Conversions per transformation #8
(Reynolds et al., 1995), 42, 36	Marital status	TRANS	Marital status reported as 65% single or divorced. Inverted to 35% married per transformation #9.

Note. All conversions were made per predefined transformations. If viable outcomes were included but not listed here, then they did not require any transformations. EXCL = excluded; TRANS = transformed. IG = intervention group; CG = control group; HSG = high school graduate.

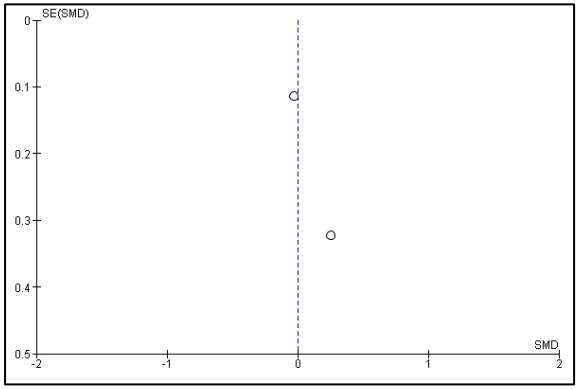
Secondary Data Categorisations

Study citation	Outcome	Reported as	Categorised as
(Chang et al., 1999)	Sampling location	Obstetrics clinic	Specialist clinic
	Ethnic majority	Caucasian	White
	Use measurement period	Base-to-delivery, x̄ 22.4 (±5.6) weeks	≥2-months
(Chang et al., 2005)	Sampling location	Obstetrics clinic	Specialist clinic

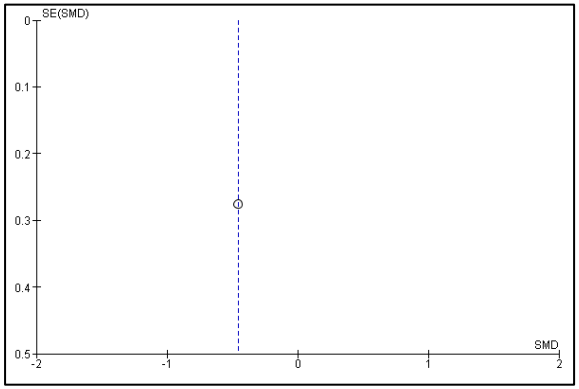
Study citation	Outcome	Reported as	Categorised as
	Use measurement period	Base-to-delivery, $\bar{x}$ 158 days	≥2-months
(Handmaker et al., 1999)	Use measurement period	2-months	≥2-months
	Control type	Letters about risks of antenatal alcohol consumption and referrals	Reduced care
(Joya et al., 2016)	Ethnic majority	Spanish	Other
	Use measurement period	Base-to-delivery	≥2-months
(Ondersma et al., 2015)	Sampling location	Prenatal care clinic	Specialist clinic
	Use measurement period	90-days	≥2-months
	Bonus item	Three tailored letters	Bonus item = yes
	Control type	Less-interactive brief computer intervention about infant nutrition	Reduced care
(Osterman & Dyehouse, 2012)	Sampling location	Prenatal care clinic	Specialist clinic
	Ethnic majority	Non-Hispanic African American	African American
	Use measurement period	4-6 weeks	<2-months
(Osterman et al., 2014)	Sampling location	Prenatal care clinic	Specialist clinic
	Ethnic majority	Black	African American
	Use measurement period	1-month	<2-months
	Intervention theory	Motivational interviewing based on self-determination theory	Motivational interviewing
(Reynolds et al., 1995)	Sampling location	Maternity clinic	Specialist clinic
	Use measurement period	2-months	≥2-months
	Intervention theory	Education and social cognition theory	Education
	Bonus item	Self-help manual and one-minute call to check manual progress	Bonus item = yes
(Tzilos et al., 2011)	Sampling location	Prenatal care clinic	Specialist clinic
	Use measurement period	1-month	<2-months
	Control type	Computer activity about television preferences	Reduced care

Appendix Q: Meta-Analyses Funnel Plots

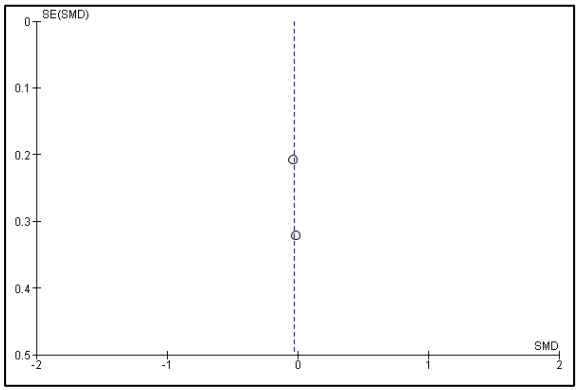
Funnel Plot for Frequency at Follow-Up Outcome



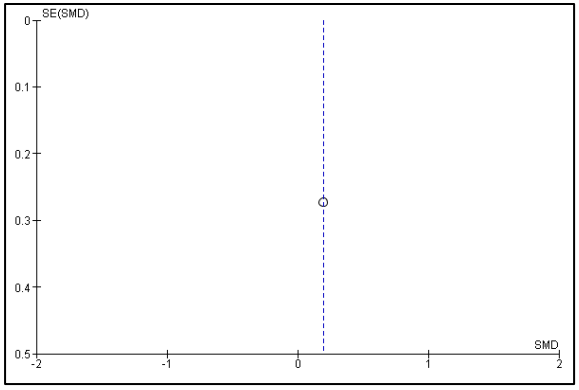
Funnel Plot for Frequency Change Outcome



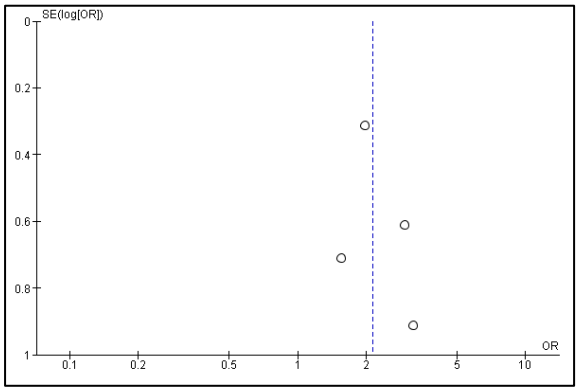
Funnel Plot for Quantity at Follow-Up Outcome



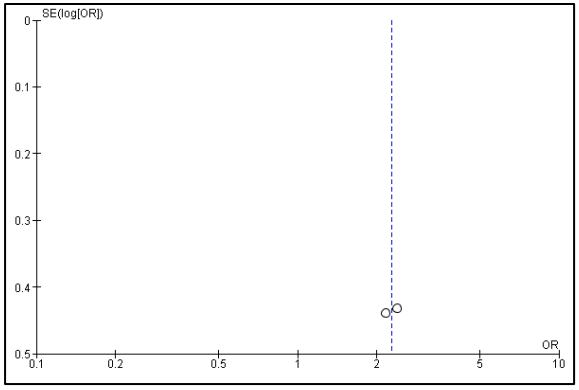
Funnel Plot for Quantity Change Outcome



Funnel Plot for Abstinence Rate Outcome



Funnel Plot for Abstinence Maintenance Rate Outcome

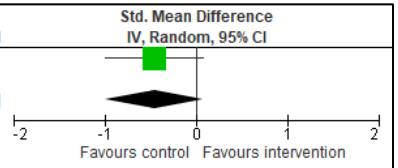


Appendix R: Meta-Analyses Ineligible Comparison Tables

Comparison Table for Frequency Change Outcome

Study or Subgroup	Control group Mean	SD	Total	Intervention group Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Osterman & Dyehouse, 2012 (1)	-5.52	5	26	-3.5	3.68	28	100.0%	-0.46 [-1.00, 0.09]
Total (95% CI)			**26**			**28**	**100.0%**	**-0.46 [-1.00, 0.09]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.65 (P = 0.10)

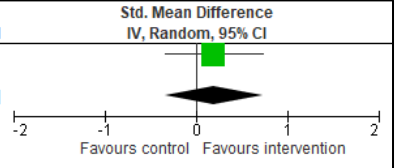


Footnotes
(1) Data required transformation, sample sizes are estimates based on baseline values and number of total exclusions

Comparison Table for Quantity Change Outcome

Study or Subgroup	Control group Mean	SD	Total	Intervention group Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Osterman & Dyehouse, 2012 (1)	-1.93	1.11	26	-2.18	1.4	28	100.0%	0.19 [-0.34, 0.73]
Total (95% CI)			**26**			**28**	**100.0%**	**0.19 [-0.34, 0.73]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.71 (P = 0.48)



Footnotes
(1) Data units assumed to be drinks per drinking day (reported as drinks per day), sample sizes are estimates based on baseline values and number of total...

Appendix S: Subgroup Analyses Results

Viable Subgroup Analysis Outcomes

Outcome type	Subgroup number	Study variable(s) assessed	Number of participants (studies)	Relative effect [95% CI]	Test for overall effect	Test for subgroup differences
FFU	2.1.1	[1]	304 (1)	Hedges' *g* = -0.03 [-0.26, 0.19]	*Z* = 0.27, *p* = 0.79	-
	2.1.2	[1]	39 (1)	Hedges' *g* = 0.26 [-0.37, 0.89]	*Z* = 0.80, *p* = 0.43	Chi^2 = 0.71, *df* = 1 (*p* = 0.40)
QFU	3.1.1	[1]	39 (1)	Hedges' *g* = -0.01 [-0.64, 0.62]	*Z* = 0.04, *p* = 0.97	-
	3.1.2	[1]	93 (1)	Hedges' *g* = -0.03 [-0.44, 0.37]	*Z* = 0.16, *p* = 0.87	Chi^2 = 0.00, *df* = 1 (*p* = 0.96)
AR	4.1.1	Pilot study	73 (2)	Odds ratio = 2.05 [0.68, 6.13]	*Z* = 1.28, *p* = 0.20	-
	4.1.2	Non-pilot	240 (2)	Odds ratio = 2.15 [1.24, 3.72]*	*Z* = 2.74, *p* = 0.006	Chi^2 = 0.01 *df* = 1 (*p* = 0.94)
AR	4.2.1	Baseline abstainers included	207 (2)	Odds ratio = 2.08 [1.16, 3.73]*	*Z* = 2.47, *p* = 0.01	-
	4.2.2	Baseline abstainers excluded	72 (1)	Odds ratio = 2.96 [0.89, 9.79]	*Z* = 1.78, *p* = 0.08	Chi^2 = 0.27 *df* = 1 (*p* = 0.61)

Outcome type	Subgroup number	Study variable(s) assessed	Number of participants (studies)	Relative effect [95% CI]	Test for overall effect	Test for subgroup differences
AR	4.3.1	Sampling from specialist clinic/AA ethnic majority	111 (2)	Odds ratio = 3.03 [1.12, 8.20]*	$Z = 2.19$, $p = 0.03$	-
	4.3.2	Sampling from general hospital/other ethnic majority	168 (1)	Odds ratio = 1.98 [1.07, 3.66]*	$Z = 2.17$, $p = 0.03$	
	4.3.3	Sampling from other location/Hispanic ethnic majority	34 (1)	Odds ratio = 1.56 [0.39, 6.25]	$Z = 0.62$, $p = 0.53$	$Chi^2 = 0.74$ $df = 2$ ($p = 0.69$)
AR	4.4.1	T-ACE screen	39 (1)	Odds ratio = 3.21 [0.54, 19.11]	$Z = 1.28$, $p = 0.20$	-
	4.4.2	AUDIT screen	168 (1)	Odds ratio = 1.98 [1.07, 3.66]*	$Z = 2.17$, $p = 0.03$	
	4.4.3	No formal tool for screening	106 (2)	Odds ratio = 2.25 [0.91, 5.57]	$Z = 1.75$, $p = 0.08$	$Chi^2 = 0.27$ $df = 2$ ($p = 0.87$)
AR	4.5.1	Brief Drinker Profile use measurement tool	34 (1)	Odds ratio = 1.56 [0.39, 6.25]	$Z = 0.62$, $p = 0.53$	-
	4.5.2	No formal use measurement tool	72 (1)	Odds ratio = 2.96 [0.89, 9.79]	$Z = 1.78$, $p = 0.08$	$Chi^2 = 0.47$ $df = 1$ ($p = 0.49$)
AR	4.6.1	Face-to-face delivery	274 (3)	Odds ratio = 2.06 [1.24, 3.43]*	$Z = 2.78$, $p = 0.005$	-
	4.6.2	Computer delivery	39 (1)	Odds ratio = 3.21 [0.54, 19.11]	$Z = 1.28$, $p = 0.20$	$Chi^2 = 0.22$ $df = 1$ ($p = 0.64$)

Outcome type	Subgroup number	Study variable(s) assessed	Number of participants (studies)	Relative effect [95% CI]	Test for overall effect	Test for subgroup differences
AR	4.7.1	Motivational interviewing intervention	241 (3)	Odds ratio = 1.99 [1.17, 3.41]*	$Z = 2.52$, $p = 0.01$	-
	4.7.2	Education-based intervention	72 (1)	Odds ratio = 2.96 [0.89, 9.79]	$Z = 1.78$, $p = 0.08$	$Chi^2 = 0.35$ $df = 1$ ($p = 0.56$)
AR	4.8.1	Bonus item = yes	111 (2)	Odds ratio = 3.03 [1.12, 8.20]*	$Z = 2.19$, $p = 0.03$	-
	4.8.2	Bonus item = no	202 (2)	Odds ratio = 1.90 [1.08, 3.34]*	$Z = 2.24$, $p = 0.03$	$Chi^2 = 0.64$ $df = 1$ ($p = 0.42$)
AR	4.9.1	Usual care control	168 (1)	Odds ratio = 1.98 [1.07, 3.66]*	$Z = 2.17$, $p = 0.03$	-
	4.9.2	Reduced care control	145 (3)	Odds ratio = 2.42 [1.08, 5.44]*	$Z = 2.14$, $p = 0.03$	$Chi^2 = 0.15$ $df = 1$ ($p = 0.70$)
AR	4.10.1	Abstinence recommended	34 (1)	Odds ratio = 1.56 [0.39, 6.25]	$Z = 0.62$, $p = 0.53$	-
	4.10.2	Personalised goal setting recommended	39 (1)	Odds ratio = 3.21 [0.54, 19.11]	$Z = 1.28$, $p = 0.20$	$Chi^2 = 040$ $df = 1$ ($p = 0.53$)
AMR	5.1.1	[1]	142 (1)	Odds ratio = 2.39 [1.03, 5.57]*	$Z = 2.02$, $p = 0.04$	-
	5.1.2	[1]	101 (1)	Odds ratio = 2.17 [0.92, 5.12]	$Z = 1.76$, $p = 0.08$	$Chi^2 = 0.03$ $df = 1$ ($p = 0.87$)

Note. No individual subgroups reported heterogeneity. Not enough different categories for subgroup analyses of women with high-level use or dependencies and use measurement period for abstinence rate. FFU = frequency at follow-up; QFU = quantity at follow-up; AR = abstinence rate; AMR = abstinence maintenance rate; AA = African American; T-ACE = Tolerance, Annoyed, Cut-down, Eye-opener; NFT = no formal tool; AUDIT = Alcohol Use Disorders Identification Test.

[1] Subgroup analysis applies to several variables with the same studies (variables outlined on next page).

* *p* value <0.05.

List of Study Variables Assessed for Each Multi-Variable Subgroup Analyses

Outcome type	Subgroup number	Variables assessed	Variables excluded[1]
FFU	2.1.1, 2.1.2	2.1.1 (Chang et al., 2005): Non-pilot, women with high-level use and dependencies excluded, white ethnic majority, screen and use measurement assessed by research assistant, intervention delivered face-to-face (by first author and nurse), abstinence recommended, bonus item = no, usual care control 2.1.2 (Ondersma et al., 2015): Women with high-level use and dependencies included, African American ethnic majority, screen and use measurement self-administered, intervention delivered by computer (self-administered), personalised goal setting recommended, bonus item = yes, reduced care control	Presence of baseline abstainers, sampling location, screening tool, use measurement tool, use measurement period, and intervention theory

Outcome type	Subgroup number	Variables assessed	Variables excluded[1]
QFU	3.1.1, 3.1.2	3.1.1: Pilot study, T-ACE screen, ATFB use measurement tool, screen and use measurement self-administered, use measurement period ≥2-months, intervention delivered by computer (self-administered), bonus item = yes, reduced care control 3.1.2: Non-pilot, AUDIT screen, QDS use measurement tool, screen and use measurement assessed by research assistant, use measurement period <2-months, intervention delivered face-to-face (by first author), bonus item = no, usual care control	Presence of women with dependencies, presence of baseline abstainers, sampling location, ethnic majority, intervention theory, use recommendation
AMR	5.1.1, 5.1.2	5.1.1 (Chang et al., 1999): Women with high-level use and dependencies excluded, sampling from specialist clinic, white ethnic majority, T-ACE screen 5.1.2 (Joya et al., 2016): Women with high-level use and dependencies included, sampling from general hospital, other ethnic majority, AUDIT screen	Pilot status, presence of baseline abstainers, use measurement tool, screen and use measurement conductor, intervention conductor, intervention delivery method, intervention theory, use recommendation, presence of bonus item, control type

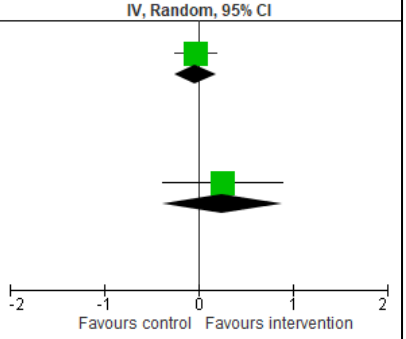
FFU = frequency at follow-up; QFU = quantity at follow-up; AMR = abstinence maintenance rate; T-ACE = Tolerance, Annoyed, Cut-down, Eye-opener; NFT = no formal tool; AUDIT = Alcohol Use Disorders Identification Test; ATFB = Alcohol Timeline Follow-back; QDS = Quick Drinking Screen.

[1] Excluded as there was not enough variance in categories reported across studies for this to be assessed.

Appendix T: Subgroup Analyses Comparison Tables

Comparison Table for Frequency at Follow-Up Subgroups 2.1.1 and 2.1.2

Study or Subgroup	Control group Mean	SD	Total	Intervention group Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.1.1 Chang et al., 2005								
Chang et al., 2005 (1)	-0.56	0.907	152	-0.532	0.907	152	100.0%	-0.03 [-0.26, 0.19]
Subtotal (95% CI)			**152**			**152**	**100.0%**	**-0.03 [-0.26, 0.19]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.27 (P = 0.79)								
2.1.2 Ondersma et al., 2015								
Ondersma et al., 2015 (2)	0.345	0.999	19	0.14	0.501	20	100.0%	0.26 [-0.37, 0.89]
Subtotal (95% CI)			**19**			**20**	**100.0%**	**0.26 [-0.37, 0.89]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.80 (P = 0.43)								



Test for subgroup differences: Chi² = 0.71, df = 1 (P = 0.40), I² = 0%

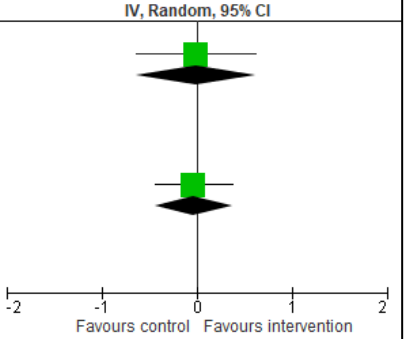
Footnotes

(1) Data required transformation, several subgroups represented

(2) Data provided by authors, data required transformation, several subgroups represented

Comparison Table for Quantity at Follow-Up Subgroups 3.1.1 and 3.1.2

Study or Subgroup	Control group Mean	SD	Total	Intervention group Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
3.1.1 Ondersma et al., 2015								
Ondersma et al., 2015 (1)	0.53	1.18	19	0.55	1.88	20	100.0%	-0.01 [-0.64, 0.62]
Subtotal (95% CI)			**19**			**20**	**100.0%**	**-0.01 [-0.64, 0.62]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.04 (P = 0.97)								
3.1.2 Osterman et al., 2014								
Osterman et al., 2014 (2)	0.04	0.29	49	0.05	0.3	44	100.0%	-0.03 [-0.44, 0.37]
Subtotal (95% CI)			**49**			**44**	**100.0%**	**-0.03 [-0.44, 0.37]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.16 (P = 0.87)								



Test for subgroup differences: Chi² = 0.00, df = 1 (P = 0.96), I² = 0%

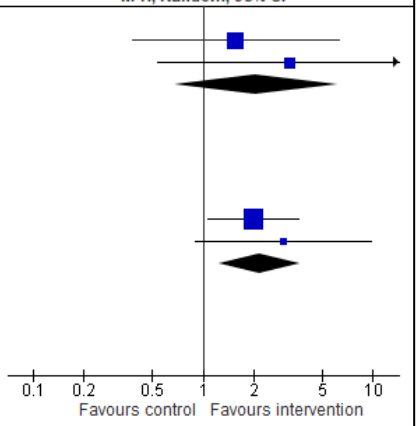
Footnotes

(1) Data provided by authors, several subgroups represented

(2) Data units assumed to be drinks per drinking day (reported as drinks per day), several subgroups represented

Comparison Table for Abstinence Rate Subgroups 4.1.1 and 4.1.2

Study or Subgroup	Intervention group Events	Total	Control group Events	Total	Weight	Odds Ratio M-H, Random, 95% CI
4.1.1 Pilot study						
Handmaker et al., 1999 (1)	7	16	6	18	62.1%	1.56 [0.39, 6.25]
Ondersma et al., 2015 (2)	18	20	14	19	37.9%	3.21 [0.54, 19.11]
Subtotal (95% CI)		**36**		**37**	**100.0%**	**2.05 [0.68, 6.13]**
Total events	25		20			
Heterogeneity: Tau² = 0.00; Chi² = 0.40, df = 1 (P = 0.53); I² = 0%						
Test for overall effect: Z = 1.28 (P = 0.20)						
4.1.2 Non-pilot						
Joya et al., 2016	52	83	39	85	79.0%	1.98 [1.07, 3.66]
Reynolds et al., 1995 (3)	34	39	23	33	21.0%	2.96 [0.89, 9.79]
Subtotal (95% CI)		**122**		**118**	**100.0%**	**2.15 [1.24, 3.72]**
Total events	86		62			
Heterogeneity: Tau² = 0.00; Chi² = 0.34, df = 1 (P = 0.56); I² = 0%						
Test for overall effect: Z = 2.74 (P = 0.006)						



Test for subgroup differences: Chi² = 0.01, df = 1 (P = 0.94), I² = 0%

Footnotes

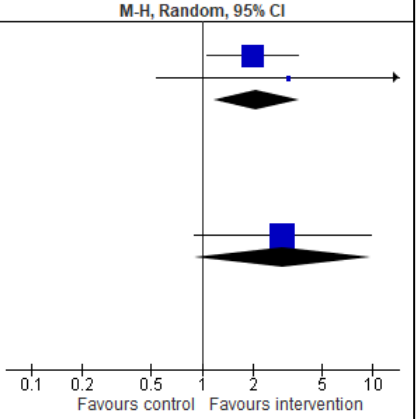
(1) Data required transformation

(2) Data provided by authors, data required transformation

(3) Data required transformation

Comparison Table for Abstinence Rate Subgroups 4.2.1 and 4.2.2

Study or Subgroup	Intervention group Events	Total	Control group Events	Total	Weight	Odds Ratio M-H, Random, 95% CI
4.2.1 Baseline abstainers included						
Joya et al., 2016 (1)	52	83	39	85	89.3%	1.98 [1.07, 3.66]
Ondersma et al., 2015 (2)	18	20	14	19	10.7%	3.21 [0.54, 19.11]
Subtotal (95% CI)		**103**		**104**	**100.0%**	**2.08 [1.16, 3.73]**
Total events	70		53			
Heterogeneity: Tau² = 0.00; Chi² = 0.25, df = 1 (P = 0.61); I² = 0%						
Test for overall effect: Z = 2.47 (P = 0.01)						
4.2.2 Baseline abstainers excluded						
Reynolds et al., 1995 (3)	34	39	23	33	100.0%	2.96 [0.89, 9.79]
Subtotal (95% CI)		**39**		**33**	**100.0%**	**2.96 [0.89, 9.79]**
Total events	34		23			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.78 (P = 0.08)						



Test for subgroup differences: Chi² = 0.27, df = 1 (P = 0.61), I² = 0%

Footnotes

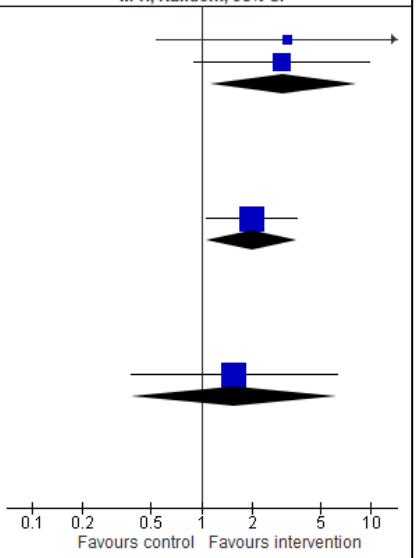
(1) Several subgroups represented

(2) Data provided by authors, data required transformation, several subgroups represented

(3) Data required transformation, several subgroups represented

Comparison Table for Abstinence Rate Subgroups 4.3.1, 4.3.2, and 4.3.3

Study or Subgroup	Intervention group Events	Total	Control group Events	Total	Weight	Odds Ratio M-H, Random, 95% CI
4.3.1 Sampling from specialist clinic/ African American ethnic majority						
Ondersma et al., 2015 (1)	18	20	14	19	31.1%	3.21 [0.54, 19.11]
Reynolds et al., 1995 (2)	34	39	23	33	68.9%	2.96 [0.89, 9.79]
Subtotal (95% CI)		**59**		**52**	**100.0%**	**3.03 [1.12, 8.20]**
Total events	52		37			
Heterogeneity: Tau² = 0.00; Chi² = 0.01, df = 1 (P = 0.94); I² = 0%						
Test for overall effect: Z = 2.19 (P = 0.03)						
4.3.2 Sampling from general hospital/Other ethnic majority						
Joya et al., 2016	52	83	39	85	100.0%	1.98 [1.07, 3.66]
Subtotal (95% CI)		**83**		**85**	**100.0%**	**1.98 [1.07, 3.66]**
Total events	52		39			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.17 (P = 0.03)						
4.3.3 Sampling from other location/Hispanic ethnic majority						
Handmaker et al., 1999 (3)	7	16	6	18	100.0%	1.56 [0.39, 6.25]
Subtotal (95% CI)		**16**		**18**	**100.0%**	**1.56 [0.39, 6.25]**
Total events	7		6			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.62 (P = 0.53)						



Test for subgroup differences: Chi² = 0.74, df = 2 (P = 0.69), I² = 0%

Footnotes

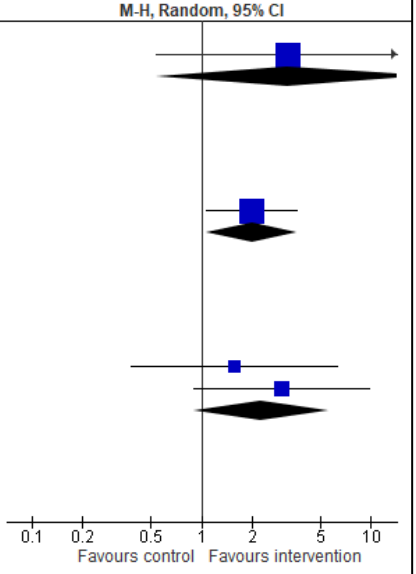
(1) Data provided by authors, data required transformation

(2) Data required transformation

(3) Data required transformation

Comparison Table for Abstinence Rate Subgroups 4.4.1, 4.4.2, and 4.4.3

Study or Subgroup	Intervention group Events	Total	Control group Events	Total	Weight	Odds Ratio M-H, Random, 95% CI
4.4.1 T-ACE screen						
Ondersma et al., 2015 (1)	18	20	14	19	100.0%	3.21 [0.54, 19.11]
Subtotal (95% CI)		**20**		**19**	**100.0%**	**3.21 [0.54, 19.11]**
Total events	18		14			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.28 (P = 0.20)						
4.4.2 AUDIT screen						
Joya et al., 2016 (2)	52	83	39	85	100.0%	1.98 [1.07, 3.66]
Subtotal (95% CI)		**83**		**85**	**100.0%**	**1.98 [1.07, 3.66]**
Total events	52		39			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.17 (P = 0.03)						
4.4.3 No formal tool for screen						
Handmaker et al., 1999 (3)	7	16	6	18	42.5%	1.56 [0.39, 6.25]
Reynolds et al., 1995 (4)	34	39	23	33	57.5%	2.96 [0.89, 9.79]
Subtotal (95% CI)		**55**		**51**	**100.0%**	**2.25 [0.91, 5.57]**
Total events	41		29			
Heterogeneity: Tau² = 0.00; Chi² = 0.47, df = 1 (P = 0.49); I² = 0%						
Test for overall effect: Z = 1.75 (P = 0.08)						



Test for subgroup differences: Chi² = 0.27, df = 2 (P = 0.87), I² = 0%

Footnotes

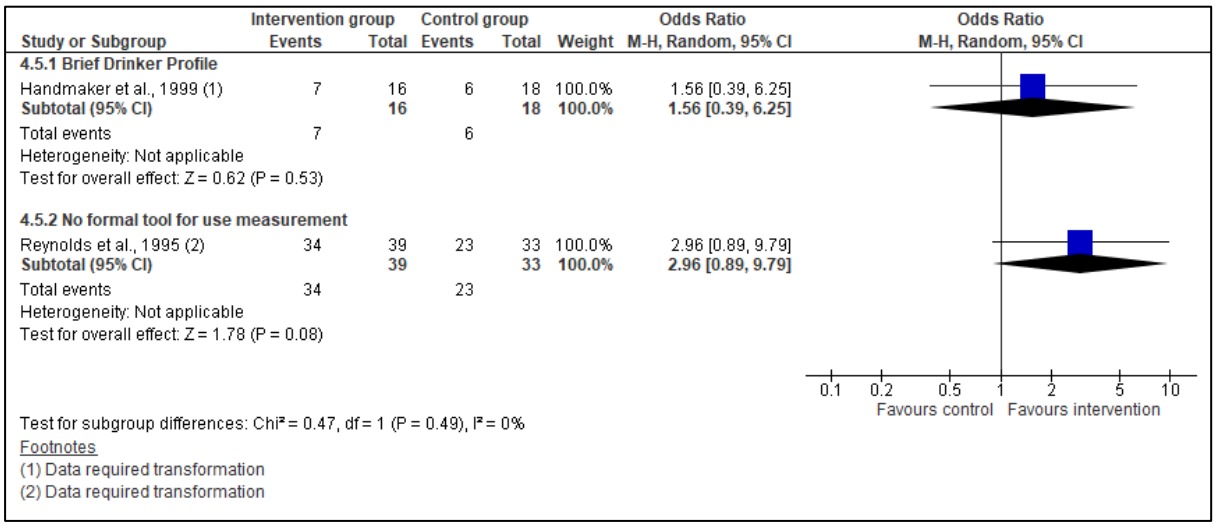
(1) Data provided by authors, data required transformation, several subgroups represented

(2) Several subgroups represented

(3) Data required transformation, several subgroups represented

(4) Data required transformation, several subgroups represented

Comparison Table for Abstinence Rate Subgroups 4.5.1 and 4.5.2



Study or Subgroup	Intervention group Events	Total	Control group Events	Total	Weight	Odds Ratio M-H, Random, 95% CI
4.5.1 Brief Drinker Profile						
Handmaker et al., 1999 (1)	7	16	6	18	100.0%	1.56 [0.39, 6.25]
Subtotal (95% CI)		**16**		**18**	**100.0%**	**1.56 [0.39, 6.25]**
Total events	7		6			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.62 (P = 0.53)						
4.5.2 No formal tool for use measurement						
Reynolds et al., 1995 (2)	34	39	23	33	100.0%	2.96 [0.89, 9.79]
Subtotal (95% CI)		**39**		**33**	**100.0%**	**2.96 [0.89, 9.79]**
Total events	34		23			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.78 (P = 0.08)						

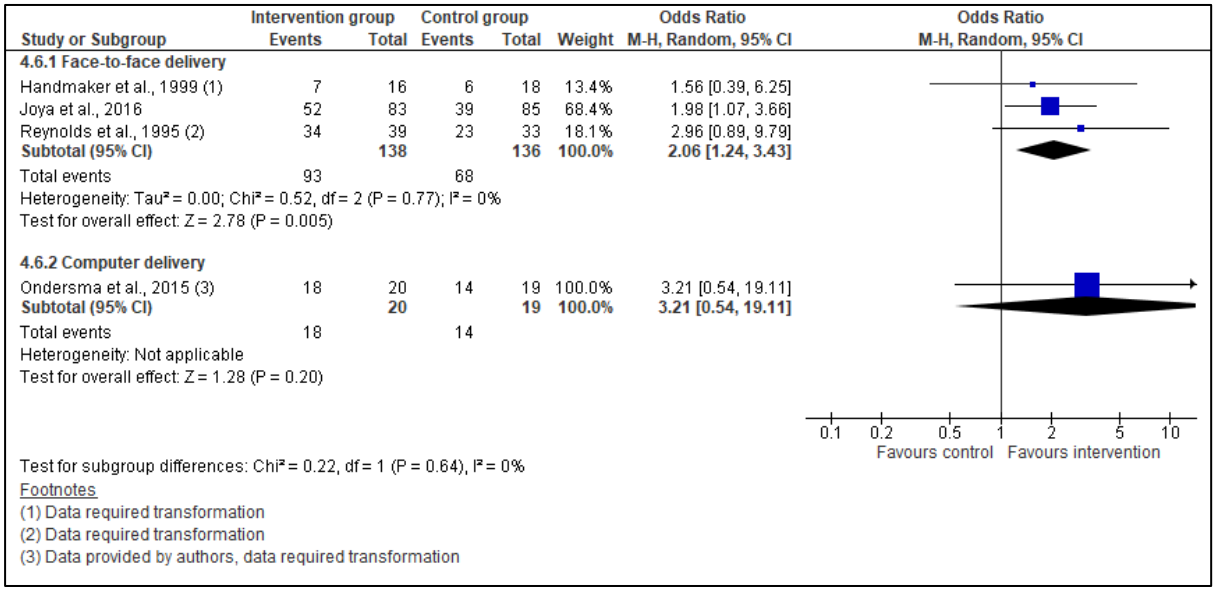
Test for subgroup differences: Chi² = 0.47, df = 1 (P = 0.49), I² = 0%

Footnotes

(1) Data required transformation

(2) Data required transformation

Comparison Table for Abstinence Rate Subgroups 4.6.1 and 4.6.2



Study or Subgroup	Intervention group Events	Total	Control group Events	Total	Weight	Odds Ratio M-H, Random, 95% CI
4.6.1 Face-to-face delivery						
Handmaker et al., 1999 (1)	7	16	6	18	13.4%	1.56 [0.39, 6.25]
Joya et al., 2016	52	83	39	85	68.4%	1.98 [1.07, 3.66]
Reynolds et al., 1995 (2)	34	39	23	33	18.1%	2.96 [0.89, 9.79]
Subtotal (95% CI)		**138**		**136**	**100.0%**	**2.06 [1.24, 3.43]**
Total events	93		68			
Heterogeneity: Tau² = 0.00; Chi² = 0.52, df = 2 (P = 0.77); I² = 0%						
Test for overall effect: Z = 2.78 (P = 0.005)						
4.6.2 Computer delivery						
Ondersma et al., 2015 (3)	18	20	14	19	100.0%	3.21 [0.54, 19.11]
Subtotal (95% CI)		**20**		**19**	**100.0%**	**3.21 [0.54, 19.11]**
Total events	18		14			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.28 (P = 0.20)						

Test for subgroup differences: Chi² = 0.22, df = 1 (P = 0.64), I² = 0%

Footnotes

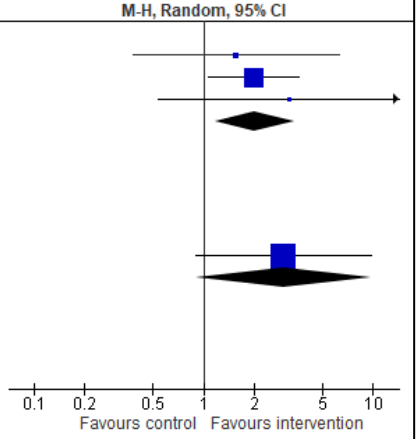
(1) Data required transformation

(2) Data required transformation

(3) Data provided by authors, data required transformation

Comparison Table for Abstinence Rate Subgroups 4.7.1 and 4.7.2

Study or Subgroup	Intervention group Events	Total	Control group Events	Total	Weight	Odds Ratio M-H, Random, 95% CI
4.7.1 Motivational interviewing intervention						
Handmaker et al., 1999 (1)	7	16	6	18	14.9%	1.56 [0.39, 6.25]
Joya et al., 2016	52	83	39	85	76.0%	1.98 [1.07, 3.66]
Ondersma et al., 2015 (2)	18	20	14	19	9.1%	3.21 [0.54, 19.11]
Subtotal (95% CI)		**119**		**122**	**100.0%**	**1.99 [1.17, 3.41]**
Total events	77		59			
Heterogeneity: Tau² = 0.00; Chi² = 0.40, df = 2 (P = 0.82); I² = 0%						
Test for overall effect: Z = 2.52 (P = 0.01)						
4.7.2 Education-based intervention						
Reynolds et al., 1995 (3)	34	39	23	33	100.0%	2.96 [0.89, 9.79]
Subtotal (95% CI)		**39**		**33**	**100.0%**	**2.96 [0.89, 9.79]**
Total events	34		23			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.78 (P = 0.08)						



Test for subgroup differences: Chi² = 0.35, df = 1 (P = 0.56), I² = 0%

Footnotes

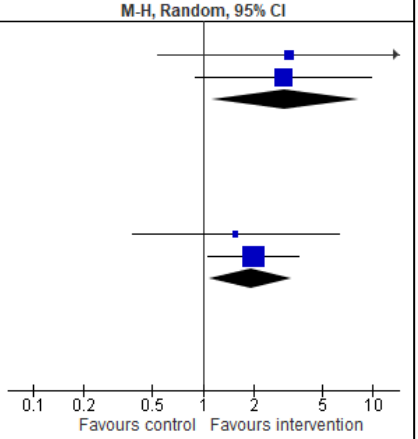
(1) Data required transformation

(2) Data provided by authors, data required transformation

(3) Data required transformation

Comparison Table for Abstinence Rate Subgroups 4.8.1 and 4.8.2

Study or Subgroup	Intervention group Events	Total	Control group Events	Total	Weight	Odds Ratio M-H, Random, 95% CI
4.8.1 Bonus item = yes						
Ondersma et al., 2015 (1)	18	20	14	19	31.1%	3.21 [0.54, 19.11]
Reynolds et al., 1995 (2)	34	39	23	33	68.9%	2.96 [0.89, 9.79]
Subtotal (95% CI)		**59**		**52**	**100.0%**	**3.03 [1.12, 8.20]**
Total events	52		37			
Heterogeneity: Tau² = 0.00; Chi² = 0.01, df = 1 (P = 0.94); I² = 0%						
Test for overall effect: Z = 2.19 (P = 0.03)						
4.8.2 Bonus item = no						
Handmaker et al., 1999 (3)	7	16	6	18	16.4%	1.56 [0.39, 6.25]
Joya et al., 2016	52	83	39	85	83.6%	1.98 [1.07, 3.66]
Subtotal (95% CI)		**99**		**103**	**100.0%**	**1.90 [1.08, 3.34]**
Total events	59		45			
Heterogeneity: Tau² = 0.00; Chi² = 0.10, df = 1 (P = 0.76); I² = 0%						
Test for overall effect: Z = 2.24 (P = 0.03)						



Test for subgroup differences: Chi² = 0.64, df = 1 (P = 0.42), I² = 0%

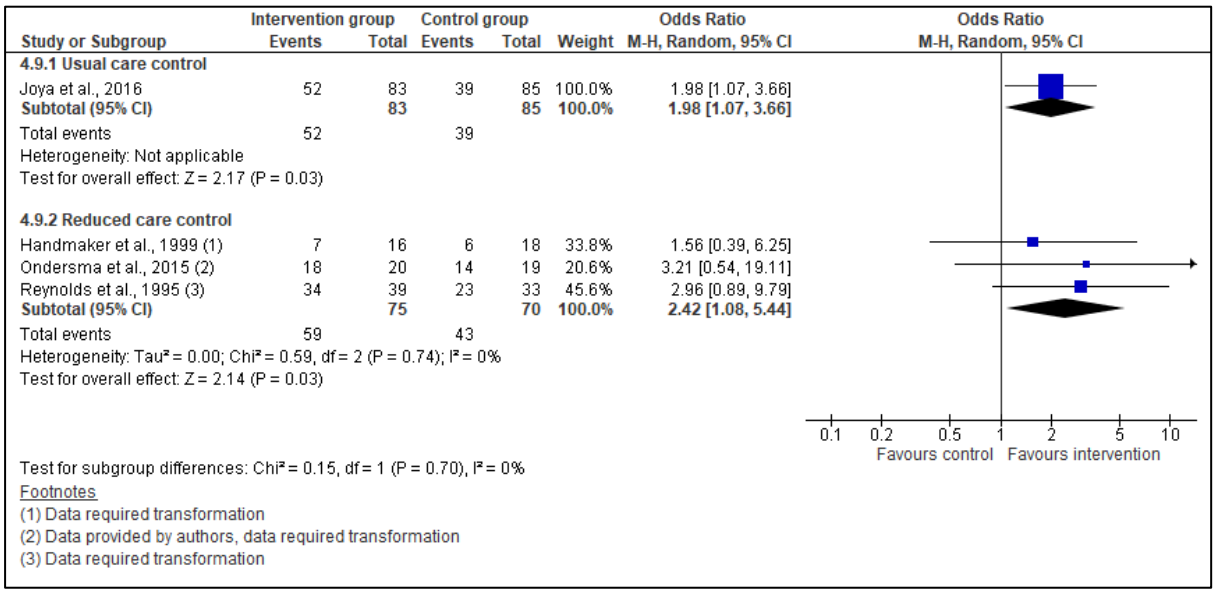
Footnotes

(1) Data provided by authors, data required transformation

(2) Data required transformation

(3) Data required transformation

Comparison Table for Abstinence Rate Subgroups 4.9.1 and 4.9.2



Study or Subgroup	Intervention group Events	Total	Control group Events	Total	Weight	Odds Ratio M-H, Random, 95% CI
4.9.1 Usual care control						
Joya et al., 2016	52	83	39	85	100.0%	1.98 [1.07, 3.66]
Subtotal (95% CI)		**83**		**85**	**100.0%**	**1.98 [1.07, 3.66]**
Total events	52		39			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.17 (P = 0.03)						
4.9.2 Reduced care control						
Handmaker et al., 1999 (1)	7	16	6	18	33.8%	1.56 [0.39, 6.25]
Ondersma et al., 2015 (2)	18	20	14	19	20.6%	3.21 [0.54, 19.11]
Reynolds et al., 1995 (3)	34	39	23	33	45.6%	2.96 [0.89, 9.79]
Subtotal (95% CI)		**75**		**70**	**100.0%**	**2.42 [1.08, 5.44]**
Total events	59		43			
Heterogeneity: Tau² = 0.00; Chi² = 0.59, df = 2 (P = 0.74); I² = 0%						
Test for overall effect: Z = 2.14 (P = 0.03)						

Test for subgroup differences: Chi² = 0.15, df = 1 (P = 0.70), I² = 0%

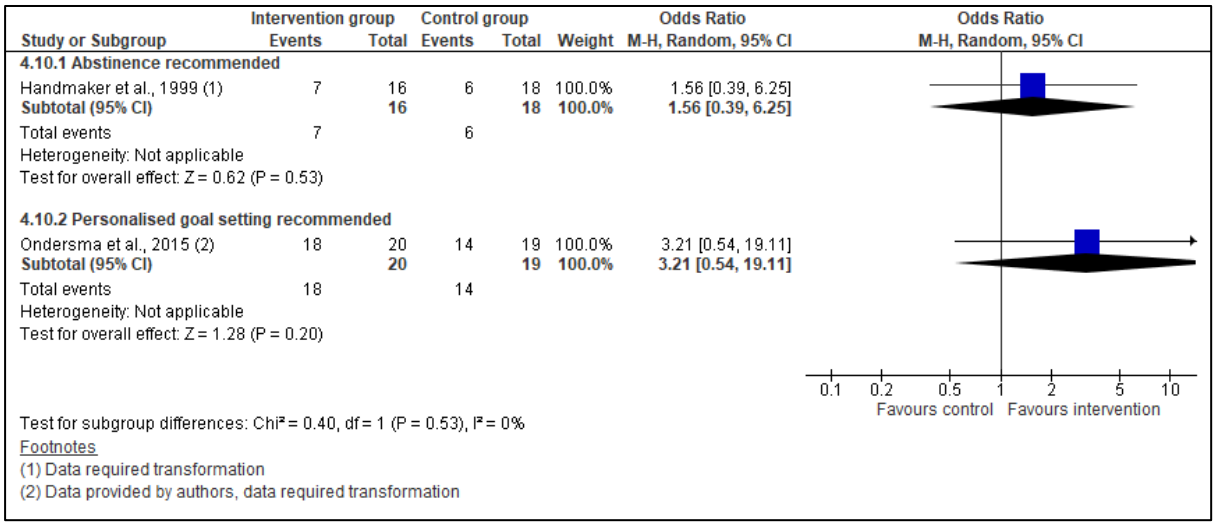
Footnotes

(1) Data required transformation

(2) Data provided by authors, data required transformation

(3) Data required transformation

Comparison Table for Abstinence Rate Subgroups 4.10.1 and 4.10.2



Study or Subgroup	Intervention group Events	Total	Control group Events	Total	Weight	Odds Ratio M-H, Random, 95% CI
4.10.1 Abstinence recommended						
Handmaker et al., 1999 (1)	7	16	6	18	100.0%	1.56 [0.39, 6.25]
Subtotal (95% CI)		**16**		**18**	**100.0%**	**1.56 [0.39, 6.25]**
Total events	7		6			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.62 (P = 0.53)						
4.10.2 Personalised goal setting recommended						
Ondersma et al., 2015 (2)	18	20	14	19	100.0%	3.21 [0.54, 19.11]
Subtotal (95% CI)		**20**		**19**	**100.0%**	**3.21 [0.54, 19.11]**
Total events	18		14			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.28 (P = 0.20)						

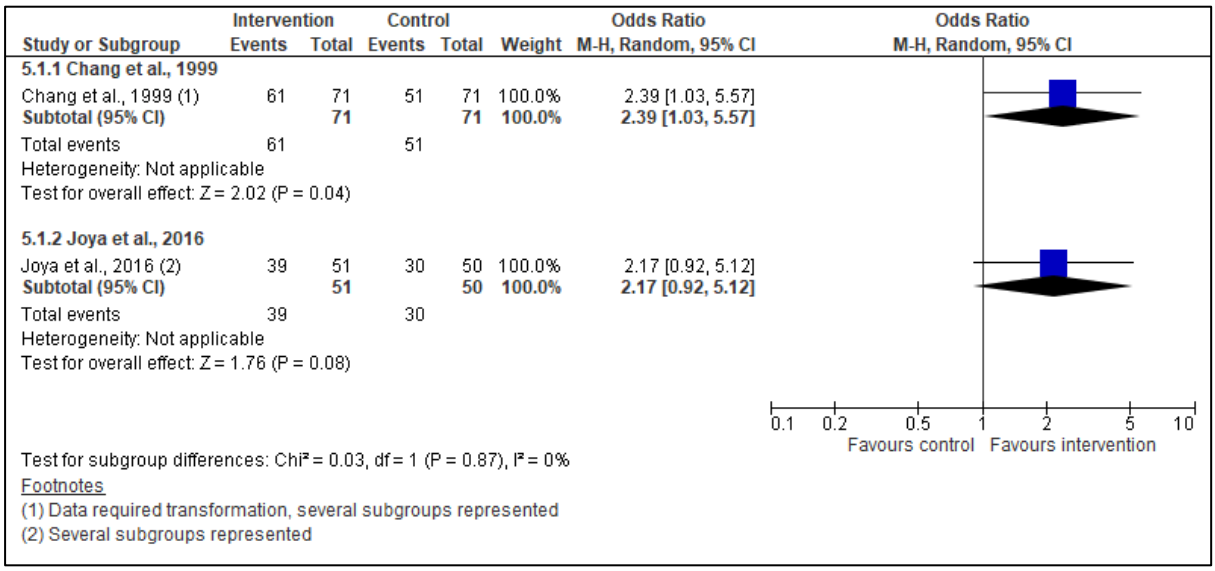
Test for subgroup differences: Chi² = 0.40, df = 1 (P = 0.53), I² = 0%

Footnotes

(1) Data required transformation

(2) Data provided by authors, data required transformation

Comparison Table for Abstinence Maintenance Rate Subgroups 5.1.1 and 5.2.1



Study or Subgroup	Intervention Events	Intervention Total	Control Events	Control Total	Weight	Odds Ratio M-H, Random, 95% CI
5.1.1 Chang et al., 1999						
Chang et al., 1999 (1)	61	71	51	71	100.0%	2.39 [1.03, 5.57]
Subtotal (95% CI)		**71**		**71**	**100.0%**	**2.39 [1.03, 5.57]**
Total events	61		51			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.02 (P = 0.04)						
5.1.2 Joya et al., 2016						
Joya et al., 2016 (2)	39	51	30	50	100.0%	2.17 [0.92, 5.12]
Subtotal (95% CI)		**51**		**50**	**100.0%**	**2.17 [0.92, 5.12]**
Total events	39		30			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.76 (P = 0.08)						

Test for subgroup differences: Chi² = 0.03, df = 1 (P = 0.87), I² = 0%

Footnotes

(1) Data required transformation, several subgroups represented

(2) Several subgroups represented

Appendix U: Meta-Regression Results

Meta-Regression Summary of Findings I: Frequency at Follow-Up and Quantity at Follow-Up

Moderator	FFU			QFU		
	B (SE [95% CI])	*Z*, *p*	*β*	B (SE [95% CI])	*Z*, *p*	*β*
Gestation cut-off (weeks)	-	-	-	-0.00265 (0.05 [-0.60, 0.59])	-0.06, 0.955	-1.00
Mean age (years)	-	-	-	-	-	-
% ≥HSG	-	-	-	-0.00923 (0.16 [-2.08, 2.06])	-0.06, 0.955	-1.00
% Married	-0.00481 (0.01 [-0.08, 0.07])	-0.86, 0.392	-1.00	0.00348 (0.06 [-0.78, 0.79])	0.06, 0.955	1.00
% Nulliparous	-	-	-	-	-	-
Mean gestation at base (weeks)	-	-	-	-0.00190 (0.03 [-0.43, 0.42])	-0.06, 0.955	-1.00
g S. drink	-	-	-	-	-	-
Length of intervention (minutes)	-0.05740 (0.07 [-0.91, 0.79])	-0.86, 0.392	-1.00	-0.00212 (0.04 [-0.48, 0.47])	-0.06, 0.955	-1.00

Note. A Negative effect for FQU and QFU outcomes shows that as the moderator increases, the level of use at follow-up decreases (beneficial effect). Missing data reflects studies reporting the same value, or only one study reporting for the moderator. FFU = frequency at follow-up; QFU = quantity at follow-up; HSG = high school graduate; g S. drink = grams of pure alcohol in a standard drink; B = unstandardised correlation coefficient; SE = standard error; *β* = standardised correlation coefficient.

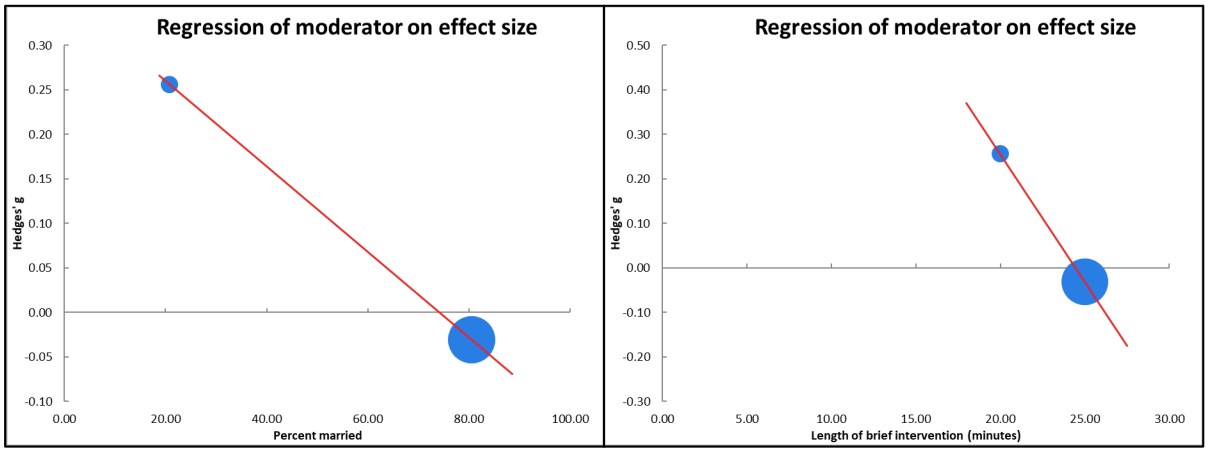
Meta-Regression Summary of Findings II: Abstinence Rate and Abstinence Maintenance Rate

Moderator	AR			AMR		
	B (SE [95% CI])	*Z*, *p*	*β*	B (SE [95% CI])	*Z*, *p*	*β*
Gestation cut-off (weeks)	-0.01 (0.18 [-2.27, 2.25])	-0.06, 0.953	-1.00	-	-	-
Mean age (years)	-0.02 (0.08 [-0.38, 0.35])	-0.19, 0.850	-0.32	0.25 (1.87 [-24.0, 23.5])	-0.13, 0.895	-1.00
% ≥HSG	-0.02 (0.04 [-0.58, 0.54])	-0.39, 0.695	-1.00	0.08 (0.58 [-7.23, 7.38])	0.13, 0.895	1.00
% Married	-0.03 (0.08 [-0.38, 0.31])	-0.43, 0.670	-0.65	-	-	-
% Nulliparous	-		-	0.01 (0.06 [-0.81, 0.83])	0.13, 0.895	1.00
Mean gestation at base (weeks)	-0.84 (14.21 [-181.38, 179.71])	-0.06, 0.953	-1.00	-	-	-
g S. drink	0.36 (0.71 [-8.67, 9.40])	0.51, 0.610	1.00	-	-	-
Length of intervention (minutes)	-0.02 (0.03 [-0.15, 0.11])	-0.65, 0.518	-0.99	-	-	-

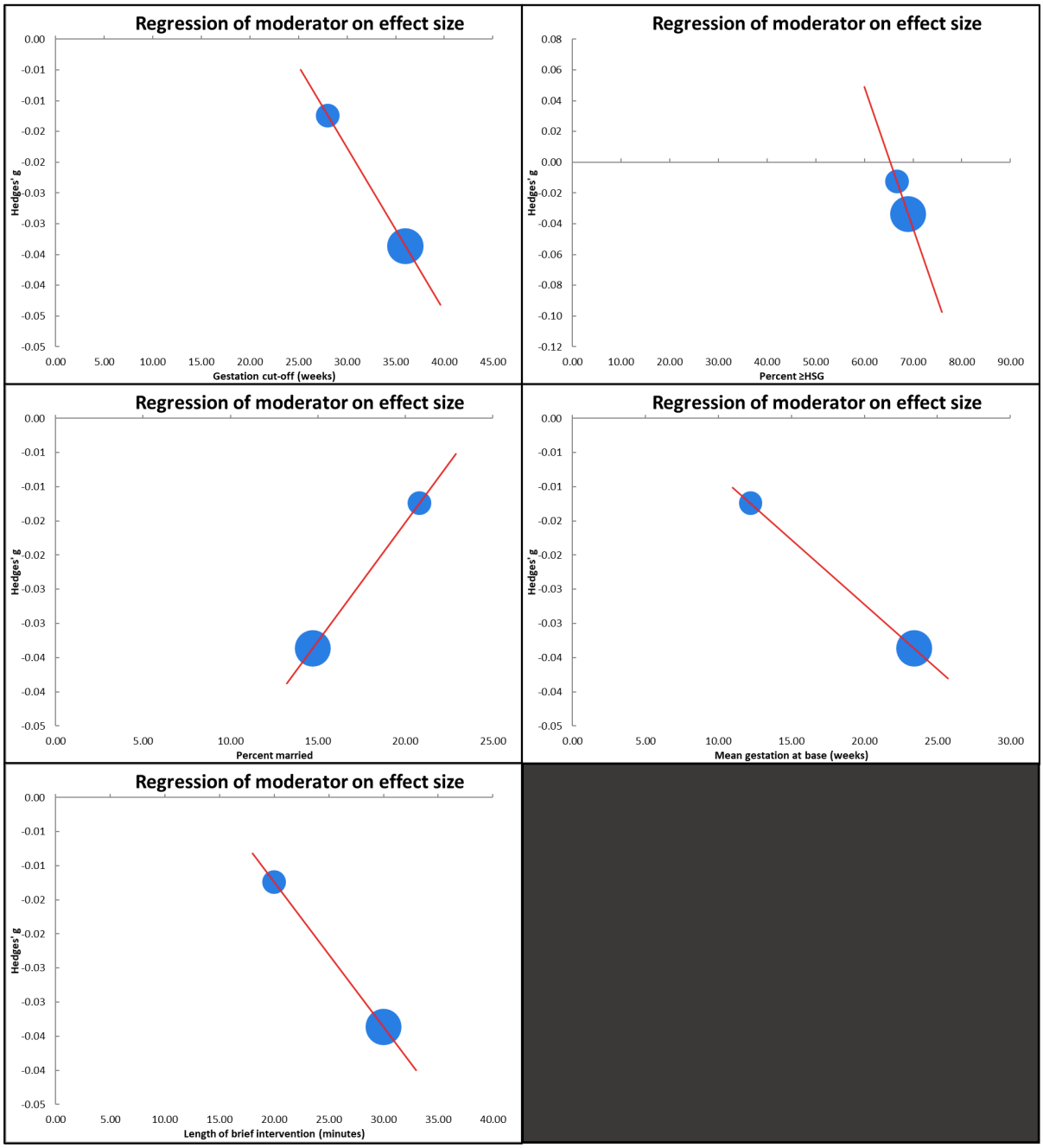
Note. A negative effect for abstinence outcomes shows that odds of abstinence decrease as the moderator increases (detrimental effect). Missing data reflects studies reporting the same value, or only one study reporting for the moderator. AR = abstinence rate; AMR = abstinence maintenance rate; HSG = high school graduate; g S. drink = grams of pure alcohol in a standard drink; B = unstandardised correlation coefficient; SE = standard error; *β* = standardised correlation coefficient.

Appendix V: Meta-Regression Correlation Diagrams

Meta-Regression Graphs for Frequency at Follow-Up Outcomes

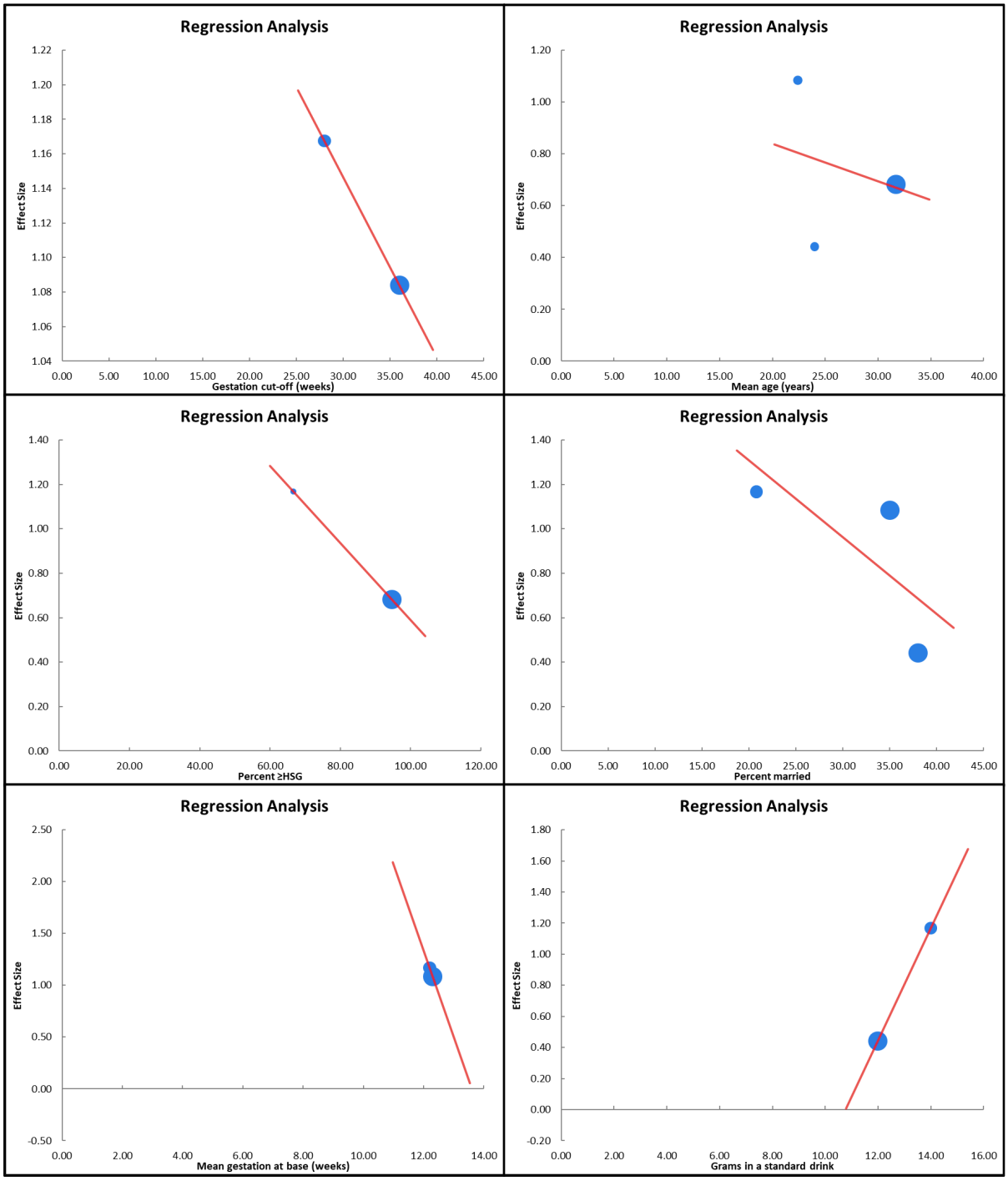


Meta-Regression Graphs for Quantity at Follow-Up Outcomes

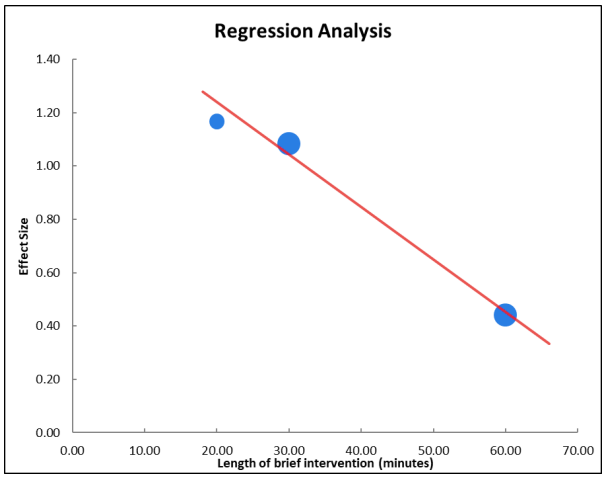


Note. HSG = high school graduate.

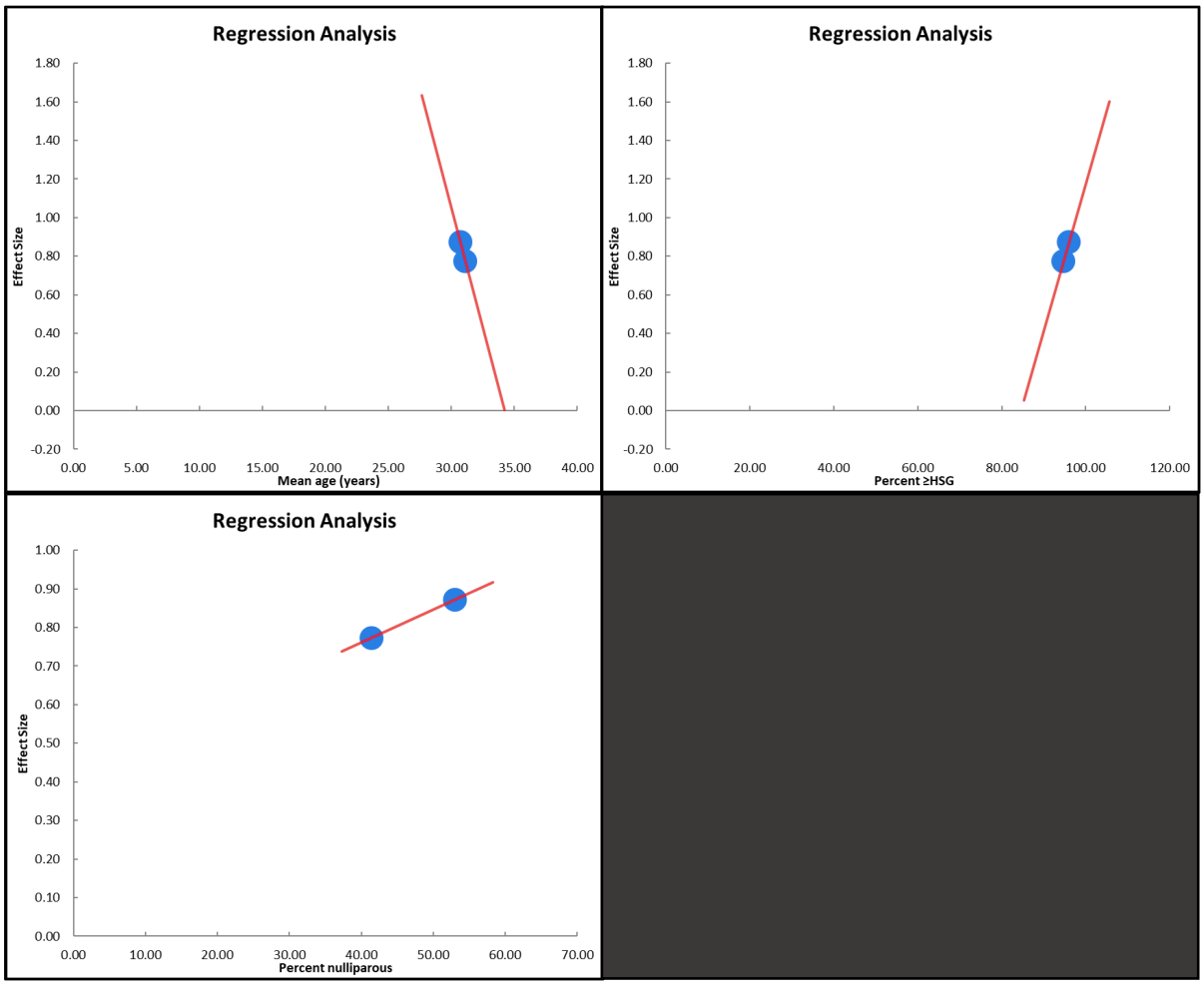
Meta-Regression Graphs for Abstinence Rate Outcomes I



Meta-Regression Graphs for Abstinence Rate Outcomes II



Meta-Regression Graphs for Abstinence Maintenance Rate Outcomes



Note. HSG = high school graduate.

Appendix W: Method and Data Reporting Framework

METHOD AND DATA REPORTING FRAMEWORK: Brief interventions for alcohol use in pregnant women	Page 1/5 *Version 1.0*

Study authors: Year:

Pilot status: □ Pilot study □ Not a pilot study

Sampling and eligibility criteria

Sampling location: □ Specialist pregnancy clinic (obstetrics, maternity, antenatal care clinic etc.)

□ General hospital □ Other: _____

Did intervention occur at the same location?: □ Yes □ No

If not, where did intervention occur?: _____

Were women reporting high-level use or dependencies eligible for the study? □ Yes □ No

Were women reporting abstinence at baseline eligible for the study? □ Yes □ No

What was the eligibility cut-off for baseline gestation?: _____ weeks

Demographics details of sample

	Total sample	Intervention group	Control group
n (at base)	_____	_____	_____
n (at analysis)	_____	_____	_____
Mean age (SD) [years]	_____ (± _____)	_____ (± _____)	_____ (± _____)
Mean gestation at base (SD) [weeks]	_____ (± _____)	_____ (± _____)	_____ (± _____)
Ethnic majority (%)	_____(_____%)	_____(_____%)	_____(_____%)
Percent high school graduated or above or equivalent	_____%	_____%	_____%
Percent married	_____%	_____%	_____%
Percent nulliparous	_____%	_____%	_____%

Notes:

_____

_____

Screen and use measurement details Page 2/5

Screening tool: _____

Tool used to assess level of alcohol consumption: _____

Grams of absolute alcohol defined as a standard drink: _____

Measurement period of intervention (only fill one):

- Aimed for measurement period of: □ _____ weeks □ _____ months □ base-to-delivery
- Mean (SD) measurement period of: □ _____ (± _____) weeks □ _____ (± _____) months
- Measurement was conducted: □ During antenatal period □ During postnatal period

Who conducted these assessments?: *Please tick relevant boxes*	Research assistant	Self-administered	Other (please provide details)
Screening tool			
Baseline use measurement			
Follow-up use measurement			
Other assessment: _____			

Notes:

_____

_____

Intervention details

Brief intervention length: _____minutes

How was the brief intervention delivered?: □ Face-to-face □ By computer □ By phone

Who delivered the brief intervention (and what were their qualifications)?:

_____

What psychotherapeutic theory/theories were employed?: _____

Control details: □ Usual care □ Other: _____

What advice was given?: □ Abstinence recommended □ Personal goal setting recommended (preferred)

Notes (including any additional aspects to the intervention):

_____

_____

Guide for standardised reporting of primary outcomes	Page 3/5

Reporting of frequency outcomes:

Frequency outcomes should be reported using a unit of measurement that can easily be converted to drinking days per 28-day month (base-to-delivery measurements are fine if reported in days, episodic data not viable).

Unit of measurement: Drinking days per □ week □ 28-day month

□ other: _____

	Total sample	Intervention group	Control group
Baseline frequency (SD)	_____ (± _____)	_____ (± _____)	_____ (± _____)
Follow-up frequency (SD)	_____ (± _____)	_____ (± _____)	_____ (± _____)
Change in frequency (SD)	_____ (± _____)	_____ (± _____)	_____ (± _____)

Reporting of quantity outcomes:

Quantity outcomes are to be reported as drinks per drinking day (episodic data not viable).

Please restate grams of pure alcohol considered in one standard drink: _____

	Total sample	Intervention group	Control group
Baseline quantity (SD)	_____ (± _____)	_____ (± _____)	_____ (± _____)
Follow-up quantity (SD)	_____ (± _____)	_____ (± _____)	_____ (± _____)
Change in quantity (SD)	_____ (± _____)	_____ (± _____)	_____ (± _____)

Reporting of abstinence outcomes:

Abstinence outcomes are to be reported as raw numbers and percentages. To assess abstinence maintenance rate, compare number of baseline abstainers who remained abstinent to number abstinent at base.

	Total sample	Intervention group	Control group
Number abstinent at base (%)	_____ (_____%)	_____ (_____%)	_____ (_____%)
Number abstinent at follow-up (%)	_____ (_____%)	_____ (_____%)	_____ (_____%)
Number of baseline abstainers who remained abstinent (%)	_____ (_____%)	_____ (_____%)	_____ (_____%)

Cochrane risk of bias adherence checklist (part 1)	Page 4/5

How to use: Whether a bias indicator was reported on is determined by the leftmost checkbox. Whether or not a bias indicator was addressed in the methodologies is determined by the Yes/No checkboxes next to each indicator.

Domain 1: Bias arising from the randomisation process

☐ Allocation sequence randomised: ☐ Yes ☐ No

Notes: _____

☐ Allocation sequence concealed: ☐ Yes ☐ No

Notes: _____

☐ Significant differences in baseline demographics/use-levels detected: ☐ Yes ☐ No

Notes: _____

Domain 2: Bias due to deviations from intended interventions

☐ Blinding of participants: ☐ Yes ☐ No

Notes: _____

☐ Blinding of screening tool/baseline use measurement tool conductor: ☐ Yes ☐ No

Notes: _____

☐ Blinding of intervention conductor: ☐ Yes ☐ No

Notes: _____

☐ Notable deviations from intended intervention: ☐ Yes ☐ No

Notes: _____

☐ Were interventions reviewed for adherence and quality?: ☐ Yes ☐ No

Notes: _____

Domain 3: Bias due to missing outcome data

☐ Were there data that were excluded from analysis?: ☐ Yes ☐ No

Notes: _____

☐ If yes, how many data points were removed?: ____/____ total participants (_____ from intervention group, _____ from control group)

Notes: _____

Cochrane risk of bias adherence checklist (part 2)	Page 5/5

Domain 4: Bias in measurement of outcome

☐ Validated screening tool used: ☐ Yes ☐ No

Notes: _____

☐ Validated use measurement tool used: ☐ Yes ☐ No

Notes: _____

☐ Data reported in viable units of analysis (defined by this framework): ☐ Yes ☐ No

Notes: _____

☐ Did measurement differ between groups?: ☐ Yes ☐ No

Notes: _____

☐ Blinding of assessor (individual who assessed follow-up use): ☐ Yes ☐ No

Notes: _____

Domain 5: Bias in selection of reported results

☐ Was a protocol developed for this study?: ☐ Yes ☐ No

Notes: _____

☐ If yes, was this protocol adhered to?: ☐ Yes ☐ No

Notes: _____

☐ Were all planned and undertaken analyses reported?: ☐ Yes ☐ No

Notes: _____

☐ Were all outcomes reported?: ☐ Yes ☐ No

Notes: _____

Risk of bias guidelines based off those from: Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A., (2021), Cochrane Handbook for Systematic Reviews of Interventions (version 6.2), Cochrane. https://training.cochrane.org/handbook/current/chapter-08